جمهوری اسلامی ایران

وزارت جهاد کشاورزی
سازمان تحقیقات، آموزش و ترویج کشاورزی
موسسه تحقیقات خاک و آب

نمونه‌برداری در نقشه‌برداری رقومی خاک با استفاده از محیط نرم‌افزار R

(جلد اول - روش‌های نمونه‌برداری مبتنی بر طرح)

نگارندگان

شاهرخ فاتحی، هیات علمی مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان کرمانشاه
محمد جمشیدی، هیات علمی موسسه تحقیقات خاک و آب، کرج
زهره مصلح قهفرخی، هیات علمی مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان چهارمحال بختیاری

نشریه فنی: 622

1401

مشخصات اثر

عنوان: نمونه‌برداری در نقشه‌برداری رقومی خاک با استفاده از محیط نرم‌افزار R (جلد اول - روش‌های نمونه‌برداری مبتنی بر طرح)

نگارندگان: شاهرخ فاتحی، محمد جمشیدی و زهره مصلح قهفرخی

ناشر: موسسه تحقیقات خاک و آب

لیتوگرافی، چاپ و صحافی: انتشارات سنا

ویراستار: زهرا محمدی

طراح جلد: راضیه محمدی

سال انتشار: 1401

نقل مطالب با ذکر منبع بلامانع است.

این اثر با شماره 63074 در تاریخ 1401/12/4 در مرکز اطلاعات و مدارک علمی کشاورزی به ثبت رسیده است.

حق چاپ برای ناشر محفوظ است.

نشانی: کرج، میدان استاندارد، جاده مشکین‌دشت، بلوار امام خمینی (ره)، موسسه تحقیقات خاک و آب

کد پستی: 3177993545 **صندوق پستی:** 311-31785

نمابر: 02636210121 **تلفن:** 36201900- 026

وب‌سایت: http://www.swri.ir **پست الکترونیکی:** info@swri.ir

مسئولیت صحت مطالب به عهده نگارندگان است.

فهرست مندرجات

عنوان **صفحه**

پیشگفتار 1
1- مقدمه 1
2- نصب و راه‌اندازی نرم‌افزارهای لازم 5
2-1- استفاده از R، RSTUDIO و بسته‌های قابل‌الحاق به محیط نرم‌افزار R 5
2-1-2- بارگیری و نصب نرم‌افزار RSTUDIO 6
2-2- نصب نرم‌افزار جاوا و RTOOLS 6
2-3- نصب بسته‌های نرم‌افزاری لازم 7
2-3-1- پرکاربردترین بسته‌های نرم‌افزاری R برای اجرای طرح‌های نمونه‌برداری در نقشه‌برداری رقومی خاک 8
2-4- دستورالعمل‌های موجود برای شروع کار با R 9
3- اجرای روش‌های نمونه‌برداری 9
3-1- نمونه‌برداری محتمل در مقابل نمونه‌برداری غیرمحتمل 10
3-2- طرح‌های نمونه‌برداری هندسی 13
3-2-1- نمونه‌برداری از شبکه منظم 13
3-2-1-1- نمونه‌برداری شبکه منظم بر روی نقشه‌های رستری بر اساس تعداد نقاط و فاصله شبکه 13
3-2-1-2- نمونه‌برداری شبکه منظم بر روی نقشه‌ی چهارگوش 18
3-3- نمونه‌برداری با استفاده KMEANS 31
3-3-1- نمونه‌برداری با استفاده KMEANS سخت 31
3-3-1-1- اجرای طرح نمونه‌برداری KMEANS سخت در R 31
3-3-2- نمونه‌برداری با استفاده KMEANS فازی 38
3-3-3- نمونه‌برداری ابر مکعب لاتین مشروط 55

3-3-3-1- اجرای طرح نمونه‌برداری ابرمکعب لاتین مشروط با بسته نرم‌افزاری SPSANN 62
3-3-3-2- اجرای طرح نمونه‌برداری ابرمکعب لاتین مشروط با بسته نرم‌افزاری CLHS 70
4- بحث و نتیجه‌گیری 78
5- فهرست منابع 79

فهرست جدول‌ها

عنوان **صفحه**

جدول 3-1- خطاهای استاندارد برآورد شده محل تقاطع (B0) و شیب (B1) برای یک نمونه تصادفی ساده (SRS) و نمونه بهینه شده برای رگرسیون خطی ساده. 12
جدول 4-1- مروری بر روش‌های نقشه‌برداری و روش نمونه‌برداری مبتنی بر طرح متناسب با آن 78

فهرست شکل‌ها

عنوان **صفحه**

شکل 3-1- از چپ به راست: نمونه تصادفی ساده (الف)، نمونه بهینه‌شده برای نقشه‌برداری با مدل رگرسیون خطی ساده (ب)..... 12
شکل 3-2- تعیین موقعیت نقاط نمونه‌برداری بر اساس تعداد نقاط به‌صورت شبکه منظم 17
شکل 3-3- تعیین موقعیت نقاط نمونه‌برداری بر اساس فاصله به‌صورت شبکه منظم 18
شکل 3-4- نمونه‌برداری با شبکه منظم 1000متر *1000 متر..... 21
شکل 3-5- نقشه تعداد لایه‌ها..... 25
شکل 3-6- طراحی نمونه‌برداری با روش پوشش مکانی 26
شکل 3- 7- نقشه نمونه‌های پیشین 27
شکل 3- 8- نقشه جدید تعداد لایه‌های حاصل مجموع داده قدیمی و جدید (407+100) 29
شکل 3-9- نقاط جدید و نقاط قدیم به همراه لایه‌ها..... 29
شکل 3-10- نمونه KMEANS سخت 20 نقطه‌ای در دشت روانسر..... 37
شکل 3-11- نمونه‌های KMEANS سخت که در نمودار پراکندگی NDVI در برابر MRVBF رسم شده‌اند..... 37
شکل 3-12- نمودار مولفه‌های اصلی اول و دوم و وضعیت متغیرهای کمکی در فضای مشخصه آن‌ها..... 42
شکل 3-13- مکان‌های نمونه‌برداری (20 مکان) با استفاده از روش KMEANS فازی در دشت روانسر 47
شکل 3-14- نمونه‌های KMEANS فازی (نمودار پراکندگی NDVI در برابر MRVBF) 48
شکل 3- 15- 20 نمونه حاصل از استفاده از روش KMEANS فازی در دره هانتر 54
شکل 3-16- نمونه‌های KMEANS فازی (نمودار پراکندگی ارتفاع در برابر شاخص توپوگرافی)..... 55
شکل 3-17- مثالی از نمونه‌برداری ابر مکعب لاتین برای 2 متغیر با توزیع نرمال. 57
شکل 3-18- نمودار پراکندگی احتمال تجمعی ارتفاع و شیب. 59
شکل 3- 19- تعداد نقاط در لایه‌های مرزی 68
شکل 3-20- نمونه ابر مکعب لاتین مشروط 20 نقطه‌ای در دره هانتر 69
شکل 3-21- نمونه‌های ابر مکعب لاتین مشروط (نقاط کروی قرمز)..... 70
شکل 3-22- نمونه‌های حاصل از اجرای روش ابر مکعب لاتین با استفاده از نرم‌افزار

CLHS بر روی نقشه ارتفاع منطقه روانسر 74
شکل 3-23- نقاط نمونه‌برداری بر اساس روش کنارد استون بر روی نقشه ارتفاع منطقه مطالعه شده.....77

پیشگفتار

ترسیم طرح نمونه‌برداری خاک نخستین گام مهم در نقشه‌برداری رقومی خاک[1] است. انتخاب یک طرح نمونه‌برداری قوی که بتواند حتی‌المقدور بهترین نماینده‌ی ناحیه مطالعه شده باشد، تاثیر بسزایی در دقت پیش‌بینی در این نوع مطالعات دارد. در این دستورالعمل، مجموعه‌ای غنی از طرح‌های نمونه‌برداری ساده تا پیشرفته برای نقشه‌برداری ویژگی‌ها و کلاس‌های خاک در قالب مطالعات نقشه‌برداری رقومی خاک در دسترس است. این دستورالعمل برای محققین و کارشناسانی تهیه شده است که با نقشه‌برداری رقومی خاک و نرم‌افزار R آشنایی داشته باشند. تمام کدهای لازم برای اجرای طرح‌های نمونه‌برداری و بسته‌های نرم‌افزاری لازم در فایل پیوست دستورالعمل ارائه شده است. مقاله‌ی "Sampling for digital soil mapping: A tutorial supported by R scripts" مبنای تهیه این دستورالعمل قرار گرفته است (Brus, 2019).

[1] Digital soil mapping

1- مقدمه

مطالعات خاک‌شناسی، شامل نمونه‌برداری صحرایی، تجزیه‌های آزمایشگاهی، پردازش داده‌ها و نقشه‌برداری، با هدف طبقه‌بندی نوع خاک و اندازه‌گیری ویژگی‌های خاک در یک منطقه مشخص و تولید نقشه‌ی آن‌ها به عنوان محصولات نهایی است (McBratney *et al.*, 2000). در مطالعات خاک‌شناسی مرسوم، نقشه‌های خاک بر اساس درک نقشه‌بردار از عوامل تشکیل خاک و با ترسیم کیفی مرزهای خاک تهیه می‌شود (Jenny, 1941). این روش بیشتر تحت تأثیر قضاوت ذهنی و تجربه عملی نقشه‌برداران خاک قرار دارد (Clifford *et al.*, 2014). برای حل مسائل مختلف کشاورزی، از جمله مدیریت مکانی خاص، ارزیابی کیفیت خاک، نظارت بر منابع طبیعی و محیط‌زیست، خطر فرسایش خاک و انتقال املاح، به نقشه‌های دقیق‌تری از خاک برای نشان دادن تغییرپذیری مکانی ویژگی‌های خاک نیاز است (Brus and Noij, 2008). افزون براین، توسعه فن‌آوری سیستم‌های موقعیت‌یاب جهانی (GPS[1])، تکنیک‌های سنجش از دور و نزدیک، تجزیه و تحلیل مکانی و پیشرفت‌های محاسباتی در سیستم‌های اطلاعات جغرافیایی (GIS[2])، نقشه‌برداری رقومی خاک با وضوح بالا را به یک رویکرد قدرتمند برای پیش‌بینی ویژگی‌های پیوسته و کیفی خاک با اطمینان بالا نسبت به نقشه‌برداری سنتی خاک تبدیل نموده‌است (McBratney *et al.*, 2003). مفهوم کلی نقشه‌برداری رقومی خاک یک رابطه جامع ریاضی و آماری (مدل پیش‌بینی) بین ویژگی‌های خاک اندازه-گیری شده و متغیرهای محیطی با وضوح بالا و سهل الوصول مانند توپوگرافی، مدل رقومی ارتفاع (DEM[3])، القای الکترومغناطیسی، مشخصات طیفی و غیره است. وقتی متغیرهای محیطی کمکی در مکان‌های مشاهداتی خاک در دسترس باشد، ویژگی‌های خاک هدف در مکان‌های جدید از نظر کمی با مدل پیش‌بینی تخمین زده می‌شوند (Scull *et al.*, 2003).

یک بخش جدایی‌ناپذیر در فرایند نقشه‌برداری رقومی خاک، ایجاد و اجرای یک طرح نمونه‌برداری جامع است که برنامه‌ای را برای جمع‌آوری نمونه‌های معرفی که منطقه مطالعه شده را پوشش می‌دهند، داشته باشد و ورودی قابل اعتمادی برای ایجاد

[1] Global Positioning System (GPS)
[2] Geographic Information System (GIS)
[3] Digital Elevation Model (DEM)

یک مدل پیش‌بینی همراه با متغیرهای محیطی را فراهم نماید (Kidd *et al.*, 2015). طرح نمونه‌برداری از اهمیت زیادی برخوردار است، زیرا بر نتایج اندازه‌گیری آزمایشگاهی و تجزیه و تحلیل داده‌های بعدی تأثیر می‌گذارد (De Zorzi *et al.*, 2008). میزان کل خطای موجود در مطالعات می‌تواند به دو گروه کلی شامل خطای نمونه‌برداری و خطای آزمایشی تفکیک شوند که میزان آن به‌ترتیب حداقل 90 و حداکثر 10 درصد است (Lame and Defize, 1993). در بسیاری از موارد خطای ایجاد شده به دلیل طراحی نامناسب نمونه‌برداری به‌مراتب بسیار بیشتر از خطای ایجاد شده در هنگام آماده‌سازی نمونه، خطای دستگاه‌های اندازه‌گیری و آنالیز داده‌ها است (Markert, 2007). از این‌رو، لزوم انتخاب یک طرح نمونه‌برداری مناسب بیش از پیش احساس می‌شود.

حجم یا تعداد نمونه[1] یک بخش اصلی طرح نمونه‌برداری است که باید از پیش تعیین شود. در بیشتر موارد، تعداد نمونه با توجه به بودجه موجود برای کار میدانی و تجزیه‌های آزمایشگاهی تعیین می‌شود. از سوی دیگر، اگر دقت، عامل غالب باشد، تعداد نمونه باید از پیش تعیین شده و تغییرات مکانی را پوشش دهد که معمولاً در این حالت به نمونه‌های بیشتری نیاز است. از این‌رو، تعادل بین بودجه و دقت قابل قبول، تعداد نمونه مناسب را تعیین می‌کند (Brungard and Boettinger, 2010). Vaˇs´at و همکاران (2012) با مقایسه تأثیر تعداد نمونه بر پارامترهای تغییرنما، نتیجه گرفتند که 50 نمونه برای تهیه نقشه درون‌یابی قابل اعتماد در یک مزرعه کشاورزی 24 هکتاری کافی است (تقریباً 2 نمونه در هکتار). موقعیت مکانی نمونه یکی دیگر از اجزای ضروری طرح نمونه‌برداری است و توسط طرح‌های نمونه‌برداری مختلف تعیین می‌شود. طراحی نمونه‌برداری در واقع روش انتخاب موقعیت مکانی نمونه است (Brus and de Gruijter, 1997) که اهداف مختلفی، مانند جستجوی مکان‌های آلوده (Theocharopoulos *et al.*, 2001)، استنباط پارامترهای جمعیت مانند میانگین و واریانس جمعیت (de Gruijter and Ter Braak, 1990)یا تخمین تغییرنما (Lark, 2002) را دنبال می‌کند. هدف اصلی طرح نمونه‌برداری در نقشه‌برداری رقومی خاک تهیه داده‌های ورودی قابل اعتماد برای مدل‌های پیش‌بینی است.

در نمونه‌برداری برای بررسی خاک می‌توان دو رویکرد کاملاً متفاوت را دنبال کرد: یکی

[1] Sample size

رویکرد مبتنی بر طرح[1] و دیگری مبتنی بر مدل[2] (Brus and de Gruijter, 1997). رویکرد نمونه‌برداری مبتنی بر طرح بر نظریه احتمال استوار است و در روش نمونه‌برداری مبتنی بر مدل از تجزیه و تحلیل زمین آماری استفاده می‌شود. با این حال، خروجی این دو روش بیشتر برای تخمین برخی از پارامترهای آماری استفاده می‌شود. در روش مبتنی بر طرح، مکان‌های نمونه‌برداری با نمونه‌برداری احتمال انتخاب می‌شوند و استنباط آماری (به عنوان مثال تخمین میانگین مکانی) بر اساس طرح نمونه‌برداری است. در روش نمونه‌برداری مبتنی بر مدل، هیچ الزامی برای روش انتخاب مکان‌های نمونه‌برداری وجود ندارد و معمولاً با استفاده از نمونه‌برداری هدفمند، به‌عنوان مثال در یک شبکه متمرکز، انتخاب نمونه صورت می‌گیرد. در استنباط آماری، یک مدل مانند کریجینگ معمولی، با فرض میانگین ثابت (ناشناخته)، یا کریجینگ عام (که در آن میانگین به عنوان یک تابع خطی از یک یا چند پیش‌بینی‌کننده در نظر گرفته می‌شود) برای پیش‌بینی تغییرات مکانی معرفی می‌شود. منبع تصادفی بودن در دو رویکرد متفاوت است. در روش مبتنی برطرح، انتخاب مکان‌های نمونه‌برداری تصادفی است، در حالی‌که در روش مبتنی بر مدل، مدل تغییرات مکانی، تصادفی بودن را معین می‌کند. در این حالت می‌توان معیارهای عدم قطعیت مانند واریانس خطای برآورد را نیز مشخص کرد.

انتخاب بین این دو رویکرد نمونه‌برداری یکی از مهم‌ترین تصمیمات در طراحی طرح‌های نمونه‌برداری است (de Gruijter et al., 1990). Brus و de Gruijter (1997) درباره جوانب مثبت و منفی هر دو رویکرد بحث کردند و قواعدی برای انتخاب میان این دو روش بیان کردند. به‌طورکلی، اگر هدف تعیین پارامترهای تابع توزیع تجمعی مکانی[3](SCDF) برای کل منطقه مطالعه شده یا تعداد محدودی از زیرناحیه‌های آن باشد و افزون‌براین، اعتبار نتیجه مهمتر از کارایی باشد، رویکرد مبتنی بر طرح بهترین گزینه است. اگر علاقه به تولید نقشه‌ای باشد که مقادیر را در مناطق بسیار کوچک نشان دهد، به عنوان مثال مقادیر در هر اندازه پیکسل و هدف پیش‌بینی دقیق مقادیر باشد؛ یعنی کارایی مدل مهم‌تر از مقدار اعتبارسنجی مدل (شاخص‌هایی مانند R^2 یا

[1] Design-basedsampling

[2] Model-basedsampling

[3] Spatial cumulative distribution function

RMSE) باشد، رویکرد مبتنی بر مدل، بهترین طرح انتخابی برای نمونه‌برداری خواهد بود.

بیشتر طرح‌های نمونه‌برداری در نقشه‌برداری رقومی خاک به گونه‌ای طراحی شده‌اند که بتوانند پوشش مکانی خوبی برای یک ناحیه یا پوشش خوبی از تغییرات ویژگی خاک مطالعه شده در آن ناحیه را فراهم نمایند. از این‌رو، به منظور پیش‌بینی مکانی با استفاده از متغیرهای محیطی، دو طرح دیگر نمونه‌برداری یعنی نمونه‌برداری در فضای جغرافیایی یا نمونه‌برداری در فضای ویژگی‌ها مناسب‌ترند (Minasny and McBratney, 2006).

طرح نمونه‌برداری را می‌توان در فضای جغرافیایی، فضای ویژگی یا هر دو بهینه کرد (Hengl *et al.*, 2003). برخی مطالعات استدلال می‌کنند که برای اهداف واسنجی، ضروری نیست پوشش مکانی مناسب در تقابل با پوشش مناسب متغیرهای محیطی باشد (Minasny and McBratney, 2006). با این وجود، بسیاری از مطالعات، پراکنش مناسب در فضای مشخصه و فضای جغرافیایی را توصیه می‌کنند. Hengl و همکاران (2003) بیان داشتند که یک طرح نمونه‌برداری بهینه باید هم‌زمان نشان‌دهنده‌ی تغییرات ویژگی‌های خاک در فضای مشخصه و جغرافیایی باشد. آن‌ها یک روش لایه‌بندی با دامنه برابر را پیشنهاد کردند که در آن دامنه متغیر پیش‌بینی‌کننده به خوشه‌هایی با عرض مساوی تقسیم شود و سپس نمونه‌ها به‌شکل تصادفی در هر خوشه با توجه به وزن‌های داده شده انتخاب می‌شوند. Brus و Heuvelink (2007) همچنین نشان دادند که گسترش خوب نمونه‌ها در فضای مشخصه، تخمین دقیق ضرایب رگرسیون را تضمین می‌کند، در حالی که میان‌یابی قابل اعتماد داده‌های نمونه بستگی زیادی به پراکندگی خوب در فضای جغرافیایی دارد. Walvoort و همکاران (2010) دریافتند که برآورد میانگین‌های مکانی متغیرهای محیطی، زمانی که مکان‌های نمونه به طور مساوی در فضای جغرافیایی گسترش می‌یابند، دقیق‌تر می‌شود. از این‌رو، انتخاب طرح‌های نمونه‌برداری در نقشه‌برداری رقومی خاک به هدف مطالعه و مزایا و معایب اساسی طرح نمونه‌برداری بستگی دارد.

برای انتخاب طرح نمونه‌برداری مناسب برای رسیدن به هدف مطالعه باید دقت شود. باید اطلاعات جامعی از طرح‌های نمونه‌برداری موجود و معمولاً استفاده شده در نقشه‌برداری رقومی خاک در اختیار محققان، مشاوران، نقشه‌برداران و برنامه‌ریزان برای

تصمیم‌گیری آگاهانه قرار گیرد. از این‌رو، هدف اصلی این دستورالعمل فنی ارائه شرح مفصلی از اصول، مزایا و کاستی‌های چندین طرح نمونه‌برداری متداول و پرکاربرد در نقشه‌برداری رقومی خاک است.

ترسیمی کلی از طرح‌های نمونه‌برداری برای نقشه‌برداری رقومی خاک که توسط de Gruijter و همکاران (2006) پیشنهاد شده که در دستورالعمل کنونی ارائه شده است. در جلد اول این دستوالعمل، طرح‌های نمونه‌برداری غیرمحتمل برای نقشه‌برداری شامل نمونه‌برداری منظم شبکه‌ای، نمونه‌برداری با پوشش مکانی، نمونه‌برداری با استفاده از میانگین‌های کا و کا فازی، نمونه‌برداری به روش ابر مکعب لاتین مشروط و نمونه‌برداری به روش کنارد-استون بررسی شده و در محیط نرم‌افزار R به طور کاربردی اجرای این روش‌ها ارائه شده است. در جلد دوم این مجموعه، نمونه‌برداری مبتنی بر مدل شرح داده می‌شود و از مدل اولیه تغییرات مکانی متغیر مورد نظر خاک برای بهینه‌سازی حجم (اندازه) نمونه و یا مختصات مکانی موقعیت‌های نمونه‌برداری استفاده می‌شود. طرح‌های نمونه‌برداری برای برآورد تغییرنما، نمونه‌برداری تودرتو، نمونه‌برداری تصادفی مستقل از جفت نقاط، طرح‌های نمونه‌برداری مبتنی بر مدل هستند که در آن واریانس کریجینگ به کمترین می‌رسد. برای اعتبارسنجی نقشه و روش نمونه‌برداری احتمالی اضافی در این دستورالعمل اطلاعات لازم کاربردی بیان شده است. از این‌رو، می‌توان برآورد بی‌طرفانه شاخص‌های کیفیت نقشه و خطاهای استاندارد آن‌ها را بدست آورد. برای تمام طرح‌های نمونه‌برداری، کدهای قابل اجرا در R، در پیوست دستورالعمل موجود است.

2- نصب و راه‌اندازی نرم‌افزارهای لازم

این دستورالعمل فنی بر نصب نرم‌افزار منبع آزاد R و بسته‌های نرم‌افزاری که در این محیط برنامه‌نویسی نصب و راه‌اندازی می‌شوند، تکیه دارد.

2-1- استفاده از R، Rstudio و بسته‌های قابل الحاق به محیط نرم‌افزار R

R یک زبان و محیط برای محاسبات آماری است و طیف گسترده‌ای از روش‌های آماری (به عنوان مثال مدل‌سازی خطی، آزمون‌های آماری، سری‌های زمانی، طبقه‌بندی،

خوشه‌بندی و...)، زمین آماری (کریجینگ معمولی، کوکریجینگ، رگرسیون کریجینگ و ...) و روش‌های گرافیکی را در بردارد و بسیار قابل گسترش است.

2-1-1- بارگیری و نصب نرم‌افزار R

بارگیری نرم‌افزار R، پرونده‌ها و دستورالعمل‌های نصب را می‌توان از شبکه آرشیو جامع (CRAN) بارگیری کرد:

1) برای بارگیری و نصب نرم‌افزار R به آدرس: https://cloud.r-project.org index.html مراجعه کنید.

2) متناسب با سیستم عامل استفاده شده (ویندوز، لینوکس یا مک او اس)، نسخه نرم‌افزار انتخاب و بارگیری می‌شود.

2-1-2- بارگیری و نصب نرم‌افزار RStudio

شروع استفاده از R برای افراد مبتدی بسیار دشوار است زیرا بدون رابط کاربری گرافیکی (GUI) است. برخی از رابط‌های کاربری گرافیکی وجود دارد مانند RStudio وجود دارد که استفاده از R را آسان‌تر می‌کند. این رابط گرافیکی شامل پنجره ویرایشگر کد، اشکال‌زدایی و پنجره نمایش گرافیکی است.

برای نصب RStudio باید مراحل زیر اجرا شود:

1) بارگیری و نصب نسخه منبع باز RStudio از آدرس اینترنتی https://www.rstudio.com/products/rstudio/download/.

2) انتخاب گزینه RStudio Desktop و Open Source License

3) متناسب با سیستم عامل استفاده شده (ویندوز، لینوکس یا مک او اس)، نسخه نرم‌افزار انتخاب و بارگیری می‌شود.

2-2- نصب نرم‌افزار جاوا و Rtools

امکان توسعه‌ی قابلیت‌های R، با افزودن بسته‌های نرم‌افزاری ایجاد شده توسط کاربران آن، یکی از ویژگی‌های مهم این نرم‌افزار است. این بسته‌های نرم‌افزاری به زبان‌های برنامه

نویسی R، LaTeX، جاوا، سی++ و فورترن نوشته شده‌اند. در محیط R، کدهای ویژوال بیسیک، سی، سی++ و فورترن قابلیت اتصال و فراخوانی هنگام اجرای برنامه را دارند و به همین منظور با نصب جاوا بر روی سیستم خود می‌توانید نرم‌افزارهایی مانند spcosa و یا برنامه‌هایی که بر پایه زبان برنامه‌نویسی جاوا نوشته شده‌اند را اجرا کنید.

2-2-1- نصب نرم‌افزار جاوا

بعد از بارگیری نرم‌افزار جاوا نسخه 15 jre-7u17-windows-i586 از سایت https://www.java.com/en، نرم‌افزار بر روی رایانه نصب می‌شود.

2-2-2- نصب نرم‌افزار Rtools

Rtools: مجموعه‌ای نرم‌افزاری برای ساخت بسته‌هایی برای R در ویندوز یا ایجاد R به تنهایی (نسخه 1.9.0 یا بالاتر) است.

1) Rtools از آدرس اینترنتی http://cran.stat.sfu.ca/bin/windows/Rtools دردسترس بوده و به صورت رایگان بارگیری می‌شود. در این دستورالعمل چون از نسخه R3.6 برای اجرای کدها استفاده شده است از این‌رو لازم است نسخه Rtools35.exe بارگیری شود.

2) پس از بارگیری کامل، می‌توان Rtools را در مسیر مورد نظر نصب نمود.

2-3- نصب بسته‌های نرم‌افزاری لازم

هنگام بارگیری R، تنها سیستم پایه‌ی R دریافت می‌شود که زبان R را پیاده‌سازی می‌کند. R با مجموعه بزرگی از بسته‌های نرم‌افزاری قابل الحاق به آن که عملکرد اساسی آن را گسترش می‌دهند، مفیدتر می‌شود. بسته‌های R توسط جامعه R ساخته می‌شوند.

منبع اصلی بسته‌های نرم‌افزاری R وب سایت رسمی CRAN[1] بوده که در حال حاضر حدود 17479 بسته در آن فهرست شده است. برای پردازش اطلاعات مکانی، بسته‌های نرم‌افزاری مختلفی موجود است. با تابع available.packages() و اجرای ان در

[1] https://cran.r-project.org

محیط R می‌توانید اطلاعات مربوط به بسته‌های نرم‌افزاری موجود را مستقیماً در CRAN بدست آورید.

2-3-1- پرکاربردترین بسته‌های نرم‌افزاری R برای اجرای طرح‌های نمونه‌برداری در نقشه‌برداری رقومی خاک

همانطور که پیش از این یادآوری شد، R را می‌توان با بسته‌های نرم‌افزاری قابل الحاق توسعه داد. این بسته‌های نرم‌افزاری مجموعه‌ای از توابع به زبان R، داده‌ها، مستندات و کدهای گردآوری شده‌است که به راحتی با دیگران به اشتراک گذاشته می‌شود. در بخش‌های زیر، پرکاربردترین بسته‌های نرم‌افزاری مربوط به نقشه‌برداری رقومی خاک ارائه می‌شود. هر کدام یا مجموعه‌ای از این بسته‌های نرم‌افزاری با استفاده از تابع (install.packages) در محیط نرم‌افزاری R نصب می‌شود. به عنوان مثال:

```
install.packages(c("sp", "gstat", "raster","rgdal"))
```

در زیر به تعدادی از این بسته‌های نرم‌افزاری که در این دستورالعمل استفاده می‌شود کوتاه اشاره شده است:

raster: توابعی در ارتباط با خواندن، نوشتن، دستکاری، تجزیه و تحلیل و مدل‌سازی داده‌های مکانی به فرمت رستری و وکتوری در خود دارد. این بسته توابع اساسی و سطح بالا را اجرا نموده و از پردازش فایل‌های بسیار بزرگ پشتیبانی می‌شود.

rgdal: پیوندهایی را برای آرشیو[1](GDAL) فراهم می‌کند.

Sp: این بسته نرم‌افزاری انواع داده‌های مکانی و روش‌های کار بر روی آن‌ها را ارائه می‌دهد.

caret: طیف گسترده‌ای از توابع برای آموزش و اعتبارسنجی و ترسیم مدل‌های طبقه‌بندی و رگرسیون در بردارد.

gstat: این بسته شامل توابع سودمندی برای مدل‌سازی و ترسیم تغییرنما و نقشه کریجینگ (کوکریجینگ) ساده، معمولی و عام نقطه‌ای یا بلوکی است.

spsann: در این بسته نرم‌افزاری روش‌هایی برای بهینه‌سازی پیکربندی نمونه‌ها با استفاده

[1] Geospatial Data Abstraction Library (GDAL)

از بازپخت مکانی شبیه‌سازی شده[1] وجود دارد. همچنین شامل توابع چندمنظوره برای اهداف مختلف مانند برآورد تغییرنما، تخمین روند مکانی و میان‌یابی مکانی است.

cluster: روش‌های تجزیه و تحلیل خوشه‌های داده را در بردارد.

fclust : این بسته نرم‌افزاری شامل الگوریتم‌های خوشه‌بندی فازی، توابعی برای تولید شاخص‌های اعتبار خوشه‌ای و نمودارهایی برای نمایش اعتبار خوشه‌ای و تجسم نتایج خوشه بندی فازی است.

ggplot2: این بسته نرم‌افزاری افزون‌بر رسم گراف‌های پایه، قابلیت ترسیم نمودارهای مفید را دارد.

e1071: این بسته نرم‌افزاری روش‌های تجزیه و تحلیل و خوشه‌بندی فازی را در بردارد.

prospectr: این نرم‌افزار دارای توابعی برای پیش پردازش داده‌های طیف سنجی و نمونه‌برداری انتخابی/واسنجی (نماینده) و روش نمونه‌برداری کنارد استون است.

2-4- دستورالعمل‌های موجود برای شروع کار با R

1) کتابچه‌های راهنمای R: http://cran.r-project.org/manuals.html
2) مستندات ارائه شده: http://cran.r-project.org/other-docs.html
3) Quick-R: http://www.statmethods.net/index.html
4) Stackoverflow R community: https://stackoverflow .com/questions /tagged/r

3- اجرای روش‌های نمونه‌برداری

براساس داده‌های سه منطقه مطالعه شده در داخل و خارج کشور، روش‌های نمونه‌برداری مبتنی بر طرح، شرح و گسترش داده شده‌اند. این داده‌ها عبارتند از:

- **داده‌های دره هانتر، نیو ساوت ولز در استرالیا**

داده‌های منطقه مطالعاتی دره هانتر، شامل نقشه‌های رستری پنج متغیر کمی ارتفاع، شیب، جهت شیب، شاخص توپوگرافی مرکب و شاخص پوشش گیاهی با اختلاف

[1] Spatial Simulated Annealing

نرمال است. از این داده برای نشان دادن نمونه‌برداری به روش میانگین‌های کا[1]و نمونه‌برداری به روش ابر مکعب لاتین مشروط[2] استفاده شده است.

- **داده‌های مناطق آلفا، چیلگا و دمبیا در اتیوپی**

داده‌های غلظت مواد آلی خاک (SOM) در افق A، از سه منطقه آلفا، چیلگا و دمبیا کشور اتیوپی جمع‌آوری شده است و بیشتر این داده‌ها، حاصل نمونه‌برداری خاک از حاشیه جاده‌ها هستند. متغیرهای کمکی برای این منطقه مطالعه شده شامل نقشه‌های رستری باند مادون قرمز نزدیک (NIR)، باند مادون قرمز مریی، دمای سطح زمین، شاخص پوشش گیاهی افزایش یافته و ارتفاع است. این مجموعه داده برای نشان دادن روش نمونه‌برداری پرنمودن مکانی[3]، بهینه‌سازی مبتنی بر مدل فاصله از یک شبکه مربع و نمونه‌برداری پرنمودن مکانی مبتنی بر مدل استفاده شده است.

- **داده‌های دشت روانسر استان کرمانشاه در ایران**

این مجموعه داده شامل نقشه‌های رستری ارتفاع، شیب، شاخص همواری دره با درجه‌ی تفکیک بالا (MRVBF[4])، شاخص خیسی توپوگرافی و شاخص اندازه ذرات مربوط به دشت روانسر واقع در استان کرمانشاه است. از این داده‌ها برای اجرای مثال نمونه‌برداری شبکه منظم، روش نمونه ابرمکعب لاتین و کنارد-استون استفاده شده است.

3-1- نمونه‌برداری محتمل در مقابل نمونه‌برداری غیر محتمل

در نمونه‌برداری تصادفی، با استفاده از مولد اعداد تصادفی، زیر مجموعه‌ای از واحدهای جمعیت به طور تصادفی از بین جمعیت انتخاب می‌شوند. نمونه‌هایی حاصل از نمونه‌برداری غیرتصادفی، نمونه‌های راحت به آن‌ها اطلاق می‌شود. در این خصوص می‌توان به نمونه‌برداری در امتداد جاده‌ها، نمونه‌برداری دلخواه یعنی نمونه‌برداری بدون

[1] K-means sampling
[2] Conditioned Latin hypercube sampling
[3] Spatial infill sampling
[4] Multi-resolution valley bottom flatness index (MRVBF)

هدف خاص در ذهن و نمونه‌برداری هدفمند اشاره نمود.

در منابع علمی، اصطلاح نمونه‌برداری تصادفی غالباً برای نمونه‌برداری اختیاری[1] استفاده شده‌است. برای جلوگیری از سردرگمی اصطلاح نمونه‌برداری محتمل[2] معرفی می‌شود. نمونه‌برداری محتمل، در صورتی نمونه‌برداری تصادفی تلقی می‌شود که دو شرط زیر را برآورده کند. اولاً اول، همه واحدهای جمعیت یک احتمال مثبت برای انتخاب شدن دارند و هیچ بخشی از جمعیت از این قاعده مستثنی نیست. دوم، احتمال انتخاب هر نمونه ممکن قابل محاسبه است. درنمونه‌برداری اختیاری این دو شرط اغلب بدست نمی‌آید.

در انتخاب نمونه، بین نمونه‌برداری محتمل یا غیرمحتمل با انتخاب نمونه بین روش مبتنی بر مدل یا مبتنی بر طرح برای استنباط آماری (برآورد،آزمایش فرضیه) ارتباط نزدیکی وجود دارد (de Gruijter and ter Braak, 1990; Papritz and Webster, 1995; Brus and de Gruijter, 1997) در روش مبتنی بر طرح، واحدها (نمونه‌ها) از طریق نمونه‌برداری محتمل انتخاب می‌شوند. برآوردها براساس احتمال انتخاب واحدهای نمونه‌برداری است که توسط طرح نمونه‌برداری تعیین می‌شود (استنباط مبتنی بر طرح). برعکس، در روش‌های مبتنی بر مدل، از یک مدل تصادفی، به عنوان مثال مدل رگرسیون خطی یا مدل کریجینگ معمولی برای برآورد استفاده می‌شود. به مثال زیر توجه کنید:

ما یک مدل خطی به شرح زیر داریم:

$$z_i = \beta_0 + \beta_i x_i + \epsilon_i \qquad (1)$$

در این معادله zi متغیر مورد نظر از واحدi، x_i یک متغیر کمکی از آن واحد،β0 و $\beta 1$ ضرایب رگرسیون وϵi خطای (باقیمانده) در واحدi، دارای توزیع نرمال با میانگین صفر و انحراف معیارثابت σ است. خطاها مستقل هستند به‌طوری که

$\mathrm{Cov}(\epsilon i,\epsilon j)=0 \text{for all } i\neq j.$

شکل 1 (الف و ب) یک نمونه تصادفی ساده و بدون جایگزینی[3](SRS) و نمونه بهینه شده برای کالیبراسیون مدل رگرسیون خطی ساده را نشان می‌دهد. هر دو نمونه بر روی

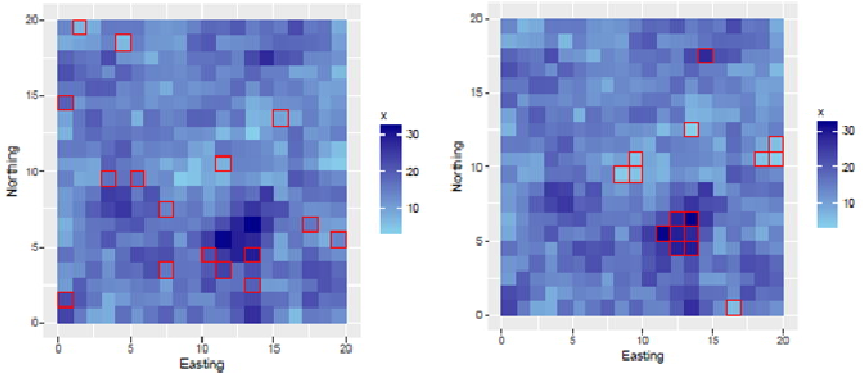
[1] Arbitrary sampling
[2] Probability sampling
[3] Simple randomsample without replacement

نقشه متغیر کمکی (پیش‌بینی‌کننده) ترسیم شده‌اند. خطاهای استاندارد هر دو ضریب رگرسیون (محاسبه‌شده برای انحراف استاندارد باقی مانده σ از 2) به‌طور قابل توجهی بزرگ‌تر از نمونه بهینه‌سازی شده است (جدول 3-1).

جدول 3-1- خطاهای استاندارد برآورد شده محل تقاطع (β0) و شیب (β1) برای یک نمونه تصادفی ساده (SRS) و نمونه بهینه شده برای رگرسیون خطی ساده (شکل 1 را ببینید).

β1	β0	
0/086	1/51	*SRS*
0/051	1/08	Optimized sample

عدم قطعیت مشترک درباره دو ضریب رگرسیون، توسط دترمینان ماتریس واریانس-کوواریانس ضرایب رگرسیون برآورد شده تعیین می‌شود که برابر 0/0002 برای SRS و 0/0001 برای نمونه بهینه شده است. از این‌رو، برای تهیه نقشه رقومی خاک با یک مدل رگرسیون خطی ساده، نمونه‌برداری تصادفی ساده گزینه خوبی نیست.



الف ب

شکل 3-1- از چپ به راست: نمونه تصادفی ساده (الف)، نمونه بهینه‌شده برای نقشه‌برداری با مدل رگرسیون خطی ساده (ب)

3-2- طرح‌های نمونه‌برداری هندسی

3-2-1- نمونه‌برداری از شبکه منظم

یک روش نمونه‌برداری ساده و عامه‌پسند در نقشه‌برداری رقومی خاک، نمونه‌برداری در یک شبکه منظم است. نمونه‌برداری شبکه‌ای، روشی است که در فواصل مساوی سطح زمین، نمونه‌های کوچک گرفته می‌شود. یک شبکه منظم می‌تواند مربع، مثلث یا چندضلعی باشد. هنگام نمونه‌برداری از یک شبکه منظم، باید درباره فواصل شبکه، یعنی فاصله بین نقاط همسایه، تصمیم‌گیری شود. بودجه موجود یا لزوم تهیه نقشه با کیفیت، عوامل اصلی تصمیم‌گیری درباره تعیین فاصله نقاط نمونه‌برداری هستند.

در ارتباط با بودجه موجود و تخمین هزینه‌های مرتبط با هر نقطه نمونه‌برداری، نخست حجم نمونه مناسب محاسبه می‌شود. سپس، برای یک شبکه مربع می‌توان فاصله شبکه را از رابطه $d = \sqrt{\frac{A}{n}}$ محاسبه کرد که در آن A مساحت منطقه و n حجم نمونه‌ی مقرون به صرفه است. در این رابطه واحد مساحت A بر حسب متر مربع، فاصله شبکه d بر حسب متر است. شبکه‌های مربعی را می‌توان با تابع (spsample) در بسته نرم‌افزاری sp در محیط R انتخاب کرد (Pebesma and Bivand, 2005). برای اجرای عملی این روش نمونه‌برداری به فایل Regular grid sampling.R در پیوست این دستورالعمل مراجعه شود.

3-2-1-1- نمونه‌برداری شبکه منظم بر روی نقشه‌های رستری بر اساس تعداد نقاط و فاصله شبکه

بارگیری بسته‌های نرم‌افزار R

```
library(sp)
```

برای ورود داده‌ها به محیط R، داده‌های اتیوپی را بارگیری کنید

```
load("yourdirectory/CovariatesThreeWoredasEthiopia.RData")
str (grdEthiopia)
```

```
Formal class 'SpatialPixelsDataFrame' [package "sp"] with 7 slots
 @.. data       :'data.frame':        ۱۰۸۴۳obs. of  5 variables:
 $.. .. dem    : num [1:10843] 1.08 1.08 1.1 1.1 1.1...
 $.. .. evi    : num [1:10843] 0.299 0.284 0.294 0.27 0.269...
 $.. .. rfl_NIR: num [1:10843] 0.246 0.235 0.253 0.232 0.227...
 $.. .. rfl_red: num [1:10843] 0.071 0.0689 0.0709 0.0716 0.0723...
 $.. .. lst    : num [1:10843] 310 310 311 311 310...
 @.. coords.nrs : num (۰)
 @.. grid       :Formal class 'GridTopology' [package "sp"] with 3 slots
 @.. .. .. cellcentre.offset: Named num [1:2] 1736 739
 -.. .. .. .. attr(*, "names")= chr [1:2] "s1" "s2"
 @.. .. .. cellsize         : Named num [1:2] 1 1
 -.. .. .. .. attr(*, "names")= chr [1:2] "s1" "s2"
 @.. .. .. cells.dim        : Named int [1:2] 162 168
 -.. .. .. .. attr(*, "names")= chr [1:2] "s1" "s2"
 @.. grid.index : int [1:10843] 27108 27109 26945 26946 26947 26948
26949 26782 2678326784...
 @.. coords     : num [1:10843, 1:2] 1789 1790 1788 1789 1790...
 -.. .. attr(*, "dimnames")=List of 2
 : $.. .. .. chr [1:10843] "1" "2" "3" "4... "
 : $.. .. .. chr [1:2] "s1" "s2"
 @.. bbox       : num [1:2, 1:2] 1736 739 1898 907
 -.. .. attr(*, "dimnames")=List of 2
 : $.. .. .. chr [1:2] "s1" "s2"
 : $.. .. .. chr [1:2] "min" "max"
 @.. proj4string:Formal class 'CRS' [package "sp"] with 1 slot
 @.. .. .. projargs: chr NA
```

با استفاده از تابع (coordinates) کلاس grdEthiopia به فرمت SpatialPointsDataFrame تغییر می‌یابد. سپس با استفاده از تابع (gridded) آن را به SpatialPixelsDataFrame تغییر دهید.

```
coordinates(grdEthiopia)<-~s1+s2
gridded(grdEthiopia)<- TRUE
str(grdEthiopia)
```

```
> str(grdEthiopia)
Formal class 'SpatialPixelsDataFrame' [package "sp"] with 7 slots
  ..@ data       :'data.frame':       10843 obs. of  5 variables:
  .. ..$ dem    : num [1:10843] 1.08 1.08 1.1 1.1 1.1 ...
  .. ..$ evi    : num [1:10843] 0.299 0.284 0.294 0.27 0.269 ...
  .. ..$ rfl_NIR: num [1:10843] 0.246 0.235 0.253 0.232 0.227 ...
  .. ..$ rfl_red: num [1:10843] 0.071 0.0689 0.0709 0.0716 0.0723 ...
  .. ..$ lst    : num [1:10843] 310 310 311 311 310 ...
  ..@ coords.nrs : num(0)
  ..@ grid       :Formal class 'GridTopology' [package "sp"] with 3 slots
  .. .. ..@ cellcentre.offset: Named num [1:2] 1736 739
  .. .. .. .. ..- attr(*, "names")= chr [1:2] "s1" "s2"
  .. .. ..@ cellsize         : Named num [1:2] 1 1
  .. .. .. .. ..- attr(*, "names")= chr [1:2] "s1" "s2"
  .. .. ..@ cells.dim        : Named int [1:2] 162 168
  .. .. .. .. ..- attr(*, "names")= chr [1:2] "s1" "s2"
  ..@ grid.index : int [1:10843] 27108 27109 26945 26946 26947 26948
26949 26782 26783 26784 ...
  ..@ coords     : num [1:10843, 1:2] 1789 1790 1788 1789 1790 ...
  .. ..- attr(*, "dimnames")=List of 2
  .. .. ..$ : chr [1:10843] "1" "2" "3" "4" ...
  .. .. ..$ : chr [1:2] "s1" "s2"
  ..@ bbox       : num [1:2, 1:2] 1736 739 1898 907
  .. ..- attr(*, "dimnames")=List of 2
  .. .. ..$ : chr [1:2] "s1" "s2"
  .. .. ..$ : chr [1:2] "min" "max"
  ..@ proj4string:Formal class 'CRS' [package "sp"] with 1 slot
  .. .. ..@ projargs: chr NA
```

1-1-1-2-3- تعیین موقعیت نقاط نمونه‌برداری با توجه به حجم نمونه

برای تعیین موقعیت نقاط نمونه‌برداری از تابع (spsample) استفاده می‌شود:

spsample (x, n, type, offset…)

این تابع موقعیت مکانی نمونه‌ها را در نواحی به شکل مربع و چند ضلعی، در یک شبکه یا در روی خط مکانی، با استفاده از روش نمونه‌گیری منظم یا تصادفی تعیین می‌کند. لازم به توضیح است که مختصات جغرافیایی باید در سیستم تصویر UTM باشد.

در این تابع:

x= داده‌ها به فرمت مکانی SpatialPolygons و SpatialGrid

n=تعداد یا حجم نمونه‌ها و type= نوع روش نمونه‌برداری را نشان می‌دهد. کاراکترهای "random" برای کاملاً تصادفی مکانی،"regular" برای نمونه‌برداری منظم (با تراز سیستماتیک) ؛"stratified" برای طبقه‌بندی تصادفی؛ "nonaligned " برای نمونه‌برداری سیستماتیک غیرمستقیم،"hexagonal" برای نمونه‌برداری از یک شبکه شش ضلعی؛"clustered"برای نمونه‌برداری خوشه‌ای استفاده می‌شوند.

offset= این شناسه برای انواع نمونه‌برداری مبتنی بر سلول مربعی (منظم، طبقه‌ای، مستقل[1]، شش ضلعی) استفاده می‌شود.

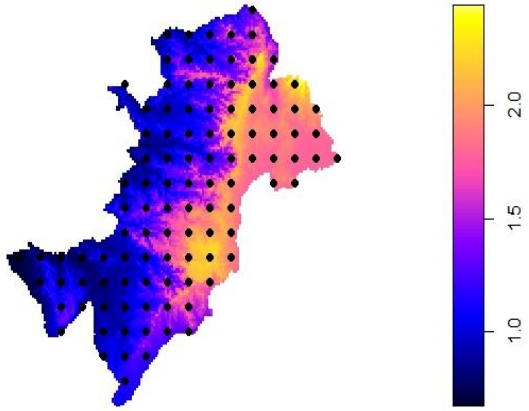
```
n<-101
mysample<-spsample(x=grdEthiopia,n=n,type="regular)
mysample <- as(mysample,"data.frame")
str(mysample)
```

```
'data.frame':100 obs. of  2 variables:
 $ x1: num  1792 1792 1803 1782 1792 ...
 $ x2: num  744 754 754 765 765 ...
```

ترسیم نقشه شبکه نقاط بر روی نقشه منطقه (شکل 3-2)

```
plot(grdEthiopia)
points(mysample, pch = 19)
```

[1] Nonaligned



شکل 3-2- تعیین موقعیت نقاط نمونه‌برداری بر اساس تعداد نقاط به‌صورت شبکه منظم

3-2-1-1-2- تعیین موقعیت نقاط بر اساس فاصله

اگر فاصله بین نقاط یا اندازه سلول یا همان اندازه پیکسل، 10/2 متر باشد. با استفاده از تابع (spsample) تعداد نمونه با استفاده از دستورات زیر محاسبه می‌شود:

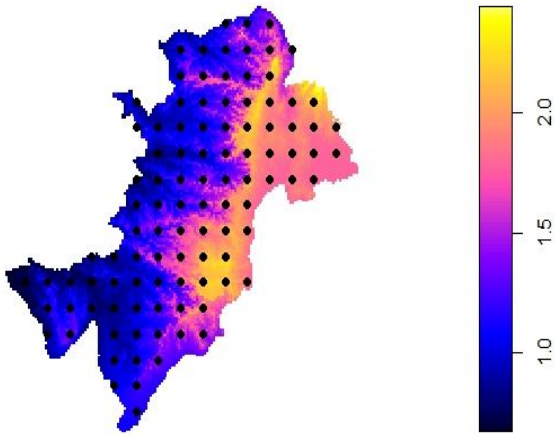
```
spacing <- 10.2
mysample2<-spsample(x=grdEthiopia,cellsize=spacing,type="regular")
mysample2 <- as(mysample2,"data.frame")
#compute sample size
nrow(mysample2)
```

تعداد نمونه برابر با 101 نقطه است

```
[1] 101
```

ترسیم نقشه شبکه نقاط بر روی نقشه منطقه

```
plot(grdEthiopia)
points(mysample2, pch = 19)
```



شکل 3-3- تعیین موقعیت نقاط نمونه‌برداری بر اساس فاصله به‌صورت شبکه منظم

3-2-1-2- نمونه‌برداری شبکه منظم بر روی نقشه‌ی چهارگوش

در این بخش، ناحیه مطالعه شده بخشی از اراضی حوضه آبریز قره‌سو در استان کرمانشاه است. نخست بسته‌های نرم‌افزاری لازم با تابع (library) فراخوانی می‌شوند:

```
library(sp)
library(rgdal)
library(raster)
```

```
> library(sp)
Warning messages:
1: package 'spcosa' was built under R version 3.5.0
2: package 'rJava' was built under R version 3.3.3
3: In fun(libname, pkgname) : bytecode version mismatch; using eval
> library(rgdal)
rgdal: version: 0.8-11, (SVN revision 479M)
Geospatial Data Abstraction Library extensions to R successfully loaded
Loaded GDAL runtime: GDAL 1.9.2, released 2012/10/08
Path to GDAL shared files: C:/Users/ECC/Documents/R/win-
library/3.3/rgdal/gdal
```

```
GDAL does not use iconv for recoding strings.
Loaded PROJ.4 runtime: Rel. 4.7.1, 23 September 2009, [PJ_VERSION: 470]
Path to PROJ.4 shared files: C:/Users/ECC/Documents/R/win-library/3.3/rgdal/proj
> library(raster)
Warning message:
no function found corresponding to methods exports from ‘raster’ for: ‘overlay’
```

فراخوان فایل نقشه پلی-گونی محدوده مطالعه شده با استفاده از تابع (shapefile) از بسته نرم‌افزاری raster

```
boundary <- shapefile('test/boundary.shp')
crs(boundary)
```

```
> boundary <- shapefile('test/boundary.shp')
Warning message:
In readOGR(dirname(filename), fn, stringsAsFactors = stringsAsFactors, :
 Z-dimension discarded
> crs(boundary)
[1] "+proj=utm +zone=38 +datum=WGS84 +units=m +no_defs +ellps=WGS84 +towgs84=0,0,0"
```

محاسبه مساحت محدوده مطالعه شده با استفاده از تابع (area).

```
area(boundary)
```

```
[1] 578536730
```

```
boundary$area_m2 <- area(boundary)
```

اگر هدف این باشد که شبکه منظمی از نقاط با فاصله 1000 متر ایجاد شود بر اساس فرمول فوق 578 نقطه نیاز خواهد بود:

$$d=\sqrt{\frac{A}{n}}\rightarrow 1000=\sqrt{\frac{578536730}{n}}\rightarrow n=\frac{578536730}{1000000}\approx 578$$

برای تعیین محل 578 نقطه به‌طور تصادفی بر روی نقشه چهار گوش دشت روانسر از تابع ()spsample از بسته نرم‌افزاری sp استفاده می‌شود.

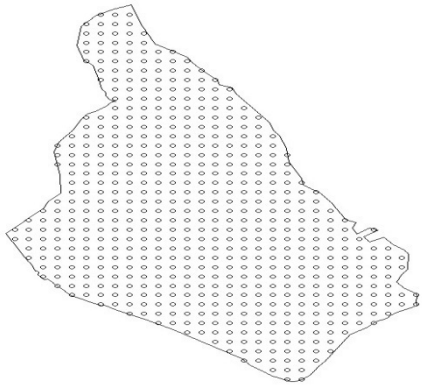
```
obs <- spsample(boundary, n = 578, "regular")
Str(obs)
```

در کادر زیر خلاصه‌ای از 578 نقطه انتخاب شده و موقعیت جغرافیایی آن‌ها ارایه شده است:

```
Formal class 'SpatialPoints' [package "sp"] with 3 slots
  ..@ coords     : num [1:576, 1:2] 650594 651594 648593 649593 650594
...
  .. ..- attr(*, "dimnames")=List of 2
  .. .. ..$ : NULL
  .. .. ..$ : chr [1:2] "x1" "x2"
  ..@ bbox       : num [1:2, 1:2] 631585 3822423 658597 3861440
  .. ..- attr(*, "dimnames")=List of 2
  .. .. ..$ : chr [1:2] "x1" "x2"
  .. .. ..$ : chr [1:2] "min" "max"
  ..@ proj4string:Formal class 'CRS' [package "sp"] with 1 slot
  .. .. ..@ projargs: chr "+proj=utm +zone=38 +datum=WGS84 +units=m
+no_defs +ellps=WGS84 +towgs84=0,0,0"
```

در مرحله آخر، نقشه نقاط بر روی نقشه پلی‌گونی محدوده اراضی منطقه با توابع زیر ترسیم می‌شود.

```
plot(boundary)
points(obs, pch = 1)
```



شکل 3-4- نمونه‌برداری باشبکه منظم 1000متر *1000 متر

3-2-2- نمونه‌برداری با پوشش مکانی و پرشده مکانی

در مناطقی با شکل نامنظم، اگر از نمونه‌برداری به روش شبکه‌ای منظم استفاده شود، گسترش جغرافیایی مکان‌های نمونه‌برداری در سراسر منطقه مطالعه شده ممکن است به حالت بهینه نباشد یعنی در قسمت‌هایی از منطقه مطالعه شده، فاصله تا نزدیک‌ترین نقطه نمونه‌برداری تعیین‌شده می‌تواند به نسبت زیاد باشد. در این موارد، اگر بخواهیم محدودیت نمونه‌برداری بر روی یک شبکه منظم را کاهش دهیم بایستی نقاط شبکه را کمی به سمت مناطق تحت نمونه‌برداری نزدیک کنیم، تا الگوی مکانی نامنظم شود. این عمل منجر به نمونه‌برداری با پوشش مکانی می‌شود که در آن معیار هندسی تعریف شده برحسب فاصله بین گره‌های یک شبکه گسسته ریز و نقاط نمونه‌برداری به حداقل (کمترین) می‌رسد (Royle and Nychka, 1998). Brus و همکاران (2007) پیشنهاد نمودند برای رسیدن به این مطلوب، میانگین مربعات کوتاه‌ترین فاصله MSSD[1]) بایستی توسط میانگین‌های کا به حداقل برسد. مختصات مکانی مراکز سلول‌های یک شبکه گسسته (پیکسل‌ها) به عنوان متغیر در خوشه‌بندی میانگین‌های کای سلول‌های شبکه استفاده می‌شود. از مراکز خوشه‌ها به عنوان نقاط نمونه‌برداری استفاده می‌شود. اگر فردی از پیش اندازه‌گیری‌هایی را در مکان‌هایی با مختصات مکانی شناخته شده (داده‌های

[1] Mean Squared Smallest Distance (MSSD)

نقطه‌ای میراثی) داشته باشد و فرض بر این باشد که این اندازه‌گیری‌ها هنوز معتبر هستند، استفاده از این داده‌ها در نقشه‌برداری می‌تواند کارآمد باشد. در این حالت، مکان‌های جدید در همسایگی مکان‌های موجود انتخاب نمی‌شوند بلکه نمونه‌برداری در مناطق تحت نمونه‌برداری تکمیل می‌شود.

شکل 3-5 لایه‌ها و شکل 3-6 یک پوشش مکانی شامل لایه‌ها و نقاط نمونه‌برداری در مرکز لایه/ها را نشان می‌دهد. نمونه‌برداری با پوشش مکانی، داده‌های میراثی را در نظر نمی‌گیرد. اگر به هر دلیلی از جمله کیفیت پایین داده‌ها از داده‌های قدیمی استفاده نشود، نمونه‌برداری با پوشش مکانی روش مناسبی است. اگر هدف استفاده از داده‌های میراثی باشد می‌توان یک طرح نمونه‌برداری از نوع نمونه‌پرشده مکانی را طراحی نمود. طراحی نمونه‌برداری با پوشش مکانی و پرشده مکانی را می‌توان با بسته نرم‌افزاری spcosa در محیط R اجرا کرد (Walvoort et al., 2010a, b).

به فایل‌های SpatialCoverageSample.R و SpatialInfillSample.R در پیوست این دستورالعمل مراجعه شود.

3-2-2-1- اجرای طرح نمونه‌برداری با پوشش مکانی در R

در این بخش از داده‌های منطقه اتیوپی استفاده شده است. مراحل انجام کار به‌صورت زیر است:

1) نصب نرم‌افزار جاوا

2) فراخوانی بسته‌های نرم‌افزاری مورد نیاز با تابع (library)

```
library(sp)
library(rJava)
library(spcosa)
```

3) بارگیری داده‌ها

```
load("Data/CovariatesThreeWoredasEthiopia.RData")
str(grdEthiopia)
```

```
'data.frame': 10843 obs. of  7 variables:
 $ s1     : num  1789 1790 1788 1789 1790 ...
 $ s2     : num  739 739 740 740 740 ...
 $ dem    : num  1.08 1.08 1.1 1.1 1.1 ...
 $ evi    : num  0.299 0.284 0.294 0.27 0.269 ...
 $ rfl_NIR: num  0.246 0.235 0.253 0.232 0.227 ...
 $ rfl_red: num  0.071 0.0689 0.0709 0.0716 0.0723 ...
 $ lst    : num  310 310 311 311 310 ...
```

4) تبدیل دیتا فریم به کلاس SpatialPointsDataFrame

```
coordinates(grdEthiopia)<-~s1+s2
```

```
Formal class 'SpatialPointsDataFrame' [package "sp"] with 5 slots
  ..@ data       :'data.frame':10843 obs. of  5 variables:
  .. ..$ dem    : num [1:10843] 1.08 1.08 1.1 1.1 1.1 ...
  .. ..$ evi    : num [1:10843] 0.299 0.284 0.294 0.27 0.269 ...
  .. ..$ rfl_NIR: num [1:10843] 0.246 0.235 0.253 0.232 0.227 ...
  .. ..$ rfl_red: num [1:10843] 0.071 0.0689 0.0709 0.0716 0.0723 ...
  .. ..$ lst    : num [1:10843] 310 310 311 311 310 ...
  ..@ coords.nrs : int [1:2] 1 2
  ..@ coords     : num [1:10843, 1:2] 1789 1790 1788 1789 1790 ...
  .. ..- attr(*, "dimnames")=List of 2
  .. .. ..$ : chr [1:10843] "1" "2" "3" "4" ...
  .. .. ..$ : chr [1:2] "s1" "s2"
  ..@ bbox       : num [1:2, 1:2] 1736 739 1897 906
  .. ..- attr(*, "dimnames")=List of 2
  .. .. ..$ : chr [1:2] "s1" "s2"
  .. .. ..$ : chr [1:2] "min" "max"
  ..@ proj4string:Formal class 'CRS' [package "sp"] with 1 slot
  .. .. ..@ projargs: chr NA
```

5) تبدیل دیتا فریم به کلاسSpatialPointsDataFrame به SpatialPixelsDataFrame

```
gridded(grdEthiopia)<-TRUE
```

6) تعیین تعداد نقاط نمونه‌برداری

```
n<-100
```

7) انتخاب موقعیت نقاط نمونه‌برداری با تابع (stratify)

این تابع شامل شناسه‌هایی به شرح زیر است :

```
stratify (object, nStrata, priorPoints = NULL, maxIterations = 1000, nTry = 1, equalArea = FALSE, verbose = getOption("verbose"))
```

object= داده‌ها به فرم"SpatialPixels"،"SpatialGrid"یا "SpatialPolygons"

nStrata= تعداد لایه‌ها (nStrata >= 1)

priorPoints= شامل نقاط نمونه‌برداری پیشین (میراثی) به فرمت"SpatialPoints"

maxIterations= بیشترین تعداد تکرارها

nTry= پیکربندی‌های اولیه را امتحان می‌کند و بهترین را نگه می‌دارد.

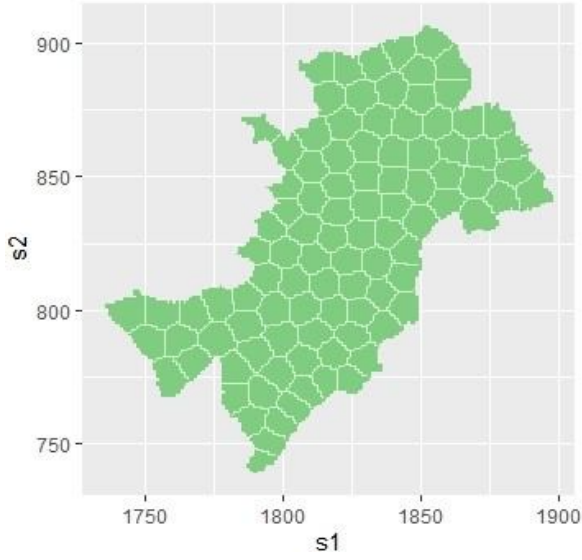
nGridCells= در صورتی که داده‌ها به فرمت"SpatialPolygons"باشند، تعداد تقریبی سلول‌های شبکه در این فرمت داده بیان می‌کند.

cellSize= در صورتی که داده‌ها به فرمت"SpatialPolygons"باشند، اندازه سلول برای گسسته‌سازی نقشه‌برداری در این فرمت داده استفاده می‌شود. توجه داشته باشید که cellize بر شناسه nGridCells برتری دارد.

equalArea= اگرFALSE انتخاب شود، الگوریتم منجر به ایجاد لایه‌های جمع و جور شده و اگر TRUE انتخاب شود، الگوریتم منجر به ایجاد لایه‌های فشرده با اندازه مساوی می‌شود.

Verbose= اگر TRUE انتخاب شود، اطلاعات پردازش یافته و نتایج میانی در دستگاه خروجی چاپ می‌شود.

```
set.seed(314)
myStrata <- stratify(grdEthiopia, nStrata = n, nTry=10)
plot(myStrata)
```



شکل3-5- نقشه تعداد لایه‌ها

8) استخراج میانگین مربعات کوتاه‌ترین فاصله

```
getObjectiveFunctionValue(myStrata)
```

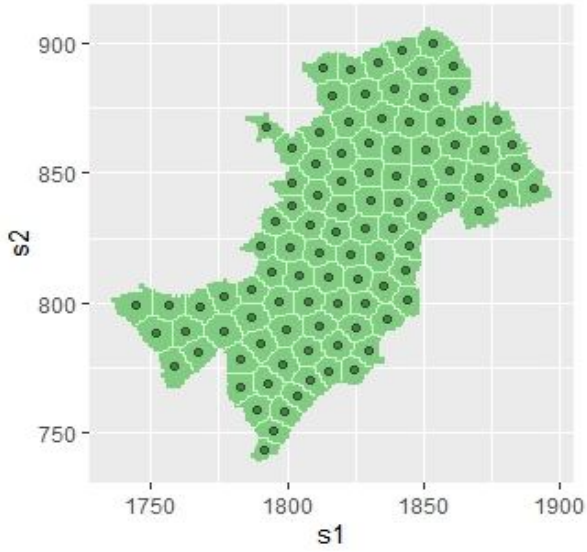
```
18.33166
```

9) انتخاب مراکز لایه‌ها با استفاده از تابع (spsample)

```
mySample <- spsample(myStrata)
```

10) رسم لایه‌ها و نقاط نمونه‌برداری (شکل 3-5)

```
plot(myStrata, mySample)
```



شکل 3-6- طراحی نمونه‌برداری با روش پوشش مکانی

2-2-3-2- اجرای طرح نمونه‌برداری پرشده مکانی در R

در این بخش از داده‌های منطقه اتیوپی استفاده شده است. مراحل انجام کار به‌صورت زیر است:

1) نصب نرم‌افزار جاوا

2) فراخوانی بسته‌های نرم‌افزاری مورد نیاز با تابع()library

```
library(sp)
library(ggplot2)
library(rJava)
library(spcosa)
```

3) بارگیری داده‌ها

```
load("Data/CovariatesThreeWoredasEthiopia.RData")
```

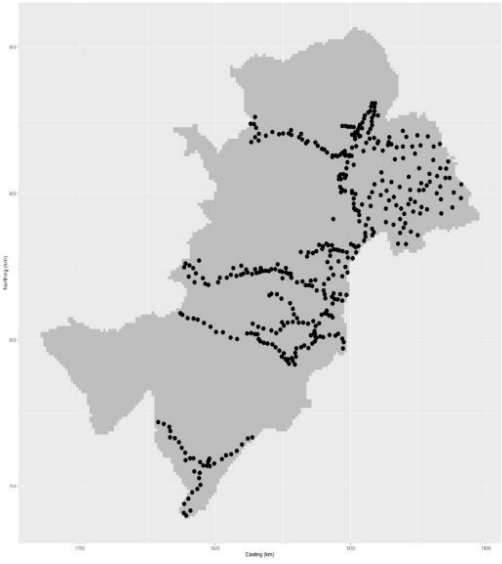
4) فراخوانی داده‌های نقطه‌ای موجود به محیط R

```
load(file="Data/DataThreeWoredasEthiopia.RData")
str(priordataEthiopia)
```

5) ترسیم نقشه نمونه‌های پیشین با استفاده از تابع ()ggplot

شکل 3-7 نشان می‌دهد داده‌های میراثی به طور عمده در امتداد جاده‌ها جمع‌آوری شده‌اند. این مثال خوبی از نمونه‌برداری آسان است.

```
ggplot(data = as.data.frame(priordataEthiopia))+
geom_raster(data=grdEthiopia,mapping = aes(x = s1,y = s2),fill="grey")+
geom_point(mapping = aes(x = s1,y = s2),size = 2)+
scale_x_continuous(name = "Easting (km)")+
scale_y_continuous(name = "Northing (km)")+
coord_equal(ratio = 1)
```



شکل 3-7- نقشه نمونه‌های پیشین

6) تبدیل دیتا فریم به کلاسSpatialPointsDataFrame

```
coordinates(grdEthiopia)<-~s1+s2
```

7) تبدیل دیتا فریم به کلاسSpatialPointsDataFrame به SpatialPixelsDataFrame

```
gridded(grdEthiopia)<-TRUE
```

8) تعیین تعداد نقاط نمونه‌برداری

```
n<-100
```

9) محاسبه اندازه کل نمونه (نمونه موجود + نمونه جدید)

```
ntot<-n+length(priordataEthiopia)
```

10) تغییر کلاس نقاط موجود از SpatialPointsDataFrame به SpatialPoints

```
priordataEthiopia<-as(priordataEthiopia,"SpatialPoints")
```

11) حذف سیستم تصویر داده‌های priordataEthiopia.

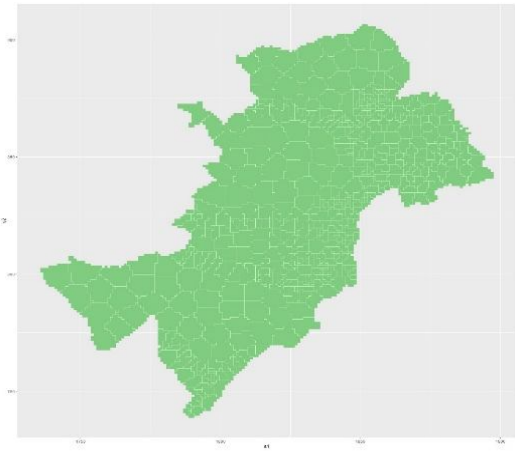
داده‌های grdEthiopia بدون ویژگی سیستم تصویرسازی است در حالی که priordataEthiopia سیستم تصویرسازی دارد. از این‌رو برای اینکه بتوان دو داده را روی هم نشان داد باید سیستم تصویر این داده‌ها را حذف کرد.

```
> proj4string(grdEthiopia)
[1] NA
> proj4string(priordataEthiopia)
[1] "+proj=laea +ellps=WGS84 +lat_0=5 +lon_0=20 +no_defs"
> proj4string(priordataEthiopia)<- NA_character_
> proj4string(priordataEthiopia)
[1] NA
```

12) انتخاب موقعیت نقاط نمونه‌برداری با تابع stratify

```
set.seed(314)
myStrata <- stratify(grdEthiopia, nStrata = ntot,        priorPoints = priordataEthiopia,
nTry=10)
```

تعداد لایه‌های حاصل از مجموع داده‌های قدیمی و جدید در شکل 3-8 نشان داده شده است.



شکل 3-8- نقشه جدید تعداد لایه‌های حاصل مجموع داده قدیمی و جدید (100+407)

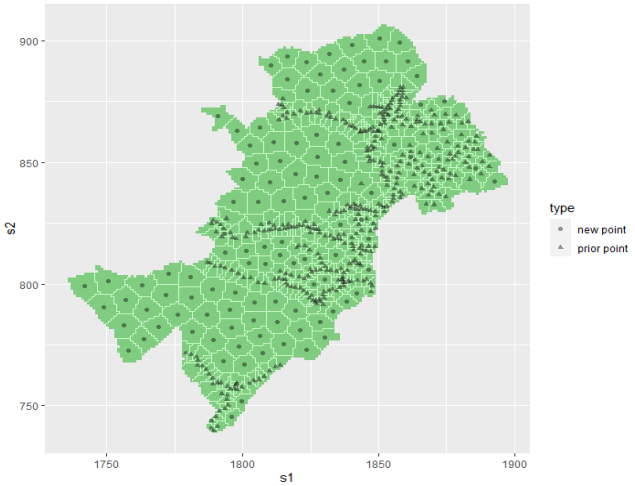
13) انتخاب مراکز لایه‌ها استفاده از تابع ()spsample

```
mySample <- spsample(myStrata)
```

14) ترسیم لایه‌ها و نقاط نمونه‌برداری (نقاط مراکز لایه‌ها و نقاط نمونه‌برداری قدیم)

شکل 3-9 نمونه پرشده مکانی 100 نقطه را در منطقه مطالعه شده اتیوپی نشان داده شده است.

```
plot(myStrata, mySample)
```



شکل 3-9- نقاط جدید و نقاط قدیم به همراه لایه‌ها

15) تعیین تعداد نقاط نمونه‌برداری شده جدید.

در mySample هم نقاط پیشین و هم نقاط جدید فهرست می‌شود. این که آیا یک نقطه قبلی است یا جدید در slot به اسم PriorPoint ذخیره می‌شود.

```
ids <- which(mySample@isPriorPoint==F)
ids
```

```
414 413 412 411 410 409 408 407 406 405 404 403 402 401 400 399 398 [1]
418 417 416 415
435 434 433 432 431 430 429 428 427 426 425 424 423 422 421 420 419 [22]
439 438 437 436
456 455 454 453 452 451 450 449 448 447 446 445 444 443 442 441 440 [43]
460 459 458 457
477 476 475 474 473 472 471 470 469 468 467 466 465 464 463 462 461 [64]
481 480 479 478
498 497 496 495 494 493 492 491 490 489 488 487 486 485 484 483 482 [85]
502 501 500 499
507 506 505 504 503 [106]
```

16) تبدیل mySample به دیتا فریم و استخراج مختصات نقاط جدید با کدهای زیر:

```
mySample <- as(mySample,"data.frame")
mySamplenew <- mySample[ids,]
str(mySamplenew)
```

```
data.frame':  110 obs. of  2 variables:
 $ s1: num  1765 1862 1806 1832 1843 ...
 $ s2: num  789 852 834 895 882 ...
```

3-3- نمونه‌برداری با استفاده میانگین‌های کا

در نمونه‌برداری با استفاده از شبکه منظم و پوشش مکانی، انتخاب مکان‌های نمونه‌برداری کاملاً براساس مختصات مکانی سلول‌های شبکه (پیکسل‌ها) صورت می‌گیرد و متغیرهایی کمکی مرتبط با خصوصیات خاک، در انتخاب مکان‌های نمونه‌برداری در نظر گرفته نمی‌شوند. اگر ویژگی خاک موردنظر با متغیرهای کمکی به عنوان مثال تصاویر سنجش از دور مرتبط باشد شیوه‌های نمونه‌برداری یادآوری شده کارآمد نخواهند بود. در زیر روش‌هایی برای انتخاب مکان‌های نمونه‌برداری بر اساس مقادیر متغیر کمکی در سلول‌های شبکه شرح داده شده است:

3-3-1- نمونه‌برداری با استفاده میانگین‌های کا سخت[1]

در نمونه‌برداری با استفاده از میانگین‌های کا سخت، از الگوریتم خوشه‌بندی میانگین‌های کا برای خوشه‌بندی سلول‌های رستری متغیرهای کمکی استفاده می‌شود. همانند نمونه‌برداری با استفاده از پوشش مکانی، میانگین مربعات کوتاه‌ترین فاصله MSSD باید به کمترین برسد، با این تفاوت که فاصله نه در یک فضای جغرافیایی بلکه در یک فضای p بعدی که توسط p تعداد متغیرکمکی (یک پلات پراکندگی چند بعدی تصور کنید که در آن متغیرهای کمکی در محورهای پلات قرار بگیرند) اندازه‌گیری می‌شود. در kmeans هر واحد فقط می‌تواند دقیقاً به یک خوشه تعلق داشته باشد. خوشه‌بندی میانگین‌های کا سخت واحدها (سلول‌های‌رستری) با استفاده از تابع kmeans در بسته نرم‌افزاری stats انجام می‌شود (به کدهای قابل اجرا در R با عنوان KMSample.R در پیوست مراجعه کنید).

3-3-1-1- اجرای طرح نمونه‌برداری kmeans سخت در R

در این بخش از داده‌های منطقه روانسر استفاده شده و مراحل انجام کار به‌صورت زیر است:

1) فراخوانی بسته‌های نرم‌افزاری لازم با استفاده از تابع ()library

[1] Hard k-means

```
library(sp)
library(fields)
library(ggplot2)
library(raster)
```

```
> library(sp)
Warning messages:
1: package ‘spcosa’ was built under R version 3.5.0
2: package ‘rJava’ was built under R version 3.3.3
3: In fun(libname, pkgname) : bytecode version mismatch; using eval
> library(fields)
Loading required package: spam
Loading required package: dotCall64
Loading required package: grid
Spam version 2.6-0 (2020-12-14) is loaded.
Type 'help( Spam)' or 'demo( spam)' for a short introduction
and overview of this package.
Help for individual functions is also obtained by adding the
suffix '.spam' to the function name, e.g. 'help( chol.spam)'.

Attaching package: ‘spam’

The following objects are masked from ‘package:base’:

backsolve, forwardsolve

See https://github.com/NCAR/Fields for
 an extensive vignette, other supplements and source code
Warning message:
failed to assign RegisteredNativeSymbol for toeplitz to toeplitz since
toeplitz is already defined in the ‘spam’ namespace
> library(ggplot2)
> library(raster)
```

2) خواندن داده‌های متغیرکمکی با اندازه پیکسل 90متر*90 متر

```
covariates <- read.table("km/covariates.txt",header = T)
summary(covariates)
```

```
clipdem90_.no_sinks.    MRVBF           Slope
 Min.   :1260         Min.   :0.000   Min.   :0.00000
 1st Qu.:1340         1st Qu.:0.165   1st Qu.:0.01424
 Median :1413         Median :1.212   Median :0.05594
 Mean   :1451         Mean   :1.965   Mean   :0.09363
 3rd Qu.:1516         3rd Qu.:3.910   3rd Qu.:0.14796
 Max.   :2176         Max.   :5.956   Max.   :0.74214
 Topographic_Wetness_Index grain_size_index       ndvi
 Min.   : 4.317            Min.   :-0.7892   Min.   :-0.0684
 1st Qu.: 7.150            1st Qu.:-0.6848   1st Qu.: 0.2043
 Median : 8.644            Median :-0.6695   Median : 0.2327
 Mean   : 9.243            Mean   :-0.6735   Mean   : 0.2472
 3rd Qu.:10.482            3rd Qu.:-0.6587   3rd Qu.: 0.2745
 Max.   :23.980            Max.   :-0.5882   Max.   : 0.7110
      s1              s2
 Min.   :630930   Min.   :3822090
 1st Qu.:639480   1st Qu.:3832890
 Median :644250   Median :3838470
 Mean   :644588   Mean   :3839717
 3rd Qu.:649380   3rd Qu.:3846120
 Max.   :659460   Max.   :3862320
```

3) انتخاب تعداد مکان‌های نمونه‌گیری

```
n<-20
```

4) خوشه‌ها با استفاده از تابع ()kmeans محاسبه می‌شوند. لازم به یادآوری است باید متغیرهای کمکی با تابع ()scale مقیاس‌بندی شوند.

این تابع خوشه‌بندی میانگین‌های کا را روی ماتریس داده‌ها انجام می‌دهد.

شناسه‌های تابع ()kmeans به شرح زیر است:

kmeans(x, centers, iter.max, nstart, algorithm = c("Hartigan-Wong",

"Lloyd", "Forgy", "MacQueen"), trace=FALSE)

x= ماتریس داده‌ها

centers = تعدادخوشه‌ها

iter.max= حداکثر (بیشترین) تعداد تکرارهای مجاز

nstart= تعداد پارتیشن‌ها برای شروع عملیات خوشه‌بندی به‌شکل تصادفی،1 <nstart اغلب توصیه می‌شود.

algorithm= کاراکتری است که نوع الگوریتم خوشه‌بندی را تعیین می‌کند. مانند "Hartigan-Wong"و"Lloyd"

مرسوم‌ترین الگوریتم‌ها توسط MacQueen (1967)، Lloyd (1957) و Forgy (1965) ارائه شده‌است. الگوریتم Hartigan و Wong (1979) به‌طور کلی عملکرد بهتری نسبت به بقیه دارد، اما اغلب چندین شروع تصادفی (nstart > 1) توصیه می‌شود.

خروجی تابع kmeans لیستی با چندین بیت اطلاعات است. مهمترین آن‌ها عبارتند از:

cluster= بردار اعداد صحیح (از 1تاk) نشان‌دهنده خوشه‌ای است که هر نقطه به آن اختصاص داده شده است.

centers= ماتریسی از مراکز خوشه.

totss= جمع کل مربعات.

withinss= جمع مربعات درون خوشه‌ای.

tot.withinss= جمع کل مربعات درون خوشه‌ای.

betweenss= جمع مربعات بین خوشه‌ای.

size: تعداد نقاط در هرخوشه.

```
set.seed(314)
myClusters <- kmeans(scale(covariates[,c(2,3,4,5,6,7)]), centers=n,
iter.max=100,nstart=10)
```

5) انتخاب مکان‌های نزدیک به مراکز خوشه‌ها

برای این منظور تابع rdist() از بسته نرم‌افزاری fields، استفاده می‌شود. این تابع فاصله اقلیدسی بین جفت نقاط در دو مجموعه داده مکانی را تا یک حد آستانه مشخص محاسبه می‌کند.

```
covariates$clusters <- myClusters$cluster
rdist.out <-
rdist(x1=myClusters$centers,x2=scale(covariates[,c(3,4,5,6,7)]))
ids.mindist <- apply(rdist.out,MARGIN=1,which.min)
mySample <- covariates [ids.mindist,]
```

6) ترسیم خوشه‌ها و نقاط نمونه‌برداری

تابع Pdf، درایور دستگاه گرافیک رایانه را برای تولید گرافیک به فرمت پی دی اف pdf راه‌اندازی می‌کند.

صورت عمومی این تابع برای فرمت‌های pdf یا png به شکل زیر است.

```
pdf (filename, width=7, height=7)
png (filename, width=480, height=480)
```

filename= مسیر و نام پوشه‌ای که به پسوند pdf. یا png. ختم می‌شود

width عرض تصویر (بر حسب اینچ).

height= ارتفاع تصویر (بر حسب اینچ).

7) ترسیم نمودار با استفاده از توابع معمول در R. کدنویسی با استفاده از تابع dev.off() به پایان می‌رسد. توابع tiff،bmp و jpeg نیز وجود دارند هرچند ثابت شده است که عملکرد jpeg ثبات کمتری نسبت به بقیه دارد.

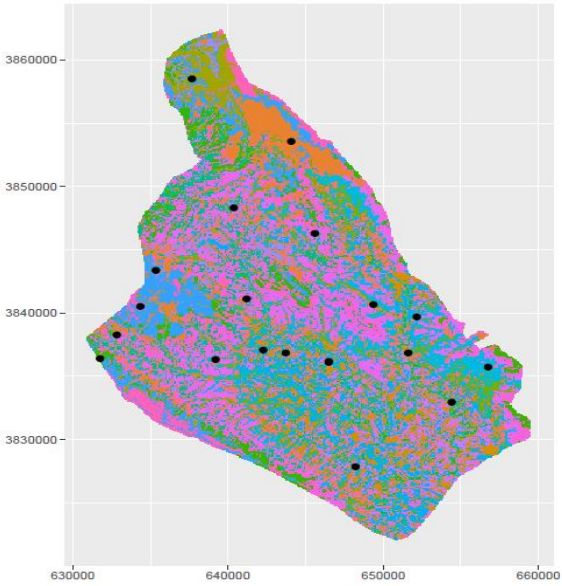
از تابع dev.off() برای اتمام اجرای کد استفاده می‌شود.

```
#plot clusters and sampling points
pdf(file = "FKM/FKMSample_ravansar.pdf", width = 7, height = 7)
ggplot(covariates) +
```

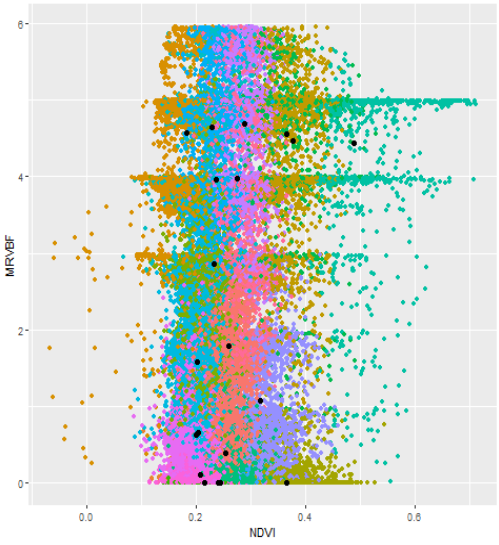
```
 geom_raster(mapping = aes(x = s1, y = s2, fill = factor(cluster))) +
 scale_fill_discrete(name = "cluster") +
 geom_point(data=myFKMSample,mapping=aes(x=s1,y=s2),size=2) +
scale_x_continuous(name = "") +
 scale_y_continuous(name = "") +
 coord_fixed() +
 theme(legend.position="none")
dev.off()

pdf (file = "Scatterplot_FKMSample_ravansar.pdf", width = 7, height = 7)
ggplot(grdHunterValley) +
 geom_point (mapping=aes(y=s2, x=s1,colour=factor(cluster))) +
 geom_point (data=myFKMSample, mapping=aes (y=MRVBF, x=ndvi),
size=2) +
 scale_y_continuous (name = "MRVBF") +
 scale_x_continuous (name = "NDVI") +
 theme (legend. position="none")
 dev.off ()
```

شکل 3-10، نقشه محل نقاط انتخاب‌شده به روش میانگین‌های کا سخت برای منطقه مطالعاتی روانسر را نشان می‌دهد. در اینجا از پنج متغیرکمکی ارتفاع، شیب، جهت شیب، شاخص توپوگرافی مرکب و شاخص پوشش گیاهی نرمال‌شده به عنوان متغیر کمکی استفاده شده است. تعداد نمونه، 20 تا در نظر گرفته شد. از این‌رو، 20 خوشه با استفاده از kmeans سخت ساخته شده است. توجه داشته باشید که تعداد خوشه‌ها براساس تعداد نمونه مورد نیاز است (تعداد خوشه‌ها برابر است با تعداد مکان‌های نمونه‌برداری). نخست داده‌های متغیرهای کمکی مقیاس‌بندی می‌شوند به گونه‌ای که انحراف معیار آن‌ها برابر با 1 شود. سلول‌های رستری (پیکسل‌هایی) با کمترین فاصله اقلیدسی در فضای مشخصه متغیر کمکی تا مرکز هسته خوشه‌ها، به عنوان نقاط نمونه‌برداری انتخاب می‌شوند. شکل 3-11 نمونه انتخاب شده را در نمودار

پراکندگی ارتفاع در برابر شاخص توپوگرافی مرکب نشان می‌دهد.



شکل 3-10- نمونه میانگین‌های کاسخت 20 نقطه‌ای در دشت روانسر



شکل 3-11- نمونه‌های میانگین‌های کاسخت که در نمودار پراکندگی NDVI در برابر MRVBF رسم شده‌اند.

2-3-3- نمونه‌برداری با استفاده میانگین‌های کا فازی[1]

بر خلاف میانگین‌های کا سخت، میانگین‌های کا فازی که به آن میانگین‌های کا نرم[2] نیز گفته می‌شود اجازه می‌دهد تا واحدها (سلول‌های رستری) به یک یا چند خوشه تعلق داشته باشند. یک بردار حاوی k عدد است که به هر واحد اختصاص داده می‌شود، اعداد در فاصله [صفر،یک] قرار دارند. مجموع این اعداد با عدد 1برابر است. اعداد نشان‌دهنده درجه متعلق بودن یک واحد به هر خوشه است. از آن‌ها به عنوان درجه عضویت یاد می‌شود. در میانگین‌های کا فازی، مرکز یک خوشه، میانگین وزنی متغیرها کمکی در تمام واحدها است. درجه‌های عضویت خوشه به عنوان اوزان در نظر گرفته می‌شوند. سلول‌های رستری (پیکسل‌ها) با کمترین فاصله اقلیدسی در فضای متغیر کمکی تا مرکز این خوشه‌های فازی به عنوان نقاط نمونه‌برداری انتخاب می‌شوند. این مکان‌ها با بیشترین عضویت در زیر مجموعه‌های فازی k....1 است. خوشه‌بندی kmeans یک روش شناخته شده برای انتخاب یک نمونه فرعی از یک نمونه بزرگتر با طیف‌سنجی NIR و vis-NIR است. برای مطالعه بیشتر می‌توان به مقالات Ramirez-Lopez و همکاران (2014) و Debaene همکاران (2014) مراجعه شود. در این دستورالعمل از تابع ()cmeans از بسته نرم‌افزاری e1071 برای انتخاب نمونه با استفاده از روش میانگین‌های کا فازی استفاده می‌شود. (به کدهای قابل اجرا در R با عنوان FKM.R در پیوست مراجعه کنید).

3-3-2-1- اجرای طرح نمونه‌برداری kmeans فازی در R

در این بخش از داده‌های منطقه روانسر استفاده شده و مراحل انجام کار به‌صورت زیر است:

1) فراخوانی بسته‌های نرم‌افزاری لازم با استفاده از تابع ()library

```
library(cluster)
library(fclust)
library(e1071)
library(ggplot2)
```

[1] Fuzzy k-means
[2] Soft k-means

2) فراخوانی متغیرهای کمکی منطقه مطالعه شده

```
Covariates <- read.table("FKM/covariates.txt",header = T)
```

3) بررسی ساختار داده‌ها

```
str(covariates )
```

```
'data.frame':71426 obs. of  9 variables:
 $ id                    : int  97 98 414 415 416 417 418 730 731 732 ...
 $ clipdem90_.no_sinks.  : num  1854 1886 1784 1809 1838 ...
 $ MRVBF                 : num  8.78e-07 6.76e-07 2.62e-05 5.82e-06 2.10e-06 ...
 $ Slope                 : num  0.355 0.349 0.279 0.353 0.336 ...
 $ Topographic_Wetness_Index: num  5.98 5.51 7.7 6.45 5.72 ...
 $ grain_size_index      : num  -0.694 -0.682 -0.688 -0.697 -0.695 ...
 $ ndvi                  : num  0.283 0.277 0.341 0.325 0.293 ...
 $ s1                    : int  639480 639570 639300 639390 639480 639570
639660 639030 639120 639210 ...
 $ s2                    : int  3862320 3862320 3862230 3862230 3862230 3862230
3862230 3862140 3862140 3862140 ...
```

4) بررسی همبستگی بین متغیرهای کمکی برای شناسایی هم راستایی خطی[1] بین متغیرهای کمکی

```
cor(covariates[,c(2,3,4,5,6,7)])
```

```
                        clipdem90_.no_sinks.      MRVBF      Slope
Topographic_Wetness_Index grain_size_index       ndvi
clipdem90_.no_sinks.            1.00000000 -0.6773996  0.7003682
-0.4533475       0.2919168 -0.06434285
MRVBF                          -0.67739960  1.0000000 -0.7093139
0.5638868      -0.4945830  0.29031679
Slope                           0.70036819 -0.7093139  1.0000000          -
```

[1] Collinearity

```
0.5721267       0.2327621 -0.11780471
Topographic_Wetness_Index       -0.45334747  0.5638868 -0.5721267
1.0000000      -0.2632819  0.15908201
grain_size_index                0.29191684 -0.4945830  0.2327621
-0.2632819       1.0000000 -0.67919026
ndvi                        -0.06434285  0.2903168 -0.1178047
0.1590820      -0.6791903  1.00000000
```

5) انجام PCA یکی از کاربردهای اصلی PCA در عملیات کاهشِ ابعاد[1] (ویژگی) و فایق آمدن بر مسئله هم راستایی خطی متغیرهای کمکی است. تجزیه به مولفه‌های اصلی (PCA) همان‌طور که از نامش پیداست می‌تواند مولفه‌های اصلی را شناسایی کند و به ما کمک کند تا به‌جای اینکه تمامیِ ویژگی‌ها را بررسی کرده، یک سری ویژگی‌هایی را که ارزشِ بیشتری دارند، تحلیل کنیم. در واقع PCA آن ویژگی‌هایی را که ارزش بیشتری فراهم می‌کنند برای ما استخراج می‌کند.

از تابع (prcomp) در بسته نرم‌افزاری stats برای انجام مولفه اصلی استفاده می‌شود:

prcomp(formula, data, scale. = FALSE...)

formula= یک فرمول بدون متغیر پاسخ، فقط به متغیرهای کمّی اشاره دارد

data= یک یک دیتا فریم شامل متغیرهای موجود در فرمول است.

Scale= یک مقدار منطقی است که نشان می‌دهد آیا متغیرها پیش از تجزیه و تحلیل باید مقیاس‌بندی شوند تا دارای واریانس واحد باشند. به‌شکل پیش فرض مقیاس‌بندی FALSE در نظر گرفته شده است، اما به‌طورکلی مقیاس‌بندی توصیه می‌شود.

```
covariates.pc <-
prcomp(~clipdem90_.no_sinks.+MRVBF+Slope+Topographic_Wetness_I
ndex+grain_size_index+ndvi, scale=TRUE, covariates)
x <- as.data.frame(covariates.pc$x[,c(1:6)])
str(x)
```

```
Standard deviations:
```

[1] Dimensionality Reduction

```
0.4601147 0.5230901 0.5831241 0.7585299 1.1915644 1.7830996 [1]
Rotation:
PC1 PC2 PC3 PC4 PC5 PC6
clipdem90_.no_sinks. -0.4414418 0.28338923 -0.47588029 -0.1856649 0.65006948 0.20321378
MRVBF 0.5021768 -0.05710735 0.13775040 0.1182583 0.66340425 -0.52104875
Slope -0.4609797 0.30095833 -0.07954465 0.5880471 -0.15770068 -0.56561959
Topographic_Wetness_Index 0.4022828 -0.17594594 -0.84968272 0.2071021 -0.20363148 -0.02989681
grain_size_index -0.3463775 -0.56896456 -0.13620152 -0.5152406 -0.07250922 -0.51674041
ndvi 0.2421497 0.68642242 -0.08788543 -0.5453255 -0.25634275 -0.31523385
```

```
x <- as.data.frame(covariates.pc$x[,c(1:6)])x <-
as.data.frame(covariates.pc$x[,c(1:6)])
```

```
data.frame':
71426obs. of 6 variables:
$ PC1: num -3.04 -3.4 -2.09 -2.61 -2.88...
$ PC2: num 2.84 2.55 2.83 3.26 2.91...
$ PC3: num -0.756 -0.798 -1.059 -0.766 -0.614...
$ PC4: num 0.85231 0.50161 -0.00449 0.66409 0.69459...
$ PC5: num 1.035 1.215 0.447 0.626 0.966...
$ PC6: num 0.0151 -0.1585 -0.0891 -0.1763 0.0791...
```

(1) نشان دادن رابطه بین متغیرهای کمکی و مولفه‌های اصلی با استفاده از تابع biplot()

```
biplot(covariates.pc, arrow.len=0.1,xlabs=rep(".",
length(covariates.pc$x[,1])),main="PCA biplot")
```

نتایج حاصل از اجرای تابع biplot را می‌توان براساس سه ویژگی حاصل از آن تفسیر نمود:

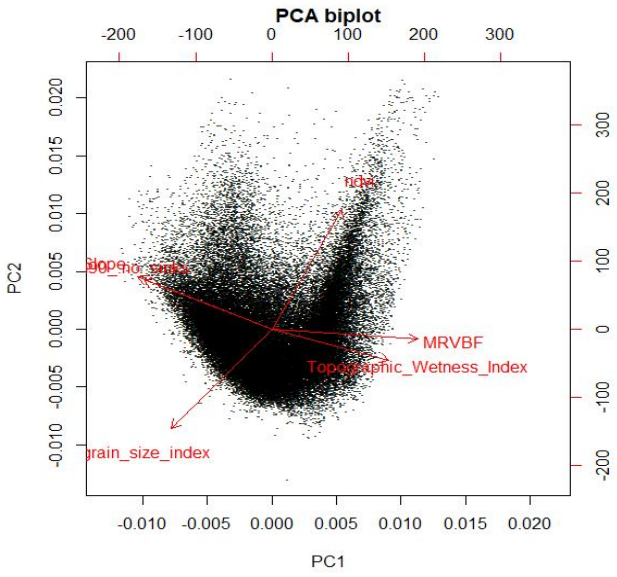
1- جهت بردار: با توجه به فضای تجزیه و تحلیل مولفه‌های اصلی، هرچه بردار متغیر

با محور فضای تجزیه و تحلیل مولفه‌های اصلی (در اینجا PC1 و PC2) به موازی نزدیک‌تر باشد، این متغیر سهم بیش‌تری در آن‌ها خواهد داشت.

2- طول‌بردار، هر چه طول بردار متغیر بیشتر باشد تغییرپذیری این متغیر که توسط PC1 و PC2 نشان داده می‌شود، بیشتر خواهد بود.

3- زوایای بین بردارهای متغیرهای مختلف: این زوایا همبستگی بین آن‌ها را نشان می‌دهد. زوایای کوچک نشان‌دهنده همبستگی مثبت بالا، زاویه راست نشان دهنده عدم همبستگی است، زوایای مخالف نشان‌دهنده همبستگی منفی بالا هستند.

براین اساس شکل 3-12 نشان می‌دهد که تمام 5 متغیر کمکی همبستگی مثبت یا منفی با هم دارند. و متغیرکمکی MRVBF و شاخص خیسی توپوگرافی سهم بیشتری در مولفه PC1 دارند و شاخص NDVI به دلیل طول بردار بیشتر دارای تغییرپذیری زیاد در هر دو مولفه اصلی اول ودوم است.



شکل 3-12- نمودار مولفه‌های اصلی اول و دوم و وضعیت متغیرهای کمکی در فضای مشخصه آن‌ها

```
x <- as.matrix(x)
str(x)
```

```
num [1:71426, 1:6] -3.04 -3.4 -2.09 -2.61 -2.88 ...
 - attr(*, "dimnames")=List of 2
  ..$ : chr [1:71426] "1" "2" "3" "4" ...
  ..$ : chr [1:6] "PC1" "PC2" "PC3" "PC4" ...
```

7) استفاده از تابع ()cmeans از بسته نرم‌افزاری e1071 برای خوشه‌بندی متغیرهای کمکی

```
kmeansx20 <- cmeans(x, 20, iter.max= 100, verbose = T, dist =
"euclidean", method = "cmeans", m = 1.3)
```

در این تابع

x: ماتریس داده‌های حاصل از تجزیه به مولفه‌های اصلی که ستون‌ها با مولفه‌های اصلی و ردیف‌ها با مشاهدات مطابقت دارند.

مراکز: تعداد خوشه‌ها یا مقادیر اولیه برای مراکز خوشه در اینجا برای این شناسه عدد 20 درج شده است.

iter.max: بیشترین تعداد تکرار.

verbose: اگر درست است، در حین یادگیری مقداری نتیجه بگیرید.

dist: نوع فاصله که محاسبه می‌شود اگر "اقلیدسی[1]" باشد، میانگین خطای مربع، اگر "منهتن[2]" باشد، میانگین خطای مطلق محاسبه می‌شود.

7) در مرحله بعد با دستورات زیرتعداد مکان‌های نمونه‌برداری انتخاب می‌شود.

```
m <- as.data.frame(kmeansx20$membership)
str(m)
```

```
'data.frame':71426 obs. of  20 variables:
 $ 1 : num  0.006594 0.000831 0.000798 0.009897 0.011261 ...
 $ 2 : num  4.72e-05 4.27e-06 1.10e-05 1.05e-04 8.88e-05 ...
 $ 3 : num  5.90e-05 6.33e-06 1.90e-05 1.32e-04 1.08e-04 ...
```

[1] Euclidean
[2] Manhattan

```
$ 4 : num  0.00328 0.000274 0.002362 0.010291 0.007682 ...
$ 5 : num  2.19e-05 1.73e-06 8.71e-06 7.98e-05 4.50e-05 ...
$ 6 : num  0.001938 0.000268 0.00055 0.003243 0.003737 ...
$ 7 : num  0.01626 0.00325 0.00597 0.02152 0.03194 ...
$ 8 : num  5.60e-05 4.77e-06 2.15e-05 1.69e-04 1.15e-04 ...
$ 9 : num  1.23e-04 1.44e-05 3.54e-05 2.50e-04 2.30e-04 ...
$ 10: num  5.77e-05 5.81e-06 1.78e-05 1.39e-04 1.12e-04 ...
$ 11: num  0.000156 0.000016 0.000059 0.000364 0.000302 ...
$ 12: num  4.11e-04 4.56e-05 1.34e-04 8.47e-04 8.14e-04 ...
$ 13: num  4.64e-05 4.29e-06 1.68e-05 1.22e-04 8.84e-05 ...
$ 14: num  1.94e-05 1.64e-06 7.40e-06 5.90e-05 3.66e-05 ...
$ 15: num  1.97e-05 1.89e-06 6.05e-06 4.84e-05 3.52e-05 ...
$ 16: num  3.47e-04 3.03e-05 2.26e-04 1.07e-03 7.50e-04 ...
$ 17: num  4.73e-04 6.16e-05 1.57e-04 8.47e-04 8.72e-04 ...
$ 18: num  0.04947 0.00227 0.98192 0.71556 0.1861 ...
$ 19: num  0.9199 0.9928 0.0075 0.2338 0.7542 ...
:20 $ num  7.65e-04 8.44e-05 1.81e-04 1.49e-03 1.46e-03...
```

8) برای هر پیکسل، خوشه‌ای که بالاترین عدد عضویت دارد با تابع()apply انتخاب می‌شود.

```
covariates$cluster <- apply(m,MARGIN=1,which.max)
```

```
int [1:71426] 19 19 18 18 19 19 19 7 18 18...
```

9) با دستور زیر مکان‌های با بیشترین عضویت در هر خوشه انتخاب می‌شوند.

```
units <- apply(m,MARGIN=2,FUN=which.max)
myFKMSample <- covariates[units,]
```

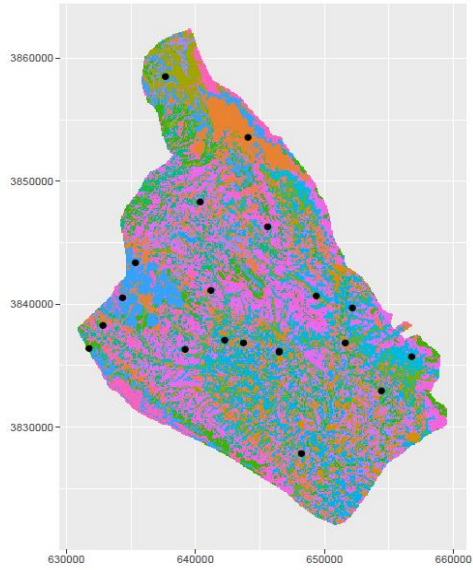
```
> myFKMSample
 id clipdem90_.no_sinks.      MRVBF      Slope
```

Topographic_Wetness_Index	grain_size_index	ndvi	s1	s2	cluster
42020	91883	1554.889	0.003894923	0.322369665	
6.068032	-0.6658767	0.2146383	631740	3836400	1
42776	92685	1336.578	4.571703911	0.006634983	
9.651773	-0.6998886	0.1821776	646500	3836220	2
55294	106667	1360.333	2.635218859	0.031008109	
12.907186	-0.6614328	0.2072551	641640	3832260	3
37465	86798	1522.000	0.334779620	0.192594543	
7.405242	-0.6868527	0.2898148	633450	3837840	4
40774	90509	1332.889	4.442118168	0.010937987	
10.706120	-0.7365576	0.4889618	651630	3836850	5
4381	25280	1569.778	0.158531412	0.176398203	
6.769922	-0.6582038	0.2260864	637860	3855210	6
25156	70243	1739.889	0.001788124	0.192899436	
6.763721	-0.6593929	0.2508611	636420	3842520	7
45480	95575	1333.444	4.817758083	0.004309775	
9.628441	-0.7022347	0.3302745	648210	3835410	8
30229	77717	1408.333	1.007661343	0.065333307	
9.056614	-0.6531885	0.1942598	648750	3840450	9
42471	92357	1339.778	4.751153946	0.008837905	
9.258477	-0.6735559	0.2495519	645690	3836310	10
42398	92284	1416.000	1.871305346	0.053229570	
10.221756	-0.6686104	0.2416867	639120	3836310	11
26473	72266	1451.111	0.493144095	0.085223943	
7.491136	-0.6647186	0.2163476	646230	3841980	12
32156	80307	1340.816	3.975913286	0.011247104	
13.120984	-0.6866525	0.2740398	652170	3839730	13
44029	94022	1326.778	4.603761673	0.016782627	
16.907812	-0.7148346	0.3600535	651990	3835860	14

```
39798  89448          1349.111 3.969752789 0.005857578
17.832022      -0.6790840 0.2356097 642270 3837120      15
15808  54005          1435.556 1.456426263 0.053994160
9.395054       -0.6829991 0.3069856 639210 3847110      16
5226   28213          1592.778 0.568466365 0.086807422
9.618626       -0.6542226 0.2074151 643440 3854400      17
3996   23711          1707.778 0.025753500 0.220623642
6.392500       -0.7003768 0.3608467 640200 3855660      18
57906 109800          1826.667 0.000009800 0.367148340
6.043368       -0.6696168 0.2270947 636510 3831360      19
17620    57206                    1440.667  0.228595883  0.188346744
6.804443       -0.6638125 0.2067102 640200 3846210      2
```

10) نقشه خوشه‌ها و نقاط نمونه‌برداری انتخاب شده از تابع ()pdf و تابع ()ggplot و با استفاده از دستورات زیر ترسیم می‌شود (شکل 3-13).

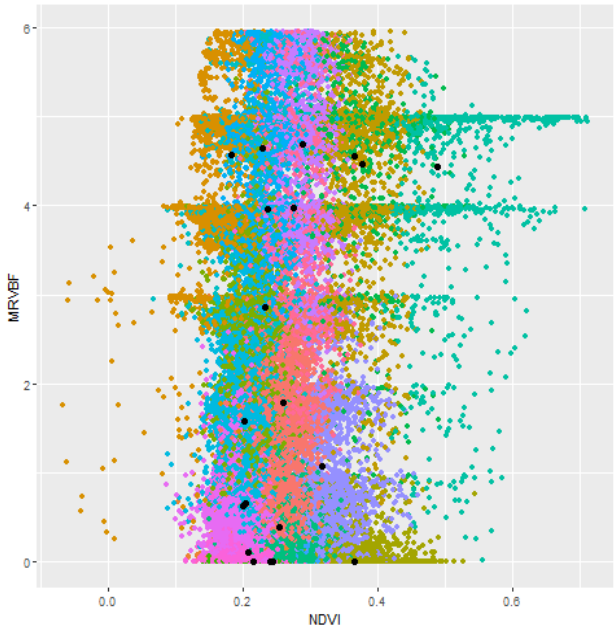
```
pdf(file = "FKM/FKMSample_ravansar.pdf", width = 7, height = 7)
ggplot(covariates)+
geom_raster(mapping = aes(x = s1, y = s2, fill = factor(cluster)))+
scale_fill_discrete(name = "cluster")+
geom_point(data=myFKMSample,mapping=aes(x=s1,y=s2),size=2)+
scale_x_continuous(name = "")+
scale_y_continuous(name = "")+
coord_fixed+ ()
theme(legend.position="none")
dev.off()
```



شکل 3-13- مکان‌های نمونه‌برداری (20 مکان) با استفاده از روش میانگین‌های کا فازی در دشت روانسر

11) نشان دادن نمونه‌های میانگین‌های کا فازی در نمودار پراکندگی NDVI در برابر MRVBF با استفاده از توابع بند 10 (شکل 3-14).

```
pdf (file = "FKM/Scatterplot_FKMSample_ravansar.pdf", width = 7, height = 7)
ggplot(covariates)+
geom_point(mapping=aes(y=s2, x=s1,colour=factor(cluster)))+
geom_point(data=myFKMSample,mapping=aes(y=MRVBF,x=ndvi),size=2)+
scale_y_continuous(name = "MRVBF")+
scale_x_continuous(name = "NDVI")+
theme(legend.position="none")
dev.off()
```



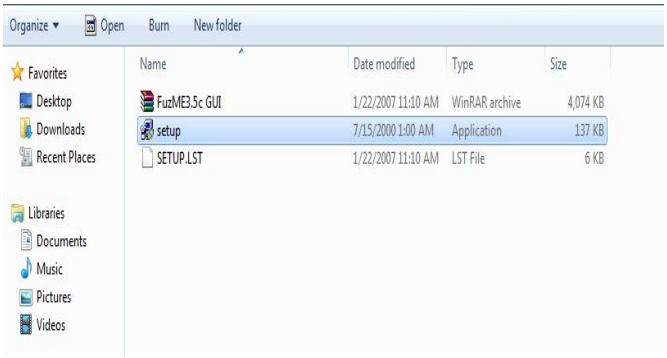
شکل 3-14- نمونه‌های *kmeans* فازی (نمودار پراکندگی NDVI در برابر MRVBF)

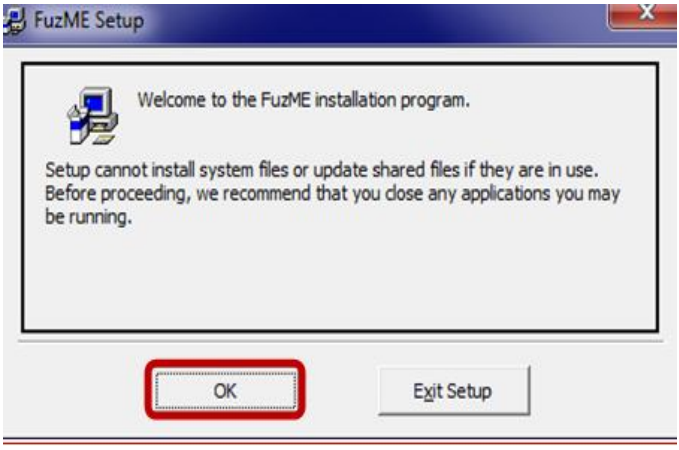
همچنین تجربه نشان داده است هنگامی که تعداد پیکسل‌های منطقه مطالعه شده زیاد باشد زمان محاسبه مکان‌های نمونه‌برداری با بسته‌های نرم‌افزار R طولانی می‌شود. از این‌رو، برای پردازش داده‌های بزرگ، به رایانه با حافظه مناسب نیاز خواهیم داشت. در این شرایط توصیه می‌شود که از نرم‌افزار FuzME استفاده شود. این نرم‌افزار از آدرس اینترنتی زیر قابل بارگیری است:

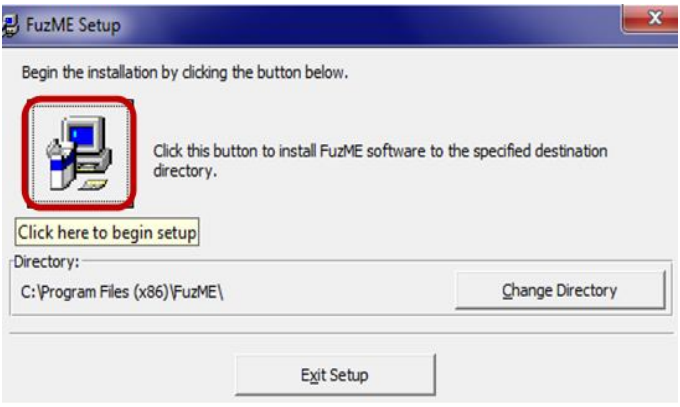
https://precision-agriculture.sydney.edu.au/resources/software/fuzme

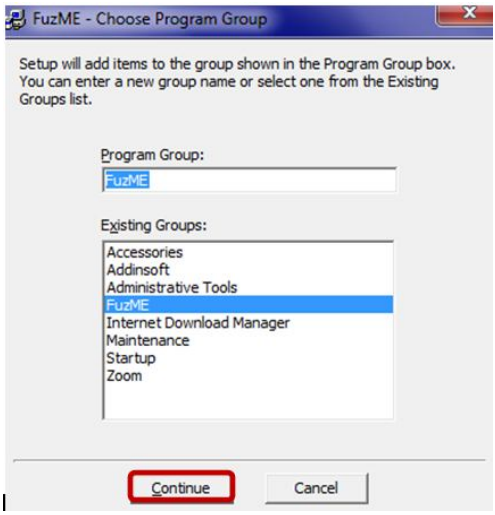
در ادامه مراحل انجام نمونه‌برداری به روش kmeans فازی با استفاده از نرم‌افزار FuzME به تفصیل بیان شده است.

1) پس از دانلود نرم‌افزار FuzME و خروج آن حالت زیپ، بر روی فایل setup کلیک شود و مراحل زیر انجام تا نرم‌افزار نصب شود.


Organize
Open
Burn
New folder
Favorites
Desktop
Downloads
Recent Places
Libraries
Documents
Music
Pictures
Videos
Name
Date modified
Type
Size
FuzME3.5c GUI
1/22/2007 11:10 AM
WinRAR archive
4,074 KB
setup
7/15/2000 1:00 AM
Application
137 KB
SETUP.LST
1/22/2007 11:10 AM
LST File
6 KB

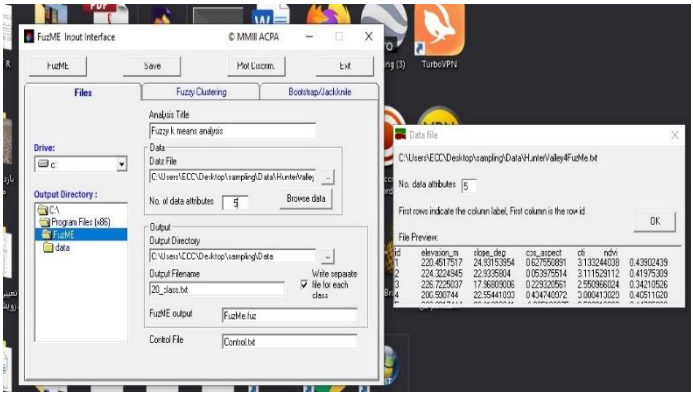

FuzME Setup
Welcome to the FuzME installation program.
Setup cannot install system files or update shared files if they are in use. Before proceeding, we recommend that you close any applications you may be running.
OK
Exit Setup


FuzME Setup
Begin the installation by clicking the button below.
Click this button to install FuzME software to the specified destination directory.
Click here to begin setup
Directory:
C:\Program Files (x86)\FuzME\
Change Directory
Exit Setup



2) براساس شکل زیر، از ایکن Files، Data File انتخاب می‌شود و فایل مورد نظر با عنوان، HunterValley4FuzMe.txt بارگیری می‌شود.

C:\Users\ECC\Desktop\sampling\Data\HunterValley4FuzMe.txt



بررسی شود که پنج متغیرکمکی elevation_m، slope_deg، cos_aspect، cti و ndvi وجود دارند.

3) دایرکتوری خروجی که همان پوشه Data است را انتخاب کنید:

C:\Users\ECC\Desktop\sampling\Data

4) به صفحه برگه خوشه‌بندی فازی (fuzzy clustering) بروید و براساس شکل زیر عمل کنید:

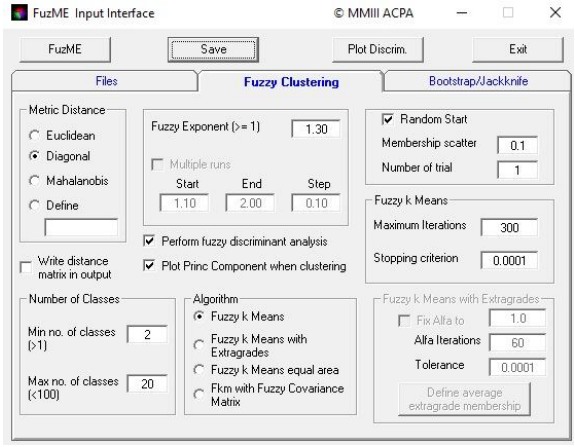
الف) Diagonal در جعبه Metric Distance تیک بزنید.

ب) در قسمت Numbers of Class کمترین کلاس را روی عدد 2 و بیشترین کلاس را روی عدد 20 قرار دهید.

ج) نمای فازی را روی1/3 تنظیم کنید.

د) تغییرات را با کلیک روی آیکن save ذخیره کنید.

ه) بر روی دکمه FuzME کلیک کنید.



از بسته نرم‌افزاری fuzme برای خوشه‌بندی با استفاده از فواصل ماهالانوبیس استفاده می‌شود.

5) ماتریس مربوط به عضویت‌ها از فایل خروجی خوانده می‌شود. فایل کدهای قابل اجرا در R که در پیوست با عنوان FKMSample_FuzMe است را اجرا کنید.

6) بارگیری داده‌های دره هانتر

```
load(file="Data/HunterValley4Practicals.RData")
```

7) بارگیری فایل خروجی FuzMe به نام FuzMeout.txt که شامل درجه عضویت‌های محاسبه شده برای هر کلاس است.

```
m <- read.csv(file=" Data /20_class.txt",sep="")
str(m)
```

با دستور خلاصه آماری مشخص می‌شود دیتا فریم شامل 23 ستون به ترتیب شامل id، بیشترین کلاس برای هر ردیف MaxCls و شاخص درهمی CI و بقیه ستون‌ها از X20a تا X20t درجه عضویت برای 20 خوشه (یا کلاس انتخاب شده) است.

```
> str(m)
'data.frame':                    22124 obs. of  23 variables:
 $ id    : int  1 2 3 4 5 6 7 8 9 10 ...
 $ MaxCls: chr  "20i" "20i" "20i" "20i" ...
 $ CI    : num  0.0629 0.0106 0.1908 0.2536 0.2367 ...
 $ X20a  : num  0.00007 0.00001 0.00003 0.00018 0.00016 0.00001 0.00002
0.00012 0.00087 0.00278 ...
 $ X20b  : num  0.00019 0.00003 0.00016 0.0006 0.00105 ...
 $ X20c  : num  0.00012 0.00002 0.00007 0.00035 0.00041 ...
 $ X20d  : num  0.00007 0.00001 0.00004 0.00017 0.00016 0.00001 0.00002
0.00015 0.00118 0.00345 ...
 $ X20e  : num  0.00027 0.00003 0.00024 0.00068 0.00054 ...
 $ X20f  : num  0.02621 0.00386 0.08128 0.10733 0.02613 ...
 $ X20g  : num  0.00035 0.00004 0.00027 0.00111 0.00078 ...
 $ X20h  : num  0.00007 0.00001 0.00004 0.00021 0.00032 0.00001 0.00003
0.00009 0.00174 0.0061 ...
 $ X20i  : num  0.963 0.993 0.89 0.854 0.862 ...
 $ X20j  : num  0.00007 0.00001 0.00004 0.00018 0.00029 0.00001 0.00003
0.00011 0.0021 0.00656 ...
 $ X20k  : num  0.00002 0 0.00001 0.00004 0.00005 0 0.00001 0.00003
0.00023 0.0006 ...
 $ X20l  : num  0.00011 0.00001 0.00006 0.0003 0.00025 0.00002 0.00003
0.00028 0.00203 0.00607 ...
 $ X20m  : num  0.00005 0.00001 0.00003 0.00015 0.00023 0.00001
0.00002 0.00008 0.00119 0.00399 ...
```

```
 $ X20n  : num  0.0064 0.0023 0.0232 0.0265 0.0985 ...
 $ X20o  : num  0.00003 0 0.00001 0.00008 0.00009 0 0.00001 0.00003 0.00025 0.00079 ...
 $ X20p  : num  0.00009 0.00001 0.00004 0.00026 0.00022 0.00001 0.00002 0.00016 0.00129 0.00423 ...
 $ X20q  : num  0.00021 0.00003 0.00019 0.00059 0.0007 ...
 $ X20r  : num  0.00148 0.00021 0.00242 0.00478 0.00283 ...
 $ X20s  : num  0.00015 0.00002 0.00014 0.00039 0.0007 ...
 $ X20t  : num  0.00072 0.00015 0.00123 0.00234 0.00483 ...
```

8) جداکردن ستون کلاس‌ها با کد زیر

```
m <- m[,-c(1,2,3)]
```

9) تعیین خوشه‌ای با بالاترین درجه عضویت برای هر سلول شبکه (پیکسل)

```
grdHunterValley$cluster <- apply(m,MARGIN=1,which.max)
```

10) محاسبه مکان‌های با بالاترین درجه عضویت برای کلاس‌ها یا خوشه‌های 1 تا 20

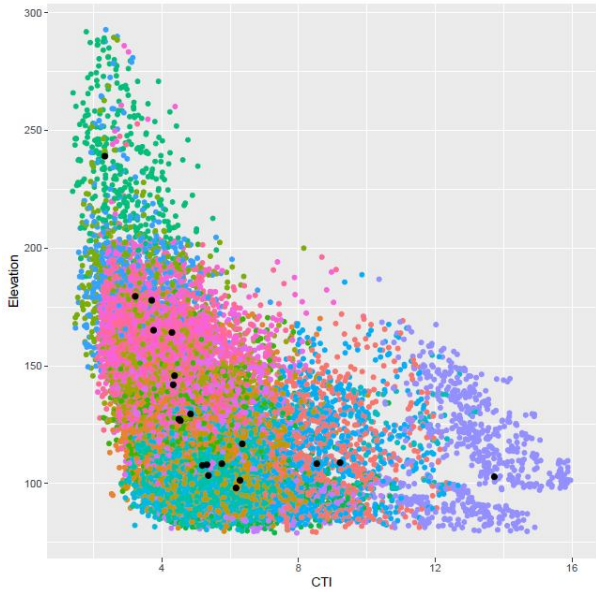
```
units <- apply (m, MARGIN=2, FUN=which.max)
myFKMSample <- grdHunterValley[units,]
```

11) با استفاده از کدهای زیر که پیش از این شرح داده شده‌اند نقشه نقاط انتخاب شده به روش kmeans فازی (شکل 3-15) و همین نقاط در فضای مشخصه ارتفاع در برابر شاخص توپوگرافی (شکل 3-16) تولید می‌شوند.

```
pdf(file = "Data/FKMSample_phi13_HunterValley.pdf", width = 7, height = 7)
ggplot(grdHunterValley) +
  geom_raster(mapping = aes(x = Easting, y = Northing, fill = factor(cluster))) +
scale_fill_discrete(name = "cluster") +

geom_point(data=myFKMSample,mapping=aes(x=Easting,y=Northing),size
=2) +
scale_x_continuous(name = "") +
```

```
scale_y_continuous(name = "") +
 coord_fixed() +
  theme(legend.position="none")
dev.off()
pdf(file = "Data/Scatterplot_FKMSample_phi13_HunterValley.pdf", width = 7, height = 7)
ggplot(grdHunterValley) +
   geom_point(mapping=aes(y=elevation_m,x=cti,colour=factor(cluster))) +
geom_point(data=myFKMSample,mapping=aes(y=elevation_m,x=cti),size=2) +
  scale_y_continuous(name = "Elevation") +
  scale_x_continuous(name = "CTI") +
  theme(legend.position="none")
dev.off()
```

شکل 3- 15- 20 نمونه حاصل از استفاده از روش *kmeans* فازی در دره هانتر



شکل 3- 16- نمونه‌های *kmeans* فازی (نمودار پراکندگی ارتفاع در برابر شاخص توپوگرافی)

لازم به یادآوری است خوشه‌بندی kmeans فازی را می‌توان با تابع FKM در بسته نرم‌افزاری fclust و تابع runFuzme دربسته نرم‌افزاری fuzme انجام داد. خوشه‌بندی با استفاده از فواصل ماهالانوبیس را می‌توان با تابع fanny در بسته نرم‌افزاری cluster در محیط R انجام داد.

3-3-3- نمونه‌برداری ابر مکعب لاتین مشروط[1]

نمونه‌برداری ابر مکعب لاتین یک روش نمونه‌برداری تصادفی لایه‌بندی شده است که روشی کارآمد برای نمونه‌برداری از متغیرها با توجه به توزیع چندمتغیره آن‌ها فراهم می‌کند. نخست به هدف شبیه‌سازی مونت کارلو و انتخاب کارآمد متغیرهای ورودی برای مدل‌های رایانه‌ای، توسعه یافت (McKay et al., 1979). این روش در علوم خاک و مطالعات زیست محیطی برای ارزیابی عدم قطعیت در یک مدل پیش‌بینی (Minasny and McBratney, 2002) و در زمین آمار برای شبیه‌سازی میدان‌های تصادفی گوسی

[1] Conditioned Latin hypercube sampling

استفاده شده است (Pebesma and Heuvelink, 1999).

نمونه‌برداری ابر مکعب لاتین از ایده مربع لاتین تبعیت می‌کند که در آن فقط یک نمونه در هر سطر و هر ستون وجود دارد. ابر مکعب لاتین این مفهوم را به تعداد دلخواهی از ابعاد تعمیم می‌دهد. نمونه‌برداری ابر مکعب لاتین به روش زیر کار می‌کند: با توجه به تعداد k متغیر $X_1, \ldots . X_K$ دامنه هر متغیر X به n فاصله محتمل برابر (لایه‌ها) تقسیم می‌شود، سپس برای هر متغیر در هر فاصله (لایه) یک نمونه تصادفی اخذ و مقادیر n بدست آمده برای هر یک از متغیرها به صورت تصادفی یا براساس برخی قوانین با با یکدیگر جفت می‌شوند.

در آخر n نمونه وجود دارد، جایی که نمونه‌ها n فاصله مربوط به همه متغیرها را پوشش می‌دهند. از این‌رو، این طرح نمونه‌برداری برای ابعاد (متغیرهای) بیشتر به نمونه‌های بیشتری نیاز ندارد. این روش تضمین می‌کند که هر یک از متغیرهای X به صورت کاملا لایه‌بندی شده نشان داده شوند. نمونه‌ای برای داده‌های دو بعدی در شکل3-17 آورده شده است. الگوریتم LHS به شرح زیر است:

- تقسیم توزیع هر متغیر X به n فاصله با احتمال برابر
- برای هر فاصله (لایه)، احتمال تجمعی نمونه را می‌توان به صورت زیر نوشت:

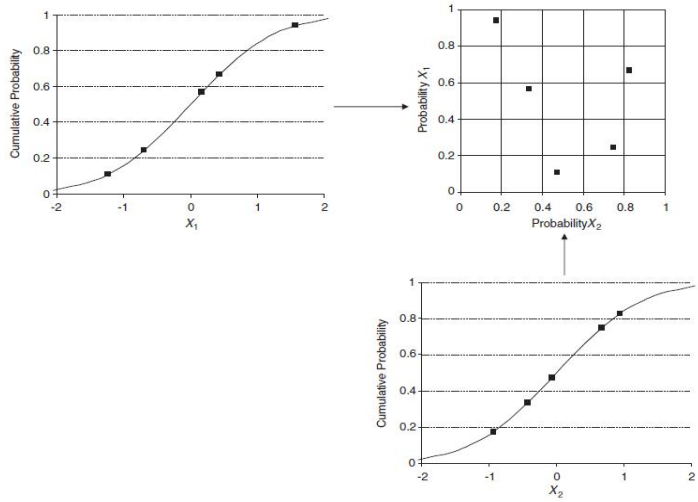
$$prob_i = (1/n)r_u + (i-1)/n \quad (2)$$

که در آن r_u یک عدد تصادفی یکنواخت است با دامنه 0 تا 1؛

- تبدیل احتمال به مقدار x نمونه‌برداری‌شده با استفاده از معکوس تابع توزیع F^{-1}

$$x = F^{-1}(prob) \quad (3)$$

مقادیر n (حجم نمونه) بدست آمده برای هر متغیر x به‌شکل تصادفی یا به ترتیب معین شده با مقادیر n متغیرهای دیگر جفت می‌شوند.



شکل 3- 17- مثالی از نمونه‌برداری ابر مکعب لاتین برای 2 متغیر با توزیع نرمال، احتمال تجمعی برای هر متغیر به 5 لایه برابر تقسیم می‌شود و از هر لایه یک نمونه تصادفی گرفته می‌شود، سپس پنج نمونه از هر متغیر به صورت تصادفی جفت و یک مربع لاتین تشکیل می‌دهند.

Minasny و McBratney (2006) ابر مکعب لاتین را برای مطالعات مشاهده‌ای سازگار نمودند. از این سازگاری به عنوان ابر مکعب لاتین مشروط (cLHS) یاد می‌شود. برای هر متغیر کمکی یک سری فواصل (لایه‌های مرزی) تعریف شده است. این لایه‌های مرزی به گونه‌ای انتخاب می‌شوند که تعداد پیکسل‌ها در آن‌ها برابر باشد. این کار را می‌توان با استفاده از مقادیر متناظر با احتمالات تجمعی با فاصله برابر به عنوان فواصل لایه‌ای انجام داد. به عنوان مثال، برای پنج لایه مرزی ما از تعداد متناظر با احتمالات تجمعی 0/2، 0/4، 0/6 و 0/8 استفاده می‌کنیم. در نمونه‌برداری ابر مکعب لاتین مشروط، توزیع مرزی متغیرهای کمکی در نمونه به توزیع آن‌ها در جمعیت نزدیک است. این می‌تواند برای روش‌های نقشه‌برداری رقومی خاک که به روابط خطی متکی نیستند مانند تکنیک‌های یادگیری ماشین شامل درختان طبقه‌بندی و رگرسیون (CART) و جنگل‌های تصادفی سودمند باشد. تئوری این روش در زیر به تفصیل آمده است:

برای نمونه‌برداری از داده‌های کمکی موجود، نمی‌توانیم مستقیم نمونه‌برداری مکعب لاتین را در توزیع چند متغیره اعمال کنیم. نمونه‌های انتخاب شده توسط نمونه‌برداری

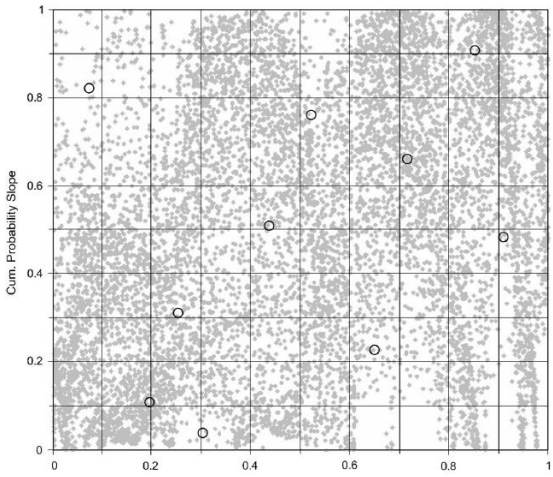
مکعب لاتین معمولی، ترکیبی از متغیرهای چندمتغیره را نشان می‌دهد که ممکن است در دنیای واقعی وجود نداشته باشد. به عنوان مثال، داده‌های ارتفاع و شیب یک سایت در دره هانتر را در نظر بگیرید. شکل 3-18 نمودار پراکندگی احتمال تجمعی داده‌های ارتفاع و شیب را به صورت نقاط خاکستری نشان می‌دهد. برای ترسیم 10 نمونه ابر مکعب لاتین از الگوریتم نمونه‌برداری مکعب لاتین معمولی همانطور که در بخش الگوریتم نمونه‌برداری مکعب لاتین ارائه شده است استفاده کردیم. 10 نمونه حاصل به صورت دایره در شکل 3-18 نشان داده شده‌است، می‌توانیم ببینیم که از 10 نمونه تنها 4 نمونه با نقاط داده منطبق هستند و 6 نمونه دیگر در دنیای واقعی وجود ندارند. این مشکل هنگام مواجهه با تعداد بیشتری از متغیرها و سایت‌های نمونه بارزتر خواهد شد. یک راه حل، ادامه تکرار نمونه‌برداری مکعب لاتین معمولی از توزیع و جستجوی داده‌ها و شناسایی داده‌هایی است که از نظر طبقه‌بندی با مقادیر انتخاب شده بیشترین شباهت را دارند. راه حل بهتر دیگر، انتخاب از سایت‌هایی است که ابر مکعب لاتین را در فضای ویژگی ایجاد می‌کند.

این مسئله به طریق زیر بهینه‌سازی می‌شود:

نمونه‌برداری ابر مکعب لاتین مشروط تلاش می‌کند توزیع چند بعدی مربوط به مجموعه‌ای از متغیرهای پیش‌بینی‌کننده با استفاده از نمونه‌برداری تصادفی طبقه‌ای را پوشش دهد. در یک فضای تک بعدی، توزیع تجمعی متغیرهای X به n (تعداد نقاط نمونه‌برداری) طبقه تقسیم می‌شود و ایده این است که یک نمونه در هر قشر انتخاب شود. با این حال، در فضای چند متغیره این کار پیچیده‌تر می‌شود. در ابر مکعب لاتین مشروط، زیرمجموعه واسنجی x با n نمونه گرفته شده از مجموعه X با N نمونه (که در آن $N>>n$) باید یک ابر مکعب لاتین را تشکیل دهد. این نیاز به یک روش جستجو دارد. Minasny و همکاران (2006) cLHS یک الگوریتم جستجو بر اساس برنامه بازپخت[1] پیشنهاد کردند (Metropolis et al., 1953). در ریاضیات و علوم کامپیوتر، مساله بهینه‌سازی، در واقع مساله پیدا کردن بهترین راه حل، از میان کلیه راه حل‌های ممکن برای مساله است. نوع خاصی از مسائل بهینه‌سازی وجود دارند که در آن‌ها، به دلیل زیاد

[1] Annealing schedule

شدن تعداد مشاهدات، مدیریت کردن آن‌ها با استفاده از «روش‌های ترکیبی امکان‌پذیر نیست برای حل چنین مسائلی، از الگوریتم شبیه سازی بازپخت استفاده می‌شود.



شکل 3- 18- نمودار پراکندگی احتمال تجمعی ارتفاع و شیب برای منطقه دره هانترنقاط خاکستری نشان دهنده داده‌ها در منطقه است و دایره‌های سیاه نشان دهنده نمونه‌هایی است که با استفاده از ابر مکعب لاتین معمولی کشیده شده است.

در متالورژی و علم مواد، بازپخت یا انیلینگ، به عملیات حرارتی گفته می‌شود که طی آن مشخصات فیزیکی و گاهی، مشخصات شیمیایی ماده تغییر می‌کند. این کار با هدف افزایش شکل‌پذیری یا کاهش سختی ماده انجام می‌شود. در این فرایند، نخست فلز گرم شده، سپس در یک درجه حرارت خاص نگه داشته و در پایان، به تدریج سرد می‌شود. با گرم کردن فلز، مولکول‌ها آزادانه به هر سوی حرکت می‌کنند. با سرد کردن تدریجی ماده، این آزادی کاهش پیدا می‌کند اگر فرایند سرد کردن به اندازه کافی کند باشد که بتوان اطمینان حاصل کرد فلز در هر مرحله در تعادل ترمودینامیکی قرار دارد، می‌توان مطمئن شد که انرژی گرمایی به طور یکنواخت در جسم توزیع شده و بهترین ساختار بلوری در آن وجود دارد که متقارن و مقاوم است. در الگوریتم برنامه بازپخت، از فرایند یاد شده الگوبرداری شده است (https://blog.faradars.org). این الگوریتم هم

روی داده‌های پیوسته و هم کیفی عمل می‌کند. برای متغیرهای پیوسته، نخست حجم نمونه nرا تعریف می‌کنیم. هر مولفه X با (N*K اندازه) بر اساس توزیع آن‌ها به n لایه محتمل مساوی تقسیم می‌شود و x با (n*k اندازه) زیرمجموعه‌ای از X است. ماتریس η تعریف می‌شود که تعداد x را که در داخل هر لایه تعریف شده قرار می‌گیرد، محاسبه می‌کند:

$$\eta = \begin{pmatrix} \eta_{11} & \cdots & \eta_{1k} \\ \vdots & \ddots & \vdots \\ \eta_{n1} & \cdots & \eta_{nk} \end{pmatrix} \quad (4)$$

در اینجا ردیف‌ها نشان دهنده لایه‌های $s_1, \dots, s_n$ هستند و ستون‌ها متغیرهای $X_1, \dots, X_k$ را نشان می‌دهد. یک نمونه‌برداری مکعب لاتین واقعی برای تمام سلول‌های ماتریس η مقادیر 1 خواهد داشت. چالش این است که x را پیدا کنید که مقدار η نزدیک یا برابر 1 داشته باشد. الگوریتم نمونه‌برداری ابر مکعب لاتین مشروطرا به صورت زیر اجرا :

(1) توزیع کمی X را به n لایه تقسیم می‌شود و توزیع کمی برای هر یک از متغیرهای $q_j^i, \dots, q_j^{n+1}$ برای $j = 1, \dots, k$ و $i = 1, \dots, n+1$ محاسبه می‌شود و محاسبه C، ماتریس همبستگی X انجام می‌شود.

(2) انتخاب n نمونه تصادفی از $N, x(1, \dots, n)$، که سایت‌های نمونه‌برداری است و $r(i = 1, \dots, N - n)$ که مکان‌های نمونه‌برداری نشده است، T ماتریس همبستگی x را محاسبه می‌شود.

(3) محاسبه تابع هدف برای متغیرهای پیوسته:

$$O_1 = \sum_i^n \sum_{j=1}^k \left| \eta\left(q_j^i \le x_j \le q_j^{i-1}\right) - 1 \right| \quad (5)$$

در اینجا k تعداد متغیرهای کمکی، $\eta\left(q_j^i \le x_j \le q_j^{i-1}\right)$ تعدادی از نمونه‌ها در x_j است که توزیع تجمعی آن‌ها در بین کمیت‌های q_j^i و q_j^{i-1} قرار می‌گیرد.

برای داده‌های کاتگوری تابع هدف با توزیع احتمال برای هر کلاس مطابقت دارد:

$$O_2 = \sum_{j=1}^C \left| \frac{\eta x_j}{n} - k_j \right| \quad (6)$$

در اینجا ηx_j تعدادی از نمونه x است که متعلق به کلاس j در داده‌های نمونه‌برداری

شده هستند و k_j نسبتی از کلاس j در X است.

برای اطمینان از این که همبستگی متغیرهای نمونه‌برداری شده جایگزینی از داده‌های اولیه است تابع هدف دیگری اضافه می‌شود:

$$O_3 = \sum_{i=1}^{k}\sum_{j=1}^{k}|c_{ij} - t_{ij}| \quad (7)$$

در اینجا c ماتریس همبستگی X و t ماتریس همبستگی x است.

تابع هدف کلی از رابطه زیر بدست آید:

$$O = w_1O_1 + w_2O_2 + w_3O_3 \quad (8)$$

در اینجا w وزن اختصاص یافته به هر مولفه از تابع هدف می‌باشد برای کاربردهای عمومی مقدار آن برای تمام مولفه‌های تابع هدف 1 در نظر گرفته می‌شود.

(4) یک برنامه بازپخت اجرا می‌شود:

$$Metro = exp^{[-\Delta O/T]} \quad (9)$$

که درآن ΔO تغییرات تابع هدف را نشان می‌دهد و T یک خنک کننده دما است (با مقدار بین 0 و 1) که در هر تکرار با فاکتور d کاهش می‌یابد.

(5) $rand$ یک عدد تصادفی یکنواخت بین 0 و 1 ایجاد می‌کند، اگر $rand < Metro$ باشد مقادیر جدید را می‌پذیرید، درغیراین صورت تغییرات را کنار می‌گذارد.

(6) تغییرات زیر باید انجام شود:

تولید یک عدد تصادفی یکنواخت $rand$

اگر $rand < p$ باشد، نمونه‌ای را به شکل تصادفی از x انتخاب کرده و آن را بایک نمونه تصادفی از مخزن r مبادله می‌کند.

مقدار p بین 0 تا 1 بوده که نشان‌دهنده احتمال جستجو به صورت تصادفی است، در غیر از این، نمونه‌هایی را از x که دارای بیشترین مقدار $\eta\left(q_j^i \le x_j \le q_j^{i-1}\right)$ است حذف نموده و آن را با نمونه(های) تصادفی از مخزن r جایگزین می‌کند.

(7) رفتن به مرحله 3

تکرار مراحل (3)-(7) تا زمانی که مقدار تابع هدف فراتر از یک معیار توقف یا تعداد

مشخصی از تکرار باشد.

این تکنیک چندین برتری دارد:

1. متغیرهای پیوسته و همچنین کیفی می‌توانند در آن وارد شوند.

2. نمونه‌برداری بر اساس توزیع تجربی داده‌های اصلی است، از این‌رو غیرپارامتری است.

3. مختصات مکانی را می‌توان برای اطمینان از گسترش خوب نقاط نمونه‌برداری درصورت لزوم استفاده کرد.

4- محدودیت‌های اضافی می‌توانند بر تابع هدف اعمال شوند، به عنوان مثال. فاصله از جاده‌ها و مرزهای مزرعه.

نمونه‌های با روش نمونه‌برداری ابر مکعب لاتین مشروط را می‌توان با استفاده از تابع optimCLHS در بسته نرم‌افزاری spsann و تابع ()clhs از بسته نرم‌افزاری clhs درمحیط R انتخاب کرد (Samuel-Rosa, 2016).

1-3-3-3- اجرای طرح نمونه‌برداری ابرمکعب لاتین مشروط با بسته نرم‌افزاری spsann

1) فراخوانی بسته‌های نرم‌افزاری لازم با استفاده از تابع ()library

```
library(sp)
library(reshape)
library(ggplot2)
library(spsann)
```

2) خواندن داده‌ها

```
load(file="Data/HunterValley4Practicals.RData")
grid <- grdHunterValley
rm(grdHunterValley)
Str(grid)
```

```
'data.frame':        22124 obs. of  9 variables:
 $ Easting      : num  337610 337610 337610 337635 337635 ...
 $ Northing     : num  6368041 6368066 6368091 6368016 6368041 ...
```

```
 $ elevation_m       : num  220 224 227 207 210 ...
 $ slope_deg         : num  24.9 22.9 18 22.6 22.8 ...
 $ cos_aspect        : num  0.628 0.054 0.229 0.435 -0.825 ...
 $ cti               : num  3.13 3.11 2.55 3.81 3.53 ...
 $ ndvi              : num  0.439 0.42 0.342 0.465 0.447 ...
 $ landuse_class     : Factor w/ 4 levels "Native forest",..: 2 2 2 1 1 2 2 2 2 2 ...
 $ soil_landscape_unit: Factor w/ 5 levels "3 Ways","Branxton",..: 4 4 4 1 4 4
4 4 4 4 ...
```

3) داده‌های شبکه‌ای دره هانتر دارای نقاط زیادی است (برای جلوگیری از مشکل حافظه رایانه)، یک زیرشبکه از داده‌های اصلی انتخاب می‌شود.

```
coordinates(grid) <- c("Easting","Northing")
gridded(grid) <- T
subgrid<- spsample(grid,type="regular",cellsize=50,offset=c(0.5,0.5))
subgriddata <- (subgrid %over% grid)
grd <- data.frame(coordinates(subgrid),subgriddata)
```

4) نقاط نمونه‌برداری لازم و متغیرهای کمکی را که باید در نمونه‌برداری ابر مکعب لاتین استفاده شود مشخص می‌شوند

```
candi <- grd[,1:2]
names(candi) <- c("x","y")
covars <- grd[, 3:7]
head(covars)
```

```
head(covars)
  elevation_m slope_deg cos_aspect      cti       ndvi
1    154.3400  9.093675 -0.5021247 5.326304 0.00552486
2    149.2310  5.553246  0.9999970 6.017144 0.13812155
3    183.9763 10.422250 -0.9850157 6.414457 0.20000000
4    175.7363 10.860250 -0.9569191 6.889666 0.17985612
5    169.7970  7.610561 -0.8156319 6.786407 0.21088435
6    174.1300 11.003140  0.5580999 2.185342 0.14124294
```

5) برنامه زمان‌بندی برای بازپخت شبیه‌سازی شده را تعریف کنید. توجه داشته باشید که هم میزان پذیرش اولیه و هم دمای اولیه[1] تنظیم شوند ممکن است عجیب به نظر برسد زیرا نرخ پذیرش تابعی از دمای اولیه است:

نرخ پذیرش اولیه[2] به عنوان مقدار آستانه استفاده می‌شود. اگر دمای اولیه‌ای انتخاب شود که منجر به میزان پذیرش کوچکتر از مقدار انتخاب شده برای نرخ پذیرش اولیه شود، در اینصورت بهینه‌سازی متوقف می‌شود. دراین حالت باید مقدار بزرگتری برای دمای اولیه انتخاب شود. در حالت پیش فرض نرخ پذیرش اولیه 0/8 و دمای اولیه 0/08 در نظر گرفته می‌شود.

برای اجرای برنامه زمان بندی از تابع *scheduleSPSANN ()* به شرح زیر استفاده می‌شود:

```
scheduleSPSANN (initial.acceptance = 0.95, initial.temperature = 0.001,
temperature.decrease = 0.95, chains = 500, chain.length = 1, stopping = 10,
x.max, x.min = 0, y.max, y.min = 0, cellsize)
```

initial.acceptance = دارای مقدار عددی بین صفر و یک است که احتمال پذیرش اولیه را مشخص می‌کند، یعنی نسبت پیکربندی‌های سیستم پیشنهادی که باید در زنجیره اول پذیرفته شود. در صورت دستیابی نداشتن به این مقدار، بهینه سازی متوقف و هشدار صادر می‌شود. مقدار پیش فرض اولیه initial.acceptance 0/95 است.

initial.temperature = مقدار آن بزرگتر از صفر است. این شناسه درجه حرارت اولیه سیستم را تعیین می‌کند. درجه حرارت اولیه پایین، همراه با پذیرش اولیه پایین، باعث می‌شود الگوریتم به عنوان یک الگوریتم حریص رفتار کند، یعنی فقط تنظیمات بهتر سیستم پذیرفته می‌شود به طور پیش فرض مقدار آن 0/001 است.

temperature.decrease = مقدار آن بین صفر تا یک به عنوان یک عامل ضرب برای کاهش دما در انتهای هر زنجیره مارکوف استفاده می‌شود. به طور پیش فرض مقدار آن 0/95 است.

[1] *Initial temperature*
[2] *Initial acceptance*

chains = عددی صحیح که حداکثر تعداد زنجیره‌ها را تعریف می‌کند. به طور پیش فرض مقدار آن 500 است.

chain.length = عددی صحیح که طول هر زنجیره مارکوف را نسبت به تعداد نقاط نمونه تعریف می‌کند به طور پیش فرض مقدار آن 1 است.

stopping = عددی صحیح که حداکثر تعداد مجاز زنجیره مارکوف را بدون بهبود مقدار تابع هدف تعریف می‌کند. به طور پیش فرض مقدار آن 10 است.

x.max, x.min, y.max, y.min = عددی است که کمترین و بیشترین مقدار نویز تصادفی تعریف شده را به مختصات x و y اضافه می‌کند.

cellsize= اندازه پیکسل

```
schedule <- scheduleSPSANN(initial.acceptance = 0.8,initial.temperature
=0.08,temperature.decrease=0.95,chains=1000,chain.length=4,stopping=1
0, x.min=10,y.min=10, cellsize=50)
```

6) اوزان و تعداد نمونه را تعیین نموده و سپس الگوریتم بازپخت شبیه‌سازی شده اجرا می‌شود.

برای اجرای برنامه بهینه‌سازی از تابع optimCLHS به شرح زیر استفاده می‌شود:

optimCLHS (points, candi, covars, use.coords = FALSE, clhs.version = c("paper", "fortran", "update"), schedule = scheduleSPSANN(),progress = "txt", weights)

points= تعداد نمونه‌ها.

candi= مختصات جغرافیایی محدوده مطالعه شده x و y (باید مختصات تصویرسازی شده و *UTM* باشد) به فرمت دیتا فریم یا ماتریس.

covars= به فرمت دیتا فریم یا ماتریس که شامل متغیرهای کمکی است.

use.coords= گزینه منطقی شامل FALSE یا TRUE که استفاده یا عدم استفاده از مختصات مکانی x و y را فراهم می‌سازد.

clhs.version= شناسه اختیاری که مشخص می‌کند از کدام نسخه CLHS باید استفاده شود. گزینه‌های موجود عبارتند از: "paper"، شامل فرمول‌های O1، O2 و O3 که در مقاله اصلی میناسنی و مک برتنی (2006) ارائه شده است. "fortran" ، برای فرمول‌های O1 و O3 که شامل یک عامل مقیاس‌بندی است که در زبان Fortran کد نویسی شده و توسط Minasny اجرا شده است (حدود 2015) ؛ و "update"، برای فرمول‌هایO1 ، O2 و O3 که شامل تغییرات پیشنهادی نویسندگان این بسته در سال 2018 است (به پایین مراجعه کنید) به صورت پیش‌فرض نرم افزار نسخه "paper" در نظر می‌گیرد.

Schedule = در بخش پیشین شرح داده شده است

Track= گزینه منطقی شامل FALSE یا TRUE است. آیا باید تکامل حالت انرژی ثبت و همراه با نتیجه بازگردانده شود؟ اگر مسیر FALSE = (به طور پیش فرض)، فقط حالت‌های شروع و پایان انرژی همراه با نتایج برمی‌گردند.

progress= شناسه (اختیاری) نوع نوار پیشرفت کار را تعیین می‌کند.

Weights= لیستی شامل وزن‌های اختصاص داده شده به هر یک از توابع هدف O1 ، O2 و O3 است. وزن‌ها باید برابر یا بزرگتر از صفر و مجموع آن‌ها به 1 برسد.

```
weights <- list(O1 = 0.5, O3 = 0.5)
samplesize<-20
set.seed(314)
res <- optimCLHS(points = samplesize, candi = candi, covars = covars, use.coords = FALSE ,schedule = schedule, track=TRUE, weights = weights, progress=NULL)
```

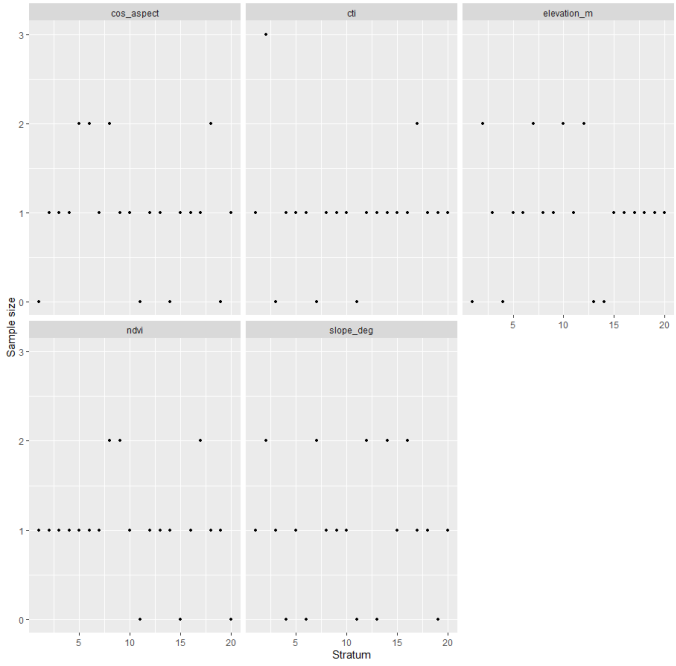
```
estimating jittering parameters from 'candi'...
low temperature: 26% of acceptance in the 1st chain
running time = 21.82 seconds
```

2) محاسبه تعداد نقاط در لایه‌های مرزی

```
by=1/samplesize
probs<-seq(from=0,to=1,by=by)
lb <- apply(grd[,3:7],MARGIN=2,FUN=function(x)
quantile(x,probs=probs,type=7))
lb <- lb[-length(probs),]
probs<-seq(from=0,to=1,length.out = samplesize + 1)
breaks <- apply(covars,MARGIN=2,FUN=function(x)
quantile(x,probs=probs,type=3))
mySample <- res$points
mySample <- data.frame(mySample,grd[mySample$id,3:7])
mySample <- data.frame(grd[res$points$id,1:7])
p<-ncol(covars)
stratum<-matrix(nrow=nrow(mySample),ncol=p)
for ( i in 1:p) {
  stratum[,i]<-findInterval(mySample[,i+4],lb[,i],left.open = TRUE)
}
counts <- lapply(1:ncol(covars), function (i)
  hist(mySample[, i+2], breaks[,i], plot = FALSE)$counts
)
counts<-matrix(nrow=nrow(lb),ncol=p)
for (i in 1:nrow(lb)) {
  counts[i,]<-apply(stratum, MARGIN=2, function(x,i) sum(x==i), i=i)
}
countslf <- data.frame(counts=unlist(counts))
countslf$covariate <- rep(names(covars),each=samplesize)
countslf$stratum<-rep(seq(1:samplesize),times=ncol(covars))
```

8) ترسیم حجم نمونه در لایه‌های مرزی (شکل 3-19)

```
ggplot(countslf) +
geom_point(mapping = aes(x=stratum,y = counts), colour = "black",size=1) +
  facet_wrap(~covariate) +
  scale_x_continuous(name = "Stratum") +
  scale_y_continuous(name = "Sample size",breaks=c(0,1,2,3))
```



شکل 3- 19- تعداد نقاط در لایه‌های مرزی

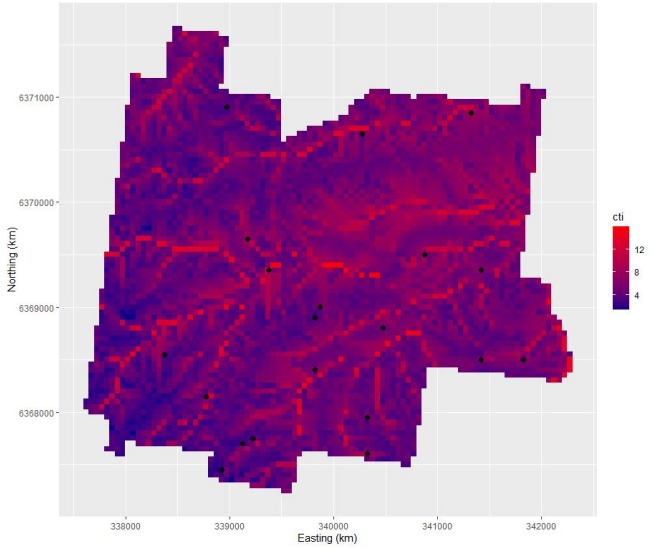
6) محاسبه معیار *O1* (به عنوان یک جمع، نه به عنوان میانگین)

```
(sum(abs(countslf$counts-1)))
```

7) ترسیم نمونه بهینه‌شده در فضای جغرافیایی

```
pdf(file = "cLHS_HunterValley.pdf", width = 7, height = 7)
ggplot(data=grd) +
geom_tile(mapping = aes(x = x1, y = x2, fill = cti))+      geom_point(data = mySample, mapping = aes(x = x1, y = x2), colour = "black",size=2) +
 scale_x_continuous(name = "Easting (km)") +
scale_y_continuous(name = "Northing (km)") +
scale_fill_gradient(name="cti",low = "darkblue", high = "red")+-
coord_fixed()
dev.off()
```

شکل 3-20 مثالی از روش نمونه‌برداری ابر مکعب لاتین مشروط برای 20 نقطه از منطقه مطالعه شده دره هانتر را نشان می‌دهد، با استفاده از همان پنج متغیر کمکی که در نمونه‌برداری به روش kmeans استفاده شدند.

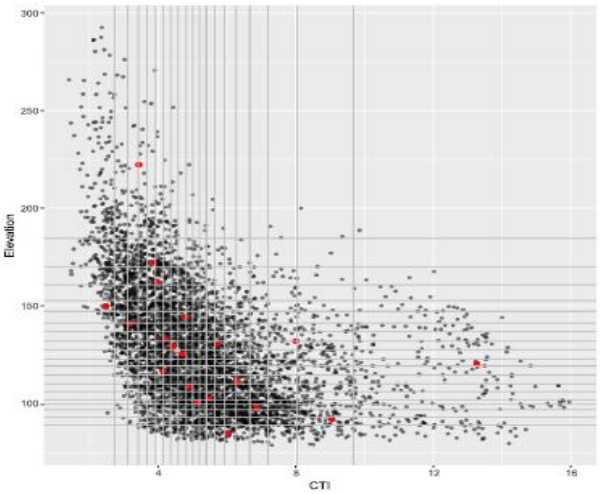


شکل 3- 20- نمونه ابر مکعب لاتین مشروط 20 نقطه‌ای در دره هانتر

8) ترسیم نمونه‌های بهینه‌شده در فضای متغیر با ساختن نمودارهای پراکندگی

```
pdf(file = "Scatterplot_cLHS_HunterValley.pdf", width = 7, height = 7)
ggplot(data=grd) +
geom_point(mapping = aes(y = elevation_m, x = cti), colour ="black",size=1,alpha=0.5) +
geom_point(data=mySample, mapping = aes(y = elevation_m, x = cti), colour = "red",size=2) +
geom_vline(xintercept=lb[-1,4],colour="grey")+
geom_hline(yintercept=lb[-1,1],colour="grey")+
scale_y_continuous(name = "Elevation") +
scale_x_continuous(name = "CTI")
dev.off()
```

در شکل 3-21، نمونه‌های حاصل از نمونه‌برداری ابر مکعب لاتین مشروط در یک نمودار پراکندگی ارتفاع در مقابل cti ترسیم شده است. افزون‌بر آن، لایه مرزی نشان داده شده است. در حالت ایده‌آل، هر ستون و هر سطر دارای یک نقطه نمونه‌برداری است.



شکل 3- 21- نمونه‌های ابر مکعب لاتین مشروط (نقاط کروی قرمز) در نمودار پراکندگی ارتفاع در برابر شاخص توپوگرافی مرکب رسم شده اند. خطوط عمودی و افقی خاکستری در این نمودار، فواصل لایه‌های مرزی هستند.

3-3-3-2- اجرای طرح نمونه‌برداری ابرمکعب لاتین مشروط با بسته نرم‌افزاری clhs

1) فراخوانی بسته‌های نرم‌افزاری مورد نیاز

```
library(sp)
library(lattice)
library(maptools)
library(rgdal)
library(raster)
library(clhs)
```

2)خواندن داده‌های متغیرهای کمکی دشت روانسر

```
Ravansar <- read.table("clhs/Ravansar.txt",header = T,sep = "\t")
str(Ravansar)
```

خلاصه داده‌ها

```
> str(Ravansar)
'data.frame':          71426 obs. of  9 variables:
 $ id                  : int  97 98 414 415 416 417 418 730 731 732 ...
 $ clipdem90_.no_sinks.    : num  1854 1886 1784 1809 1838 ...
 $ MRVBF               : num  8.78e-07 6.76e-07 2.62e-05 5.82e-06 2.10e-06
...
 $ Slope               : num  0.355 0.349 0.279 0.353 0.336 ...
 $ Topographic_Wetness_Index: num  5.98 5.51 7.7 6.45 5.72 ...
 $ grain_size_index        : num  -0.694 -0.682 -0.688 -0.697 -0.695 ...
 $ ndvi                : num  0.283 0.277 0.341 0.325 0.293 ...
 $ s1  : int  639480 639570 639300 639390 639480 639570 639660 639030
639120 639210 ...
 $ s2 : int  3862320 3862320 3862230 3862230 3862230 3862230 3862230
3862140 3862140 3862140 ...
```

انتخاب 20 نمونه به روش ابرمکعب لاتین مشروط در ناحیه مطالعه شده با استفاده از تابع زیر:

clhs (x, size, iter, temp, tdecrease, simple, progress,..)

که در آن

x: داده‌ها شامل متغیرهای کمکی به فرمت دیتا فریم یاSpatialPointsDataFrame یا Raster است که در این مثال از داده‌های با فرمت دیتا فریم استفاده شده است.

size: حجم یا تعداد نمونه مورد نظر است در این مثال تعداد 20 نمونه در نظر گرفته شده است.

progress: درحالت انتخاب TRUE نوار ابزار پیشرفت عملیات را در هنگام اجرای کد در بخش Rconsol نشان داده می‌شود و در حالت انتخاب FALSE، نوار پیشرفت نشان داده نمی‌شود.

simple: اگر این شناسه رویTRUE تنظیم شود، خروجی تنها شاخص‌های نمونه‌های انتخاب شده به شکل بردار عددی خواهد بود و اگر روی FALSE تنظیم شود، از حافظه بیشتری استفاده می‌کند اما اجازه می‌دهد تا از روش‌های cLHS_results

مانند plot.cLHS_result استفاده شود.

iter: یک عدد مثبت است، تعداد تکرارها برای فرآیند بازپخت با استفاده از الگوریتم Metropolis-Hastings نشان می‌دهد در صورت عدم انتخاب نرم‌افزار به صورت پیش فرض تعداد تکرار 10000 در نظر می‌گیرد.

3) اجرای مدل برای تولید 20 نمونه

```
sample20 <- clhs(Ravansar, size = 20, progress=FALSE, simple=FALSE,iter=100)
```

به طور خلاصه خروجی تابع به صورت فرمت لیستlist است. چون برای شناسه simple حالت FALSE در نظر گرفته شده است. در این لیست نمونه‌های انتخاب شده به صورت ردیف داده‌ها نشان داده می‌شود که در قسمتindex_sample مشاهده می‌شوند که بر اساس آن می‌توان موقعیت جغرافیایی آن‌ها را به دست آورد.

```
> str(sample20)
List of 5
 $ initial_object:'data.frame':       71426 obs. of  9 variables:
  ..$ id                  : int [1:71426] 97 98 414 415 416 417 418 730 731 732 ...
  ..$ clipdem90_.no_sinks.    : num [1:71426] 1854 1886 1784 1809 1838 ...
  ..$ MRVBF               : num [1:71426] 8.78e-07 6.76e-07 2.62e-05 5.82e-06
2.10e-06 ...
  ..$ Slope               : num [1:71426] 0.355 0.349 0.279 0.353 0.336 ...
  ..$ Topographic_Wetness_Index: num [1:71426] 5.98 5.51 7.7 6.45 5.72 ...
  ..$ grain_size_index        : num [1:71426] -0.694 -0.682 -0.688 -0.697 -0.695 ...
  ..$ ndvi                : num [1:71426] 0.283 0.277 0.341 0.325 0.293 ...
  ..$ s1                  : int [1:71426] 639480 639570 639300 639390 639480
639570 639660 639030 639120 639210 ...
  ..$ s2                  : int [1:71426] 3862320 3862320 3862230 3862230
3862230 3862230 3862230 3862140 3862140 3862140 ...
 $ index_samples : int [1:20] 863 61412 34070 4326 25723 68057 43666 54222
39993 14397 ...
 $ sampled_data  :'data.frame':     20 obs. of  9 variables:
  ..$ id                  : int [1:20] 8996 114146 82768 24989 71052 126242 93638
```

```
105380 89673 51447 ...
 ..$ clipdem90_.no_sinks.    : num [1:20] 1615 1317 1365 1724 1342 ...
 ..$ MRVBF                   : num [1:20] 0.2516 4.9896 1.446 0.0292 3.9059 ...
 ..$ Slope                   : num [1:20] 0.09678 0.00373 0.0677 0.1517 0.00616 ...
 ..$ Topographic_Wetness_Index: num [1:20] 9.46 14.96 10.07 6.15 15.33 ...
 ..$ grain_size_index        : num [1:20] -0.681 -0.687 -0.653 -0.658 -0.703 ...
 ..$ ndvi                    : num [1:20] 0.319 0.25 0.178 0.258 0.292 ...
 ..$ s1                      : int [1:20] 636510 654420 643980 640380 651810 652080
646140 640650 633810 638670 ...
 ..$ s2                      : int [1:20] 3859800 3830190 3839010 3855300 3842340
3826770 3835950 3832620 3837030 3847830 ...
 $ obj         : num [1:100] 135 135 129 129 129 ...
 $ cost        : NULL
 - attr(*, "class")= chr [1:2] "cLHS_result" "list"
```

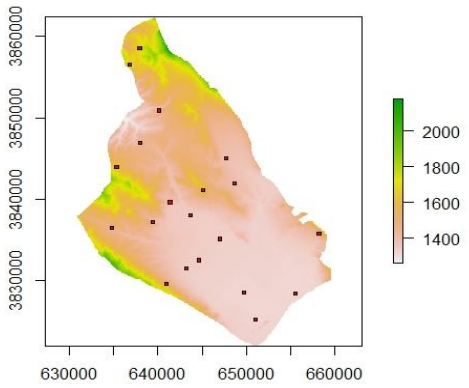
4) تعیین موقعیت جغرافیایی نمونه‌ها با کد زیر:

```
points<- Ravansar[sample20$index_samples,c("s1","s2")]
str(points)
```

```
'data.frame':          20 obs. of  2 variables:
 $ s1: int  655500 643260 640920 640110 634800 643710 658200 639390
649740 ...
 $ s2: int  3828480 3831540 3829650 3850890 3836490 3838020 3835770
3837210 3828570 3858540
```

5) در مرحله بعد با دستورات زیر، نمونه‌ها را بر روی یکی از نقشه‌های کمکی ارتفاع ناحیه مطالعه شده روانسر نشان داده می‌شوند (شکل 3-22).

```
coordinates(Ravansar) <- ~s1+s2
gridded(Ravansar) <- TRUE
fullgrid(Ravansar) <- TRUE
Ravansar <- stack(Ravansar)
plot(Ravansar[[1]])
points(Points, pch = 22, cex = 0.5, col = "black", bg = "red")
```



شکل 3-22- نمونه‌های حاصل از اجرای روش ابر مکعب لاتین با استفاده از نرم‌افزار clhs بر روی نقشه ارتفاع منطقه روانسر

نمونه‌برداری ابر مکعب لاتین مشروط یک طرح نمونه‌برداری بسیار مشهور در نقشه‌برداری رقومی خاک است. Mulder و همکاران (2013) این روش نمونه‌برداری را برای مناطقی که دسترسی به برخی از قسمت‌های آن (مانند مناطق کوهستانی و دور افتاده) دشوار است، مناسب دانستند.

3-3-4- نمونه‌برداری کنارد-استون

در آزمایشات گاهی تنها بخشی از فضای مشخصه‌ی متشکل از فاکتورها را می‌توان توسط نقاط طرح پوشش داد. برای حل این مشکل، طرح کنارد-استون(KS) از یک ماتریس محدود $N \times p$ از نقاط شروع می‌شود که فضای مشخصه فاکتور را با N تعداد نقاط انتخابی و p تعداد فاکتور مشخص می‌کند. برای انتخاب زیر مجموعه‌ای از n نقاط انتخابی که به عنوان نقاط طرح p استفاده می‌شوند، از یک معیار هندسی استفاده می‌شود. پاسخ برای ترکیبی از این فاکتورها در این نقاط طرح مشاهده می‌شود. انتخاب نقاط طرح به شرح زیراست. نخست دو نقطه انتخابی با حداکثر فاصله تفکیک در فضای مشخصه فاکتور انتخاب می‌شوند. سومین نقطه‌ای که ازگزینه $N - 2$ انتخاب می‌شود دارای بیشترین فاصله از دو نقطه طراحی اول است.

Kenard و Stone (1969) توصیه کردند که ابعاد فاکتورها، با مقیاس‌سازی آن‌ها یکنواخت‌سازی شود. آن‌ها همچنین پیشنهاد کردند که همبستگی عوامل را با تبدیل

فاکتورها، به متغیرهای متعامد و اندازه‌گیری فواصل در این فضای فاکتور تغییر یافته، در نظر بگیرند. این طرح معمولاً برای انتخاب یک زیرمجموعه از یک مجموعه با نمونه زیاد مانند داده‌های طیف‌سنجی (کتابخانه طیفی) برای واسنجی یک مدل در ارتباط با ویژگی خاک مورد نظر و طیف‌ها استفاده می‌شود، برای مثال به مقالات Viscarra Rossel و Brus (2018) و Riedel و همکاران (2018) مراجعه کنید. نمونه‌های KS را می‌توان با تابع ken.sto از بسته نرم‌افزاری prospectr در محیط R انتخاب کرد.

الگوریتم Kenard-Stone امکان انتخاب نمونه‌هایی با توزیع یکنواخت را در فضای پیش‌بینی کننده‌ها فراهم می‌کند. این کار با انتخاب جفت نقاطی که از یکدیگر فاصله دارند شروع می‌شود. سپس آن‌ها به مجموعه داده‌های واسنجی اختصاص داده و از لیست نقاط حذف می‌شوند. در ادامه‌ی اجرای این روش، با محاسبه فاصله بین هر نقطه غیراختصاص یافته و نقاط انتخاب شده براساس رابطه زیر و پیدا کردن نقطه‌ای که برای آن تعیین می‌شود، نقاط باقی مانده به مجموعه داده‌های واسنجی تنظیم شده اختصاص داده می‌شوند:

$$d_{selected} = \max_{i_a}\left(\min_{i}\left(d_{i,i_a}\right)\right) \qquad (10)$$

در این معادله از پایه نقطه‌ای انتخاب می‌شود که در داده‌های واسنجی دورترین فاصله از نزدیکترین همسایگان آن دارا است. این الگوریتم برای انتخاب نقاط از فاصله اقلیدسی استفاده می‌کند. با این حال، از فاصله ماهالانوبیس نیز می‌توان استفاده کرد. این را می‌توان با اجرای تجزیه به مولفه‌های اصلی PCA بر روی داده‌های ورودی و محاسبه فاصله با توجه به تعریف زیر از فاصله ماهالانوبیس به دست آورد:

$$H_{ij}^2 = \sum_{a=1}^{A} \frac{(\hat{t}_{ia} - \hat{t}_{ja})^2}{\hat{\lambda}_a} \qquad (11)$$

در اینجا a، $\hat{t}_{ia}$ امین نمره مولفه اصلی است، $\hat{t}_{ja}$ مقدار مولفه اصلی مربوط به نقطه j است، مقدار ویژه مولفه اصلی a است و A تعداد مولفه‌های اصلی داخل شده در محاسبه است.

مثال کاربردی تعیین نمونه‌ها به روش الگوریتم کنارد استون

1) فراخوانی بسته‌های نرم‌افزاری لازم به محیط R

```
library(sp)
library(fields)
```

```
library(ggplot2)
library(raster)
library(prospectr)
```

2) خواندن داده‌های متغیرهای کمکی دشت روانسر

```
grid <- read.table("km/covariates.txt",header = T)
```

3) بزرگ مقیاس‌سازی نقشه متغیرهای کمکی

شبکه رستری داده‌های کمکی دارای نقاط زیادی است. چون برای محاسبات در محیط R به مشکل حافظه رایانه برمی‌خوریم، از این‌رو یک زیر شبکه رستری انتخاب می‌شود و از نقشه‌های متغیر کمکی با اندازه پیکسل 90*90 متر با تابع ()aggregate به اندازه پیکسل 450*450 متر، بزرگ مقیاس‌سازی می‌شوند.

```
coordinates(grid) <- c("s1","s2")
gridded(grid) <- T
plot(grid)
grid
grid <- stack(grid)
subgrid <- aggregate(grid, fact=c(5,5), fun=mean, expand=T, na.rm=TRUE)
plot(subgrid)
subgrid <- as(subgrid,"SpatialGridDataFrame")
```

4) اجرای کد تجزیه به مولفه‌های اصلی بر روی داده‌های متغیرهای کمکی با تابع prcomp()

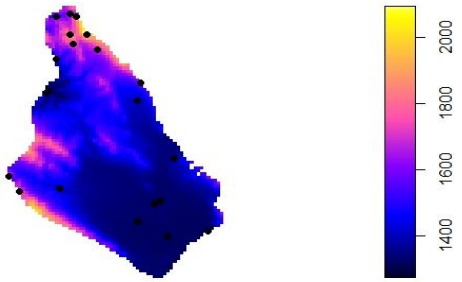
```
subgrid.pc <-
prcomp(~clipdem90_.no_sinks.+MRVBF+Slope+Topographic_Wetness_In
dex+grain_size_index+ndvi, scale=TRUE, subgrid)
str(subgrid.pc)
str(subgrid.pc$x)
biplot (subgrid.pc, arrow.len=0.1,xlabs=rep(".",
length(subgrid.pc$x[,1])),main="PCA biplot")
```

9) انتخاب 20 نمونه بر اساس روش کنارد استون با استفاده از تابع () kenStone و با استفاده از فاصله اقلیدسی

```
ken_euclid<- kenStone(X=subgrid.pc$x, k = 20, metric = "euclid")
samples <- subgrid[ken_mahal$model,c(7:8)]
```

6) ترسیم نقاط نمونه‌برداری بر روی نقشه ارتفاع منطقه مطالعه شده (شکل 3-23)

```
coordinates(subgrid) <- c("s1","s2")
gridded(subgrid) <- T
plot(subgrid)
points(samples, pch = 19, cex = 1, col = "black", bg = "grey")
```



شکل 3-23- نقاط نمونه‌برداری بر اساس روش کنارد استون بر روی نقشه ارتفاع منطقه مطالعه شده

Ramirez-Lopez و همکاران (2014) نمونه‌برداری ابر مکعب لاتین مشروط با نمونه‌برداری به روش kmeans فازی و کنارد-استون برای واسنجی مدل‌های پیش‌بینی میزانرس و غلظت کلسیم در مقیاس مزرعه‌ای و منطقه‌ای، با استفاده از داده‌های طیف سنجی خاک و روش بردار پشتیبان ماشین مقایسه کردند. نتایج این مطالعه نشان داد که خطای مدل بستگی به اندازه مجموعه داده‌های واسنجی دارد. وقتی تعداد نمونه‌های واسنجی به نسبت کم است، الگوریتم نمونه‌برداری ممکن است نقش مهمی در دقت مدل‌ها ایفا کند. از سوی دیگر، اگر اندازه مجموعه کالیبراسیون به اندازه کافی بزرگ باشد، روش نمونه‌گیری مسئله مهمی نیست. همچنین نتایج نشان داد که نمونه‌های واسنجی انتخاب شده از داده‌های طیف‌سنجی خاک با روش نمونه‌برداری ابر مکعب لاتین مشروط

و الگوریتم kmeans فازی در مقایسه با نمونه‌هایی انتخاب شده با الگوریتم کنارد -استون، توزیع اولیه داده‌های طیف سنجی خاک را بهتر انعکاس می‌دهند.

4- بحث و نتیجه‌گیری

از میان شیوه‌های مختلف نمونه‌برداری که در فصل سوم ارائه شد نمی‌توان گفت کدام یک از آن‌ها بهترین طرح نمونه‌برداری برای نقشه‌برداری رقومی خاک است. انتخاب بهترین طرح نمونه‌برداری به روش‌ها و مدل‌های زمین آماری و یادگیری ماشین بستگی دارد که برای پیش‌بینی ویژگی‌های خاک استفاده می‌شود. به طورکلی دو گزینه به منظور طراحی نمونه‌برداری برای پروژه‌های نقشه‌برداری رقومی خاک وجود دارد، یک گزینه که در آن یک یا چند متغیرکمکی داریم. در این حالت گزینه‌های نمونه‌برداری kmeans (فازی)، نمونه‌برداری ابر مکعب لاتین و روش کنارد-استون است (جدول 4-1) و تهیه نقشه با استفاده از روش‌های یادگیری ماشینی با یک یا چند متغیر کمکی انجام می‌گیرد.

گزینه دوم وضعیتی است که هیچ متغیر کمکی نداریم. در این وضعیت، ویژگی مورد نظر خاک باید با استفاده از برخی روش‌های درون‌یابی مکانی، مانند کریجینگ معمولی (OK) نقشه‌برداری می‌شود. برای درون‌یابی مکانی، نقاط نمونه‌برداری باید به طور یکنواختی در سراسر منطقه پخش شوند، که می‌توان با نمونه‌برداری در یک شبکه منظم یا روش بهتری مانند نمونه‌برداری با پوشش مکانی به این وضعیت مورد نظر دست یافت (جدول 4-1). به طور کلی برای انتخاب طرح نمونه‌برداری بر اساس وجود یا نبود متغیر کمکی می‌توان از جدول 4-1 به عنوان راهنما استفاده نمود.

جدول 4- 1- مروری بر روش‌های نقشه‌برداری و روش نمونه برداری مبتنی بر طرح متناسب با آن

آیا نقشه متغیرهای کمکی در دسترس است	روش نقشه‌برداری	طرح نمونه‌برداری
خیر	کریجینگ معمولی	شبکه‌ای منظم نمونه‌برداری پر مکانی/پوشش مکانی
بله	تکنیک‌های یادگیری ماشین	Kmeans ابرمکعب لاتین استون-کنارد

5- فهرست منابع

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): 1–44.

Brus, D.J. and Heuvelink, G.B.M. 2007. Optimization of sample patternsfor universal kriging of environmental variables. Geoderma,(138): 86–95.

Brus, D.J. and Noij, I.G.A.M. 2008. Designing sampling schemes for effect monitoring of nutrient leaching from agriculturalsoils. European Journal of Soil Science,(59): 292–303.

Brungard, C.W. and Boettinger, J.L. 2010. Conditioned Latin hypercubesampling: Optimal sample size for digital soil mappingof arid rangelands in Utah, USA. In Boettinger J L, HowellD W, Moore A C, Hartemink A E, Kienast-Brown S (eds.) Digital Soil Mapping. Springer, Dordrecht. Pp. 67–75.

Clifford, D., Payne, J.E., Pringle, M.J., Searle, R. and Butler, N. 2014. Pragmatic soil survey design using flexible Latin hypercube sampling. Computer Geoscience,(67): 62–68.

De Gruijter, J.J., Brus, D. J., Bierkens, M.F.P. and Knotters, M. 2006. *Sampling for Natural ResourceMonitoring.* Berlin: Springer-Verlag.

De Gruijter, J.J. and terBraak, C.J.F. 1990 Model-free estimation from spatial samples: a reappraisal of classical sampling theory. Mathematical Geology,(22):407-415.

Forgy, E.W. 1965. Cluster analysis of multivariate data: efficiency vs interpretability of classifications. Biometrics, (21):768--769.

Hartigan, J.A. and Wong, M.A. 1979. Algorithm AS 136: A K-means clustering algorithm. Applied Statistics, (28):100-108.

De Zorzi, P., Barbizzi, S., Belli, M., Fajgelj, A., Jacimovic, R., Jeran, Z., Sansone, U. and Van Der Perk, M. 2008. A soil sampling reference site: The challenge in defining reference material for sampling. Applied Radiation and Isotopes,(66): 1588–1591.

Hengl, T., Rossiter, D.G. and Stein, A. 2003. Soil sampling strategies forspatial prediction by correlation with auxiliary maps. Australian Journal of Soil Research,(41): 1403–1422.

Kennard, R. and Stone, L., 1969. Computer aided design of experiments. Technometrics, 11(1): 137–148.

Kidd, D., Malone, B., McBratney, A., Minasny, B. and Webb, M. 2015. Operational sampling challenges to digital soil mapping in Tasmania, Australia. Geoderma Regional, (4): 1–10.

Jenny, H. 1941. Factors of Soil Formation: A System of Quantitative Pedology. Dover Publications Inc., New York.

Lame, F.P.J. and Defize, P.R. 1993. Sampling of contaminated soil: Sampling error in relation to sample size and segregation. Environmental Science & Technology, (27): 2035–2044.

Lark, R.M. 2002. Optimized spatial sampling of soil for estimation of the variogram by maximum likelihood. Geoderma,(105): 49–80.

Lloyd, S.P. 1982. Least squares quantization in PCM. Technical Note, Bell Laboratories. Published in 1982 in IEEE Transactions on Information Theory,(28):128-137.

MacQueen, J. 1967. Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Eds L. M. Le Cam & J. Neyman, 1, Pp.281-297. Berkeley, CA: University of California Press.

Markert, B. 2007. Quality assurance of plant sampling and storage. In Quevauviller P (ed.) Qualilty Assurance in Environmental Monitoring: Sampling and Sample Preatreatment.Wiley-VCH Verlag GmbH, Weinheim.

McBratney, A.B., Odeh, I.O.A., Bishop, T.F.A., Dunbar, M.S and Shatar, T. M. 2000. An overview of pedometric techniques for use insoil survey. Geoderma,(97): 293–327.

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

McKay, M.D., Beckman, R.J. and Conover, W.J. 1979. Acomparison of three methods for selecting values of inputvariables in the analysis of output from a computer code. Technometrics,(21):239–245.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and Teller, E. 1953. Equation of state calculations by fast computing machines. Journal of Chemical Physics,(21):1087–1092.

Minasny, B. and McBratney, A.B. 2002. Uncertainty analysis forpedotransfer functions. European Journal of Soil Science,(53):417–430.

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

Minasny, B. and McBratney, A.B. 2010. Conditioned Latin Hypercube Sampling for Calibrating Soil Sensor Data to Soil Properties. In: Proximal Soil Sensing, Progress in Soil Science, pages 111-119.

Mulder, V., de Bruin, S. and Schaepman, M. 2013. Representing major soil variability at regional scale by constrained Latin hypercube sampling of remote sensing data. International Journal of Applied Earth Observation and Geoinformation,(21): 301–310.

Papritz, A. and Webster, R. 1995. Estimating temporal change in soil monitoring: I.Statistical theory. European Journal of Soil Science, (46): 1–12.

Pebesma, E.J. and Heuvelink, G.B.M. 1999. Latin hypercube samplingof Gaussian random fields. Technometrics, (41):303–312.

Ramirez-Lopez, L., Schmidt, K., Behrens, T., van Wesemael, B., Dematte, J. and Scholten, T., 2014. Sampling optimal calibration sets in soil infrared spectroscopy. Geoderma 226,140–150.

Roudier, P. 2011. Clhs: An R Package for Conditioned Latin Hypercube Sampling.Roudier, P., Beaudette, D., Hewitt, A., 2012. A Conditioned Latin Hypercube Sampling. Algorithm Incorporating Operational Constraints, Pp. 227–231.

Särndal, C.E., Swensson, B. and Wretman, J. 1992. Model-assisted Survey Sampling. (Springer: New York).

Scull. P., Franklin, J., Chadwick, O.A. and McArthur, D. 2003. Predictive soil mapping: A review. Progress in Physical Geography,(27):171-197

Papritz, A. and Webster, R. 1995. Estimating temporal change in soil monitoring: I. statistical theory. European Journal of Soil Science,(46):1-12.

Ramirez-Lopez, L., Schmidt, K., Behrens, T., van Wesemael, B., Dematte, J. and Scholten, T. 2014. Sampling optimal calibration sets in soil infrared spectroscopy. Geoderma,(226):140–150.

Riedel, F., Denk, M., Müller, I., Barth, N. and Gläßer, C., 2018. Prediction of soil parametersusing the spectral range between 350 and 15,000 nm: a case study based on thepermanent soil monitoring program in Saxony, Germany. Geoderma,(315):188–198.

Royle, J.A. and Nychka, D. 1998. An algorithm for the construction of spatial coverage designswith implementation in SPLUS. Computer Geoscience,(24): 479–488.

Schmidt, K., Behrens, T., Daumann, J., Ramirez-Lopez, L., Werban, U., Dietrich, P. and Scholten, T., 2014. A comparison of calibration sampling schemes at the field scale. Geoderma,(232): 243–256.

Theocharopoulos, S.P., Wagner, G., Sprengart, J., Mohr, M.E., Desaules, A., Muntau, H., Christou, M. and Quevauviller, P. 2001. European soil sampling guidelines for soil pollution studies. Science of the Total Environment,(264): 51–62.

Vasat, R., Boruvka, L. and Jaksık, O. 2012. Number of sampling pointsinfluences the parameters of soil properties spatial distributionand kriged maps. *In* Minasny B, Malone B P, McBratneyA B (eds.) Digital Soil Assessments and Beyond. CRC Press, London,Pp. 251–256.

Viscarra Rossel, R.A. and Brus, D.J. 2018. The cost-efficiency and reliability of two methodsfor soil organic c accounting. Land Degradation & Development, 29 (3):506–520.

Walvoort, D.J.J, Brus, D.J. and De Gruijter, J.J. 2010. An R package for spatial coverage sampling and random sampling from compact geographical strata by k-means. Comput Geoscience, (36): 1261–1267.