



نمونه‌برداری به روش اَبَر مکعب لاتین در منابع طبیعی در محیط نرم افزار R

* محسن حسینعلی‌زاده^۱، حسن احمدی^۲، سادات فیض‌نیا^۳ و فیروزه ریواز^۳

^۱ استادیار گروه آبخیزداری و مدیریت مناطق بیابانی، دانشگاه علوم کشاورزی و منابع طبیعی گرگان، آستاد گروه احیاء مناطق خشک و کوهستانی، دانشکده منابع طبیعی، دانشگاه تهران، آستادیار گروه آمار، دانشکده علوم ریاضی، دانشگاه شهید بهشتی
تاریخ دریافت: ۹۲/۴/۴؛ تاریخ پذیرش: ۹۳/۱۲/۱۴

چکیده

سابقه و هدف: نمونه‌برداری فرایندی است که هدف‌های متعدد از جمله پیش‌گویی در زمان یا مکان‌های بدون مشاهده را به دنبال دارد. بنابراین با توجه به محدودیت‌های مالی، زمانی و منابع در دسترس، راهبرد نمونه‌برداری مؤثر، یکی از مسائل هر متخصص در منابع طبیعی به‌شمار می‌رود. از این‌رو همواره متخصصان در جستجوی تکنیک‌های بهینه نمونه‌برداری هستند و یکی از راه‌های پیش‌روی آنان، استفاده از متغیرهای کمکی است. بنابراین متغیرهای کمکی به‌ویژه هنگامی که متغیرهای اولیه و اصلی محدود باشند، در بهبود پیش‌گویی‌ها از اهمیت زیادی برخوردارند. از این‌رو به‌جای نمونه‌برداری در فضای جغرافیایی که ممکن است معرف خوبی از دامنه تغییرات متغیر مورد مطالعه نباشد، بهتر است که در فضای حالت، نمونه‌برداری کرد. به‌طوری‌که در فضای حالت، متغیرهای کمکی تشکیل یک اَبَر مکعب را دهند. روش اَبَر مکعب لاتین یک روش نمونه‌برداری تصادفی طبقه‌بندی است که در آن نمونه موردنظر از توزیع چندمتغیره کمیت‌های مورد مطالعه استخراج می‌شود. در این شیوه، یک پوشش کامل از دامنه تغییرات هر یک از متغیرها بر اساس طبقه‌بندی حداکثری توزیع‌های حاشیه‌ای حاصل می‌شود. ضرورت این روش در منابع طبیعی، در نظر گرفتن مقادیر حدی در نقشه‌ها یا متغیرهای ورودی است که در نمونه‌برداری‌های مرسوم این مقادیر بنا به دلایلی در نظر گرفته نمی‌شود.

مواد و روش‌ها: در این پژوهش، به‌منظور رسیدن به حداکثر واریانس (تغییرپذیری حداکثر متغیرهای کمکی) و کاهش تعداد نمونه‌ها، روش اَبَر مکعب لاتین به‌عنوان یک راهبرد نمونه‌برداری با در نظر گرفتن داده‌های کمکی موجود (شیب‌های طولی و عرضی) در برآورد فرسایش‌پذیری خاک در زیرحوضه‌های زوجی کچیک استان گلستان مورد بحث قرار می‌گیرد. در واقع، موقعیت نمونه‌ها به‌گونه‌ای انتخاب می‌شوند که توزیع حاشیه‌ای متغیرها از حداکثر لایه‌بندی برخوردار باشند. در این راستا، نحوه اجرای این روش با تعداد نمونه‌ی معلوم و یکسان ۶۰ عدد (تعیین شده مبتنی بر یک روش دنباله‌ای) در هر یک از زیرحوضه‌های با شکل‌های متفاوت، اما مساحت تقریباً مساوی، در محیط نرم‌افزار R با استفاده از بسته تخصصی lhs ارائه می‌شود. به‌منظور اجرای این روش با استفاده از متغیرهای کمکی،

* مسئول مکاتبه: alizadeh_m2001@yahoo.com

نقشه نقطه‌ای این متغیرها با فاصله‌های ۲۰ متری تهیه شد. لازم به ذکر است که این نقشه‌ها در غالب توزیع آماری به‌عنوان ورودی الگوریتم آبرمکعب لاتین در نظر گرفته شدند. برای اجرای الگوریتم به‌منظور رسیدن به ۶۰ نمونه آبرمکعب لاتین، توزیع X_1 و X_2 مورد نیاز است. جهت برآورد پارامترهای توزیع‌ها (میانگین و واریانس)، روش حداکثر درستنمایی مورد استفاده قرار گرفت.

یافته‌ها: نتایج نشان داد که متوسط فاصله نمونه‌ها در زیرحوضه شاهد و نمونه به‌ترتیب برابر ۶۳۹ و ۶۹۳ متر بوده که دارای اختلاف ۵۴ متری نسبت به هم هستند. بیش‌ترین فراوانی فواصل نمونه‌برداری به فواصل ۴۰۰ تا ۶۰۰ متر در زیرحوضه نمونه و فواصل ۶۰۰ تا ۸۰۰ متر در زیرحوضه شاهد اختصاص دارد.

نتیجه‌گیری: این روش نمونه‌برداری از توزیع متغیرهای ورودی تبعیت می‌کند، بنابراین، مقادیر بحرانی و حدی (مقادیر حداقل و حداکثر) متغیرهای ورودی در این روش در نظر گرفته می‌شود. برخلاف روش‌های معمول که در آن‌ها واحدها یا پلی‌گون‌های با سطحی کمتر از سطح مشخص شده (کوچکترین واحد دارای قابلیت نقشه‌بندی) حذف می‌شوند، در این روش توزیع تجمعی متغیر ورودی در نظر گرفته شده و کوچکترین مساحت‌ها نیز در نمونه‌برداری لحاظ می‌شوند. بنابراین، این روش نمونه‌برداری در اکثر مطالعات منابع طبیعی از اهمیت زیادی برخوردار بوده و بیش‌ترین تغییرات را در مورد متغیرهای کمکی مرتبط با متغیر مورد مطالعه پوشش می‌دهد. بنابراین ارائه یک روش نمونه‌برداری طرح مینا در منابع طبیعی با تأکید بر در نظر گرفتن مقادیر حدی متغیرهای ورودی (نقشه‌ها) در نمونه‌برداری با تأکید نمونه‌برداری در فضای حالت (توزیع ورودی متغیرهای ورودی) بدون در نظر گرفتن توزیع جغرافیایی (مکانی) از نتایج اصلی تحقیق موجود به‌شمار می‌آید.

واژه‌های کلیدی: روش نمونه‌برداری آبرمکعب لاتین، فرسایش‌پذیری خاک، زیرحوضه‌های زوجی کوچک، نرم‌افزار R

مقدمه

نمونه‌برداری فرایندی است که هدف از اجرای آن، به‌دست آوردن داده برای مقاصدی مانند برآورد پارامترهای جامعه یا پیش‌گویی در زمان یا مکان‌های بدون مشاهده است. با توجه به محدودیت‌های مالی، زمانی و منابع در دسترس، راهبرد نمونه‌برداری مؤثر، یکی از مسائل هر متخصص در منابع طبیعی به‌شمار می‌رود. بنابراین همواره متخصصان در جستجوی تکنیک‌های بهینه نمونه‌برداری هستند. اکثر روش‌های نمونه‌برداری به‌گونه‌ای طراحی شده‌اند که بتوانند پوشش خوبی از محدوده مورد مطالعه فراهم کنند. به‌عنوان نمونه پژوهشگرانی مانند روپل و نیچکا

(۱۹۹۸)، روشی را مبتنی بر پوشش کامل محدوده مورد مطالعه پیشنهاد کردند، به‌گونه‌ای که موقعیت‌های نمونه‌ای مبتنی بر حداقل کردن یک تابع هدف که فقط تابعی از فاصله بین مکان‌های نمونه‌ای و غیرنمونه‌ای است، تعیین می‌شوند (۱۲). بروس و همکاران (۲۰۰۴)، الگویی را برای نمونه‌برداری پیشنهاد کردند که مبتنی بر تقسیم آبخیز مورد مطالعه به زیرحوضه‌ها است، به‌طوری‌که از هر زیرحوضه، یک نمونه به‌صورت تصادفی یا با هدف خاصی انتخاب می‌شود (۱). بنابراین این شیوه نمونه‌برداری یک پوشش جغرافیایی مناسب به‌ویژه برای زیرحوضه‌های با شکل‌های نامنظم فراهم می‌کند. اما آیا انتخاب شیوه

می‌شوند. در واقع در بسیاری از مسائل، نمونه‌برداری در یک محدوده را می‌توان با حضور این داده‌های کمکی بهبود بخشید. در مطالعات منابع طبیعی به‌منظور طراحی الگوی نمونه‌برداری می‌توان از تصاویر ماهواره‌ای، نقشه‌های پوشش گیاهی، نقشه‌های زمین‌شناسی، مدل‌های رقومی ارتفاع، کاربری اراضی و غیره استفاده کرد. این متغیرهای کمکی در منابع طبیعی از دیدگاه گوارتز و همکاران (۱۹۹۹)، اطلاعات ثانویه جامع نام‌گذاری شده است (۳). بنابراین متغیرهای کمکی به‌ویژه هنگامی که متغیرهای اولیه و اصلی محدود باشند، در بهبود پیش‌گویی‌های مکانی از اهمیت زیادی برخوردارند. استفاده از متغیرهای کمکی در نمونه‌برداری به این صورت است که در این روش سعی می‌شود پوشش کاملی از دامنه مقادیر هر یک از متغیرهای کمکی در طراحی شبکه نمونه‌برداری در نظر گرفته شود. بنابراین سؤال اصلی، نحوه انتخاب مکان نمونه‌هاست. از این‌رو به‌جای نمونه‌برداری در فضای جغرافیایی^۱ (موقعیت‌های مختلف یک آبخیز) که ممکن است پراکنش آن‌ها به‌گونه‌ای باشد که معرف خوبی از دامنه تغییرات متغیر مورد مطالعه نباشد، بهتر است که در فضای حالت^۲، نمونه‌برداری کرد که پوشش خوبی از دامنه تغییرات متغیرهای کمکی را فراهم می‌سازد، به‌طوری‌که در فضای حالت متغیرهای کمکی، تشکیل یک ابرمکعب^۳ را دهند. پس روش نمونه‌برداری باید به‌گونه‌ای باشد که نمونه‌ها کل فضای حالت متغیرهای کمکی را پوشش دهند. از این‌رو، روش ابرمکعب لاتین^۴ به‌عنوان یک راهبرد نمونه‌برداری برای مرتفع کردن این مشکل پیشنهاد می‌شود. در واقع، این روش یک

نمونه‌برداری به‌گونه‌ای که مقید به در نظر گرفتن کل آبخیز باشد، ضرورت دارد؟ مک‌کنزی و رایان (۱۹۹۹)، برای تعیین تغییرپذیری متغیر مورد مطالعه در محدوده نمونه‌برداری، با استفاده از خصوصیات زمینی، داده‌های زمین‌شناسی و اقلیمی، محدوده مورد بررسی را به طبقاتی تقسیم کردند و سپس واحدهای نمونه‌برداری را به‌صورت تصادفی از هر طبقه انتخاب کردند (۸) (این مورد نظیر همان واحدهای کاری است که در حال حاضر در کشور در حال انجام است). مک‌برتنی و همکاران (۱۹۹۹)، روشی را پیشنهاد کردند که در نواحی با تغییرات اندک متغیر مورد مطالعه، تعداد نمونه کم‌تری برداشته شده و در نواحی با تغییرات زیاد، تعداد نمونه بیشتر باشد (۶). هنگل و همکاران (۲۰۰۳)، بر اساس متغیرهای محیطی، روشی را پیشنهاد کردند که واریانس کل در آن حداکثر باشد، یعنی تغییرپذیری متغیر مورد مطالعه در نمونه برداشت شده، حداکثر باشد (تمامی مقادیر ممکن متغیر مربوطه را شامل شود) (۴). گورجتز و همکاران (۲۰۰۶)، طرحی را برای پیش‌گویی مکانی با استفاده از کریجینگ پیشنهاد کردند. این روش مبتنی بر حداقل کردن واریانس پیش‌گویی کریجینگ عمومی است که لازم است در آن مدل روند و باقی‌مانده‌ها معلوم باشد، اما در عمل معمولاً این اطلاعات در دسترس نیستند. به‌عبارت دیگر، آن‌ها واریانس تخمین کم‌تر را لحاظ قرار داده و در صورت وجود روند در داده‌ها، مدل‌سازی و حذف آن و همچنین تأکید بر باقی‌مانده‌ها را مهم دانسته‌اند (۵). بنابراین یکی از راه‌های پیش‌روی پژوهشگران در عمل، استفاده از متغیرهای کمکی است. به‌عنوان مثال در منابع طبیعی، همه نقشه‌های دخیل در تهیه نقشه واحدهای کاری به‌عنوان متغیرهای کمکی (ثانویه) در نظر گرفته

- 1- Space
- 2- Feature Space
- 3- Hypercube
- 4- Latin Hypercube Sampling

ضمن معرفی روش اَبَرَمکعب لاتین، نحوه تعیین مکان نمونه‌ها با این روش ارائه خواهد شد. در بخش ۳، اجرای این روش در زیرحوضه‌های زوجی کوچک استان گلستان با استفاده از دو متغیر کمکی نشان داده می‌شود. در نهایت، بحث و نتیجه‌گیری نیز در بخش ۴ ارائه خواهد شد.

مواد و روش‌ها

اَبَرَمکعب لاتین و مقایسه آن با سایر روش‌های نمونه‌برداری: اَبَرَمکعب لاتین (LHS)، یک روش نمونه‌برداری تصادفی طبقه‌بندی شده^۲ است که در آن، نمونه موردنظر به تعداد n از متغیرهای چندگانه به نحوی انتخاب می‌شوند که نمونه مرتبط با هر یک از متغیرها، از حداکثر طبقه‌بندی^۳ برخوردار باشد. در واقع حداکثر طبقه‌بندی زمانی اتفاق می‌افتد که تعداد لایه‌ها (طبقه‌ها) برابر با تعداد نمونه‌ها (n) در نظر گرفته شده و احتمال قرار گرفتن نمونه درون هر لایه یا طبقه برابر با معکوس تعداد نمونه کل ($1/n$) باشد. برای درک بیشتر این روش، k متغیر $X=(X_1, X_2, \dots, X_k)$ را در نظر بگیرید. ابتدا دامنه هر متغیر به n فاصله هم‌احتمال تقسیم شده و برای هر متغیر، یک نمونه تصادفی از هر فاصله یا طبقه انتخاب می‌شود. سپس مقادیر به‌دست آمده برای هر متغیر با یکدیگر به صورت تصادفی یا بر اساس یک قاعده خاص ترکیب شده و تشکیل n نمونه k -تایی می‌دهند که تمامی لایه‌ها را در بر گیرند. بنابراین، این روش نمونه‌برداری در حالتی که تعداد متغیرها حتی از k هم بیشتر شود، به تعداد نمونه بیشتر نیازی ندارد. این روش تضمین می‌کند که دامنه مقادیر متغیرهای موردنظر به روش حداکثر لایه‌بندی شده است.

نمونه‌برداری تصادفی طبقه‌بندی است که از توزیع متغیرها برای ایجاد طبقات استفاده می‌کند و اولین بار توسط مک‌کی و همکاران (۱۹۷۹)، برای شبیه‌سازی مونت‌کارلو مورد استفاده قرار گرفت (۷)، سپس پیسما و هویلینک (۱۹۹۹) و ژانگ و پیندر (۲۰۰۴)، از آن برای شبیه‌سازی میدان‌های تصادفی گاوسی استفاده کردند (۱۱ و ۱۴). میناسنی و مک‌برتنی (۲۰۰۶)، از آن برای ارزیابی عدم قطعیت در مدل پیش‌گویی خاک استفاده کردند (۱۰). ژو و همکاران (۲۰۰۵)، به تلفیق روش اَبَرَمکعب لاتین و شبیه‌سازی تصادفی در یک مدل شبیه‌سازی چشم‌انداز جنگلی پرداخته و نشان دادند که این روش نسبت به نمونه‌برداری تصادفی ساده^۱ (SRS) به‌ویژه در تعداد کم شبیه‌سازی، توانایی نشان‌دادن تغییرپذیری بیشتری در فضای نمونه را دارد (۱۳). البته ذکر این نکته ضروری است که در حال حاضر در مطالعات منابع طبیعی (واحدهای کاری همگن)، متغیرهای کمکی یا همان نقشه‌های پایه مورد استفاده در تهیه نقشه واحدهای کاری همگن به‌نوعی در طراحی شبکه نمونه‌برداری در نظر گرفته می‌شوند. اما در این مطالعه سعی بر آنست تا به‌جای لحاظ کردن نقشه‌ها به‌عنوان نقشه ورودی واحدهای کاری، توزیع آماری متغیرها (نقشه‌ها) لحاظ شده تا مقادیر حدی، حتی با سطح خیلی کم نیز در الگوی نمونه‌برداری در نظر گرفته شوند.

با توجه به حضور متغیرهای کمکی فراوان، به‌ویژه نقشه‌های توپوگرافی و مشتقات ناشی از مدل رقومی زمین، تصاویر ماهواره‌ای، همچنین داشتن نمونه‌هایی در فضای حالت به‌جای فضای جغرافیایی و اهمیت نرم‌افزار R در مطالعات آماری، هدف از این مطالعه معرفی و اجرای روش اَبَرَمکعب لاتین در محیط نرم‌افزار آماری R به‌منظور بهبود روش نمونه‌برداری در مطالعات منابع طبیعی است. بنابراین در بخش ۲،

2- Stratified Random Sampling
3- Marginally Maximally Stratified

1- Simple Random Sampling

شده و به نمونه‌هایی از توزیع متغیر دوم برگردانده می‌شوند. در نهایت، این نمونه‌ها به صورت تصادفی با یکدیگر جفت شده و نمونه موردنظر را می‌سازند (شکل ۱).

لازم به ذکر است که نمونه‌های اخذ شده با روش آبرمکعب لاتین به علت لایه‌بندی کامل دامنه مقادیر متغیرهای ورودی، در تهیه نمودارهای پراکنش^۱ بسیار کارا عمل می‌کنند. هم‌چنین نمونه‌های حاصل از این روش در تحلیل حساسیت حتی با تعداد نمونه کم، عملکرد مناسبی دارند. روش آبرمکعب لاتین نسبت به روش مونت‌کارلو به تعداد نمونه کم‌تری برای برآورد توزیع متغیر مورد مطالعه نیاز دارد و به‌ویژه در توزیع‌های نامتقارن، توزیع را دقیق‌تر برآورد می‌کند. این روش در مقایسه با روش نمونه‌برداری تصادفی ساده، انحراف معیار بیش‌تری را با تعداد نمونه یکسان پوشش می‌دهد که این نکته نشان‌دهنده این است که در این روش، با تعداد نمونه کم‌تری می‌توان تغییرپذیری بیش‌تری را نسبت به نمونه‌برداری تصادفی ساده نشان داد (۱۳). از طرفی، روش آبرمکعب لاتین بسیاری از ویژگی‌های مطلوب روش‌های نمونه‌برداری طبقه‌بندی^۲ و تصادفی^۳ را دارا بوده و مانند آن‌ها یک نمونه‌برداری احتمالی به حساب می‌آید.

برای استفاده از این روش، ابتدا باید توزیع‌های حاشیه‌ای X_1, X_2, \dots, X_k به n فاصله هم‌احتمال تقسیم شده، سپس از هر فاصله، نمونه‌ای به تصادف انتخاب شود. برای انجام این کار از قضیه تبدیل انتگرال احتمال استفاده می‌شود. در واقع، ابتدا فاصله $[0,1]$ را به n بازه به طول $1/n$ تقسیم کرده، سپس نمونه‌ای تصادفی از هر یک از بازه‌ها انتخاب می‌شود. در مرحله بعد، با عکس تبدیل، نمونه‌های اخذ شده به نمونه‌هایی از توزیع اصلی متغیرها برگردانده می‌شوند. الگوریتم زیر مراحل تعیین نمونه‌ها با روش آبرمکعب لاتین را نشان می‌دهد.

۱- $j=1$ در نظر گرفته شود.

۲- نمونه‌ای به صورت تصادفی از توزیع $U [0,1]$ انتخاب و u_{ij} نام‌گذاری شود. $w_{ij}=(u_{ij}+(i-1))/n$ تعیین شود.

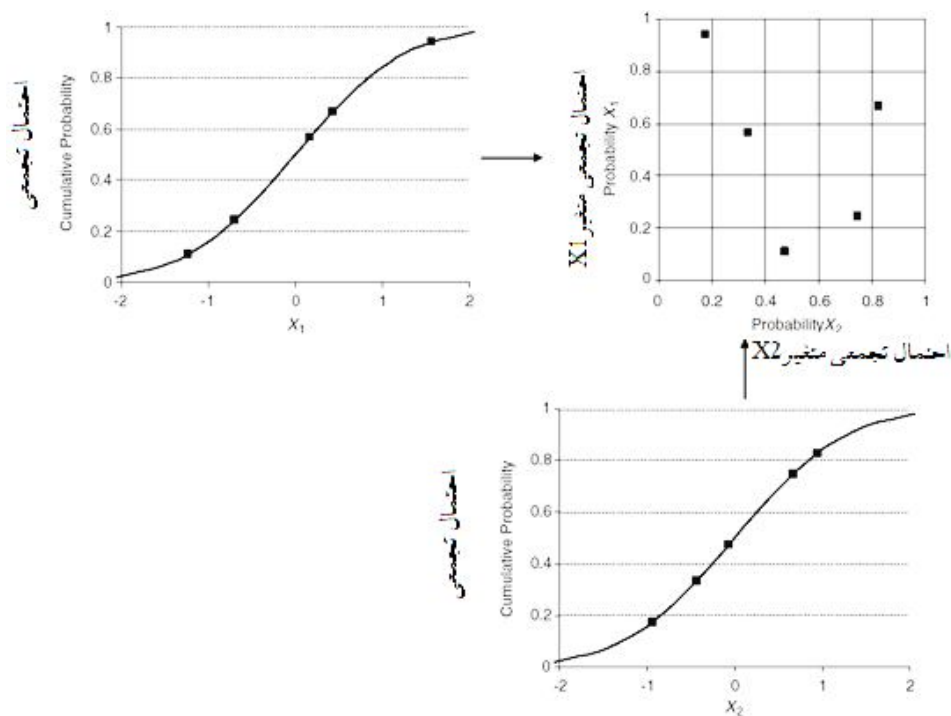
۳- i امین عضو نمونه برای متغیر j ام به صورت $x_{ij}=F_j^{-1}(w_{ij})$ حاصل می‌شود.

۴- $i=j+1$ قرار داده شود و الگوریتم به مرحله ۲ برگردد.

با تکرار این الگوریتم تا مرحله $j=k, i$ امین عضو نمونه، $i=1, \dots, n$ ، از هر یک از متغیرها حاصل می‌شود. سپس با تشکیل k تایی‌ها به صورت تصادفی یا با قاعده‌ای خاص، یک نمونه آبرمکعب به تعداد n از k متغیر به دست می‌آید (مثال ۱).

مثال ۱. فرض کنید دو متغیر ورودی با توزیع نرمال داریم که هدف: اخذ یک نمونه ۵ تایی آبرمکعب لاتین از آن‌هاست. برای این منظور ابتدا ۵ نمونه از توزیع یکنواخت روی بازه $[0,1]$ تولید کرده و با عکس تبدیل به نمونه‌هایی از توزیع متغیر اول تبدیل می‌شوند. سپس ۵ نمونه دیگر از $U [0,1]$ انتخاب

1- Scatter plot
2- Stratified Sampling
3- Random Sampling



شکل ۱- آبرمکعب لاتین با ۵ نمونه برای دو متغیر با توزیع نرمال.

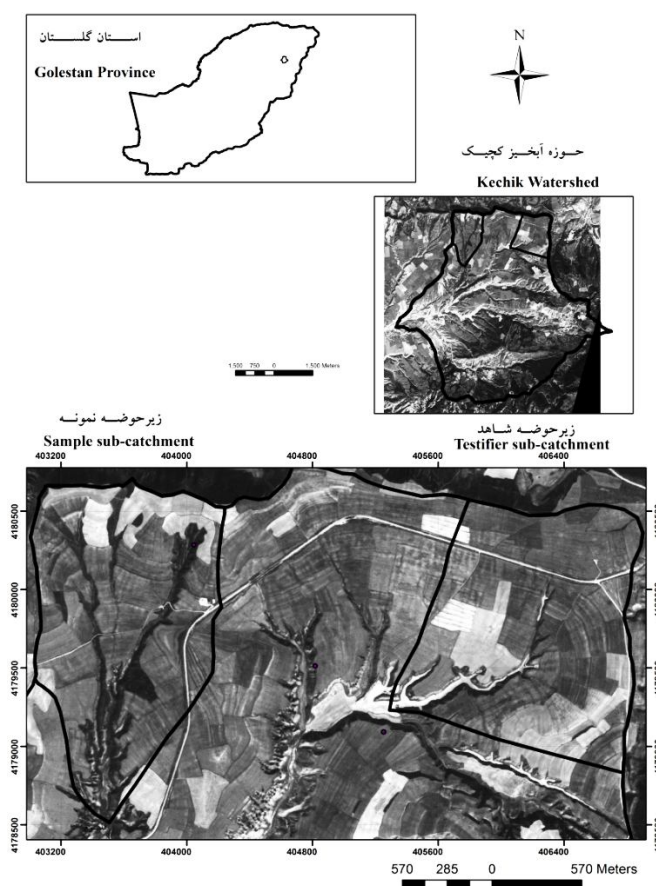
Figure 1. Latin Hypercube Sampling for 5 samples of two variables with normal distribution.

این روش در دو مطالعه موردی، آبخیز زوجی کچیک به‌عنوان یکی از حوضه‌های زوجی مهم کشور، مورد مطالعه قرار گرفت. این حوضه یکی از زیرحوضه‌های قرناوه (آبخیز گرگان‌رود) محسوب می‌شود. آبخیز کچیک با حداقل ارتفاع ۶۲۰ متر و حداکثر ارتفاع ۹۰۰ متر از سطح دریا در بین طول‌های جغرافیایی "۱۰' ۵۲" ۵۵° تا "۲۵' ۵۷" ۵۵° و عرض‌های جغرافیایی "۱۵' ۴۲" ۳۷° تا "۲۵' ۴۶" ۳۷° قرار گرفته است. این حوضه با توجه به وضعیت توپوگرافی و شبکه زهکشی و همچنین خطوط تقسیم آب از سه واحد هیدرولوژیکی تشکیل شده که زیرحوضه‌های مورد مطالعه در یک واحد هیدرولوژیکی قرار گرفته‌اند. مساحت زیرحوضه شاهد^۱ و نمونه در این مطالعه برابر ۱۹۰ هکتار است (شکل ۲).

همان‌طور که مشاهده شد با توجه به این‌که در این روش، به تعیین طبقات و احتمال‌های مربوطه نیازی نیست، اجرای این روش مانند روش نمونه‌برداری تصادفی ساده، نسبت به روش نمونه‌برداری طبقه‌بندی‌شده، ساده‌تر است. در حقیقت، نمونه‌برداری طبقه‌بندی‌شده زمانی مفید است که اطلاعات کافی در مورد طبقه‌ها و احتمال‌های آن موجود باشد. روش آبرمکعب لاتین به‌ویژه در نمونه‌برداری از توزیع‌های چندمتغیره، یک تکنیک مؤثر محسوب شده (۱۰) و کارایی آن نسبت به روش تصادفی ساده برای نمونه‌برداری از توزیع‌های تک‌متغیره بیش‌تر است (۷).

محدوده مورد مطالعه: برای اجرای روش آبرمکعب لاتین، به محدوده‌ای نیاز است که در آن اطلاعاتی راجع به کمیت‌های مورد مطالعه یا متغیرهای کمکی موجود باشد. در این مطالعه به‌علت در نظر گرفتن شکل حوضه در الگوی نمونه‌برداری و همچنین تکرار

۱- Testifier Sub-catchment: جنبه ترویجی و آموزشی داشته و عملیات آبخیزداری را شامل می‌شود.



شکل ۲- موقعیت محدوده مورد مطالعه.

Figure 2. Location of the study area.

مطالعه است. سپس نقشه توپوگرافی با مقیاس ۱:۲۵۰۰۰ تهیه شده و مدل رقومی زمین از آن استخراج شد. مشتقات ثانویه مدل رقومی زمین اعم از نقشه‌های شیب عرضی و طولی به منظور در نظر گرفتن تأثیر این متغیرهای کمکی در برآورد فرسایش پذیری خاک، نیز تهیه شد. به منظور اجرای روش آبرمکعب لاتین با استفاده از این متغیرها به عنوان متغیرهای کمکی، نقشه نقطه‌ای این متغیرها (شیب‌های طولی و عرضی) با فاصله‌های ۲۰ متری تهیه شد. لازم به ذکر است که این نقشه‌ها در غالب توزیع آماری به عنوان ورودی الگوریتم آبرمکعب لاتین در نظر گرفته شدند. برای اجرای الگوریتم به منظور رسیدن به ۶۰ نمونه آبرمکعب لاتین، توزیع X_1 و X_2 مورد نیاز است. در این خصوص، ابتدا با استفاده از نرم‌افزار R توزیع

اجرای روش نمونه‌برداری آبرمکعب لاتین در محدوده مورد مطالعه: به علت مهم بودن عامل فرسایش‌پذیری خاک و دخیل بودن آن در مدل‌های متعدد فرسایش و رسوب، سعی شد در این مطالعه، طراحی شبکه نمونه‌برداری آبرمکعب لاتین برای این متغیر با استفاده از متغیرهای کمکی شیب‌های طولی^۱ (انحنای قائم) و عرضی^۲ (انحنای عرضی) به منظور رسیدن به حداکثر واریانس (تغییرپذیری حداکثر متغیرهای کمکی) و کاهش تعداد نمونه‌ها صورت گیرد. برای این منظور ابتدا تعداد نمونه لازم در هر زیرحوضه با استفاده از یک روش دنباله‌ای، ۶۰ عدد تعیین شدند که تعیین تعداد نمونه خارج از بحث این

1- Profile Curvature

2- Plan Curvature

این الگوریتم برای هر یک از زیرحوضه‌ها، نمونه آبرمکعب لاتین به دست آمد.

نتایج

جدول ۱، آماره‌های خلاصه مربوط به شیب‌های طولی (X_1) و عرضی (X_2) زیرحوضه‌ها را نشان می‌دهد. همان‌طور که مشاهده می‌شود متوسط شیب طولی و عرضی در هر دو محدوده مورد مطالعه تقریباً با یکدیگر برابر هستند.

متغیرها در هر دو زیرحوضه برآورد شدند. لازم به ذکر است گرچه فرم توزیع‌ها مشخص شدند اما پارامترهای آن‌ها یعنی میانگین و واریانس نامعلومند که باید برآورد شوند. برای این منظور، با به‌کارگیری روش حداکثر درست‌نمایی پارامترهای موردنظر برآورد شدند. بر این اساس، ابتدا با استفاده از بسته تخصصی lhs در نرم‌افزار R (۲) نمونه‌های آبرمکعب لاتین از توزیع $U [0,1]$ تولید کرده و با عکس تبدیل به نمونه‌هایی از توزیع X_1 و X_2 تبدیل شدند. با تکرار

جدول ۱- خصوصیات آماری محدوده‌های مورد مطالعه و متغیرهای کمکی مورد استفاده در آبرمکعب لاتین.

Table 1. Statistical properties of study areas and covariables in Latin Hypercube Sampling.

زیرحوضه شاهد Testifier Sub-catchment		زیرحوضه نمونه Sample Sub-catchment		خصوصیات آماری Statistical Properties
شیب عرضی Plan Curvature	شیب طولی Profile Curvature	شیب عرضی Plan Curvature	شیب طولی Profile Curvature	
0.00	0.01	0.00	0.00	متوسط Mean
0.09	0.12	0.16	0.19	انحراف معیار SD
0.61	0.78	1.25	1.29	حداکثر Maximum
-0.41	-0.97	-1.45	-1.28	حداقل Minimum
3140		3059		تعداد نمونه Sample Number
23		27		شیب متوسط (%) Average Slope (%)
196		191		مساحت (ha) Area (ha)

جدول ۲ برآورد تابع چگالی دو متغیر شیب طولی و عرضی را در هر دو زیرحوضه مورد مطالعه نشان می‌دهد.

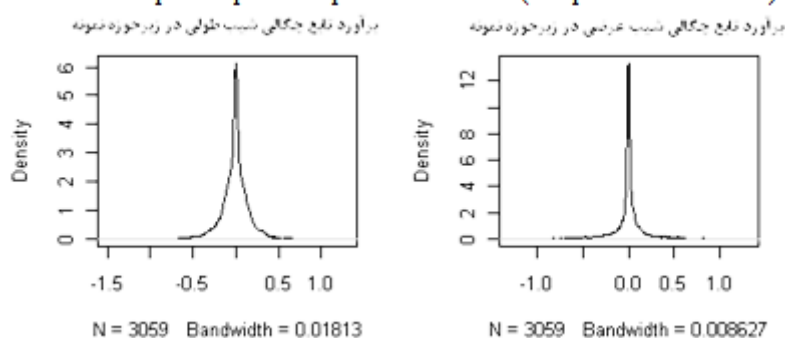
شکل ۳ نمودار توابع چگالی برآورد شده را نشان می‌دهد. همان‌طور که مشاهده می‌شود توزیع هر دو متغیر در هر زیرحوضه، توزیع نرمال است.

جدول ۲- توزیع‌های برآورد شده متغیرهای کمکی در هر دو زیرحوضه.

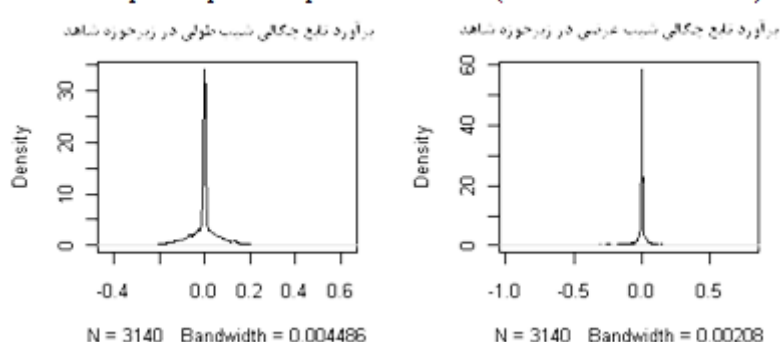
Table 2. Estimated Distributions for covariables in both sub-catchments.

برآورد تابع چگالی X_2 pdf estimation for X_2	برآورد تابع چگالی X_1 pdf estimation for X_1	
N(-0.006, 0.038)	N(-0.005, 0.025)	زیرحوضه نمونه Sample Sub-catchment
N(-0.006, 0.015)	N(0.0004, 0.008)	زیرحوضه شاهد Testifier Sub-catchment

Estimated pdf for plan and profile curvature (Sample sub-catchment)



Estimated pdf for plan and profile curvature (Testifier sub-catchment)

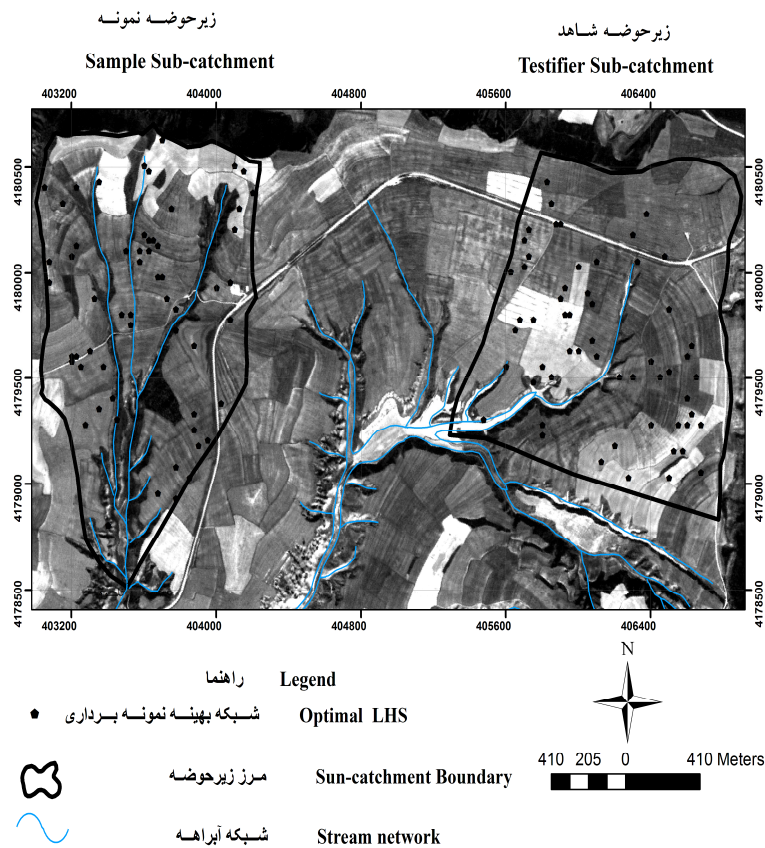


شکل ۳- نمودار توابع چگالی متغیرهای X_1 و X_2 در هر دو زیرحوضه.

Figure 3. Pdf curves of X_1 and X_2 in both sub-catchments.

متغیرها زیاد است در حالی که بعضاً در ناحیه‌ای بزرگ، تنها یک یا دو نمونه انتخاب شد. در این روش نمونه‌برداری، حداکثر تغییرپذیری در متغیرهای ورودی در نظر گرفته شده و مقادیر حدی ارزش خود را حفظ کرده‌اند. برخی از ویژگی‌های آماری فواصل نمونه‌ها در روش آبرمکعب لاتین در جدول ۳ آورده شده است.

شکل ۴، موقعیت نمونه‌ها را در هر یک از زیرحوضه‌ها نشان می‌دهد. همان‌طور که مشاهده می‌شود طرح نمونه‌برداری به لحاظ پراکندگی جغرافیایی مناسب به نظر نمی‌رسد. زیرا در این طرح سعی بر آنست تا پوشش خوبی از دامنه مقادیر دو متغیر شیب طولی و عرضی فراهم شود. بنابراین در برخی نواحی، تعداد نمونه‌ها به دلیل تغییرات مقادیر



شکل ۴- الگوی بهینه نمونه برداری آیرمکعب لاتین در زیرحوضه های شاهد و نمونه.

Figure 4. Optimum pattern of LHS in Sample and Testifier Sub-catchment.

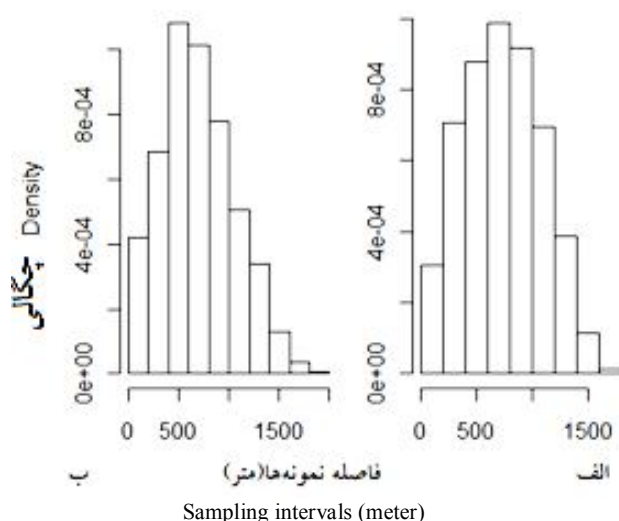
جدول ۳- آماره های توصیفی فواصل نمونه ها (متر) در شبکه نمونه برداری حاصل از روش آیرمکعب لاتین.

Table 3. Descriptive statistics samples intervals (m) in LHS.

حداکثر Maximum	متوسط Mean	حداقل Minimum	آماره های توصیفی Descriptive statistics زیرحوضه Sub-catchment
1849	693	20	نمونه Sample
1616	639	20	شاهد Testifier

متر در زیرحوضه نمونه و فواصل ۶۰۰ تا ۸۰۰ متر در زیرحوضه شاهد اختصاص دارد.

هیستوگرام فاصله های نمونه برداری در هر دو زیرحوضه در شکل ۵ آورده شده است. بیشترین فراوانی فواصل نمونه برداری به فواصل ۴۰۰ تا ۶۰۰



شکل ۵- هیستوگرام‌های فاصله‌های نمونه‌برداری در روش نمونه‌برداری آب‌مکعب لاتین در زیرحوضه نمونه (الف) و شاهد (ب).
Figure 5. Histogram of sampling intervals in LHS A). Sample sub-catchment and B). Testifier sub-catchment.

روش‌های معمول که در آن‌ها واحدها یا پلی‌گون‌های با سطحی کم‌تر از سطح مشخص شده (کوچک‌ترین واحد دارای قابلیت نقشه‌بندی^۱) حذف می‌شوند، در این روش توزیع تجمعی متغیر ورودی در نظر گرفته شده و کوچک‌ترین مساحت‌ها نیز در نمونه‌برداری لحاظ می‌شوند. از طرفی ممکن است این پلی‌گون‌های حذف شده هر چند با سطح محدود، ولی دارای تأثیر بسیار زیادی در طراحی شبکه نمونه‌برداری باشند (یک دامنه مستعد به لغزش، کاربری اراضی خاص، یک کارخانه آلاینده و غیره). به عبارت دیگر ممکن است این پلی‌گون‌های با مساحت کم، در محدوده مورد مطالعه دارای فراوانی بسیار زیادی بوده، ولی به علت فاصله زیاد از یکدیگر و هم‌چنین اجرای برخی از توابع و فرمان‌ها در محیط سیستم اطلاعات جغرافیایی، نادیده گرفته شده و حذف شوند. در صورتی که این محدودیت در روش آب‌مکعب لاتین، برطرف شده و به علت در نظر گرفتن توزیع‌های متغیرهای ورودی، تمامی مقادیر هر متغیر به منظور نمونه‌برداری در نظر گرفته می‌شوند. قابل ذکر است

بحث و نتیجه‌گیری

چون این الگوریتم بر نمونه‌برداری از فضای متغیرهای کمکی نسبت به فضای جغرافیایی تأکید دارد، طرح نمونه‌برداری حاصل نیز در فضای جغرافیایی مورد مطالعه از نظم خاصی تبعیت نمی‌کند. به عنوان مثال در زیرحوضه نمونه، در مکان‌هایی که شیب‌های طولی و عرضی بیش‌ترین تغییرات را دارند، بیش‌ترین تراکم نمونه‌برداری دیده می‌شود و بالعکس. این مسأله نیز در مورد زیرحوضه شاهد صدق می‌کند. این نکته قابل ذکر است که در تکنیک‌های رایج نمونه‌برداری در منابع طبیعی، نقشه‌های متغیرهای پیوسته (شیب، جهت، ارتفاع و غیره) که به منظور نمونه‌برداری مورد استفاده قرار می‌گیرند، طبقه‌بندی شده که این نوع طبقه‌بندی سلیقه‌ای بوده و از نظم خاصی تبعیت نکرده و خطاها و عدم قطعیت‌هایی را نیز به دنبال دارد.

در ضمن، این روش نمونه‌برداری از توزیع متغیرهای ورودی تبعیت می‌کند، بنابراین، مقادیر بحرانی و حدی (مقادیر حداقل و حداکثر) متغیرهای ورودی در این روش در نظر گرفته می‌شود. بر خلاف

1- Minimum Mapping Unit

زیرحوضه نمونه نسبت به شاهد، ۲۳۳ متر بیش‌تر است. تغییرات زیاد فاصله نمونه‌برداری در زیرحوضه نمونه نسبت به شاهد ناشی از انحراف زیاد متغیرهای ورودی (شیب‌های طولی و عرضی) در این نوع روش نمونه‌برداری است که تمامی تغییرات آن‌ها در نمونه‌برداری لحاظ شده است. قابل ذکر است که اگر روش نمونه‌برداری سنتی (روش واحدهای همگن) استفاده می‌شد، تغییرات مقادیر حدی به‌علت وسعت کم پلی‌گون‌ها، لحاظ نمی‌شدند و اهمیت آن‌ها نادیده گرفته می‌شد.

این اختلاف فراوانی علاوه بر تغییرات شیب‌های طولی و عرضی ممکن است ناشی از تفاوت شکل دو زیرحوضه باشد. بنابراین، بیش‌ترین تغییرات شیب‌های طولی و عرضی در زیرحوضه نمونه در فواصل ۶۰۰-۴۰۰ متر و در زیرحوضه شاهد در فواصل ۸۰۰-۶۰۰ متر رخ داده است. در نهایت، پیشنهاد می‌شود این روش با تعداد نقشه بیش‌تر و در نظر گرفتن نقشه‌های طبقه‌ای (توزیع آماری این نقشه‌ها) اجرا شده و معایب و مزایای آن در علوم و سطوح مختلف بررسی شود.

رهیافت ترویجی

ارائه یک روش نمونه‌برداری طرح مبنا در منابع طبیعی با تأکید بر در نظر گرفتن مقادیر حدی متغیرهای ورودی (نقشه‌ها) در نمونه‌برداری و ارائه یک روش نمونه‌برداری با تأکید نمونه‌برداری در فضای حالت (توزیع ورودی متغیرهای ورودی) بدون در نظر گرفتن توزیع جغرافیایی (مکانی).

که مقادیر بسیار کوچک و بسیار بزرگ متغیرها در قسمت دنباله‌های توزیع قرار می‌گیرند که دارای شانس یکسانی نسبت به قسمت‌های مرکزی توزیع (قسمت‌های با بیش‌ترین فراوانی) در قرار گرفتن نمونه در آن طبقه می‌باشند که این مورد از مزایای روش نمونه‌برداری مورد مطالعه می‌باشد. با توجه به این‌که زیرحوضه‌های مورد مطالعه حالت تپه‌ماهوری دارند، پراکنش نمونه‌ها نیز از این حالت تبعیت کرده و فاصله نمونه‌ها در مکان‌هایی با بیش‌ترین تغییرات متغیرهای کمکی (شیب‌های طولی و عرضی) نسبت به مکان‌های مسطح، از یکدیگر کم‌تر بوده و تراکم آن‌ها بیش‌تر است. لازم به ذکر است که این نوع نمونه‌برداری در اکثر مطالعات منابع طبیعی از اهمیت زیادی برخوردار بوده و بیش‌ترین تغییرات را در مورد متغیرهای کمکی مرتبط با متغیر مورد مطالعه پوشش می‌دهد.

از طرفی در این روش حتی می‌توان متغیرهای طبقه‌ای (واحدهای کاربری اراضی، زمین‌شناسی، پوشش گیاهی، رخساره فرسایشی و غیره) را وارد کرده و هم‌چنین می‌توان نقشه‌های متعددی را به‌عنوان نقشه‌های ورودی الگوریتم در نظر گرفت. حتی می‌توان تعداد نقشه‌های ورودی را افزایش داده تا بتوان حداکثر تغییرات را در تعیین شبکه بهینه نمونه‌برداری در نظر گرفت. این روش توسط پیسما و هویلینک (۱۹۹۹) و ژو و همکاران (۲۰۰۵) مورد تأیید قرار گرفته و برای تعداد نمونه کم پیشنهاد شده (۱۱ و ۱۲) که این پژوهش با نتایج آن‌ها هم‌خوانی دارد. بنابراین متوسط فاصله نمونه‌برداری در زیرحوضه شاهد نسبت به زیرحوضه نمونه، ۵۴ متر کم‌تر است، ولی حداکثر فاصله نمونه‌برداری در

منابع

1. Brus, D.J., De Gruijter, J.J., and Van Groenigen, J.W. 2004. Designing purposive and random spatial coverage samples by the k-means clustering algorithm. Global Workshop on Digital Soil Mapping, Montpellier, France, September 2004. 11p.
2. Carnell, R. 2009. Lhs: Latin hypercube samples. R package version 0.5. Augment a Latin Hypercube Design, 13p.
3. Goovaerts, P. 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press, Oxford, NY. 496p.
4. Hengl, T., Rossiter, D.G., and Stein, A. 2003. Soil sampling strategies for spatial prediction by correlation with auxiliary maps. *Austr. J. Soil Res.* 41: 1403-1422.
5. Gruijter, J., Brus, D.J., and Gruijter, J.J. 2006. Optimisation of sample configurations for digital soil mapping with universal kriging. *Developments in Soil Science*, 31: 137-151.
6. McBratney, A.B., Whelan, B.M., Walvoort, D.J.J., and Minasny, B. 1999. A purposive sampling scheme for precision agriculture. In: Stafford, J.J. (Ed.), *Precision Agriculture '99*. Sheffield Academic Press, Sheffield, Pp: 101-110.
7. McKay, M.D., Beckman, R.J., and Conover, W.J. 1979. Comparison of 3 methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21: 239.
8. McKenzie, N.J., and Ryan, P.J. 1999. Spatial prediction of soil properties using environmental correlation. *Geoderma*, 89: 67-94.
9. Minasny, B., and McBratney, A.B. 2002. Uncertainty analysis for pedotransfer functions. *Europ. J. Soil Sci.* 53: 417-430.
10. Minasny, B., and McBratney, A.B. 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computational Geoscience*, 32: 9. 1378-1388.
11. Pebesma, E.J., and Heuvelink, G.B.M. 1999. Latin hypercube sampling of Gaussian random fields. *Technometrics*, 41: 4. 303-312.
12. Royle, J.A., and Nychka, D. 1998. An algorithm for the construction of spatial coverage designs with implementation in SPLUS. *Computers and Geosciences*, 24: 5. 479-488.
13. Xu, C.G., He, H.S., Hu, Y.M., Changa, Y., Li, X.Z., and Bu, R.C. 2005. Latin hypercube sampling and geostatistical modeling of spatial uncertainty in a spatially explicit forest landscape model simulation. *Ecological Modeling*, 185: 255-269.
14. Zhang, Y., and Pinder, G.F. 2004. Latin-hypercube sample-selection strategies for correlated random hydraulic-conductivity fields. *Water Resources Research*, 39: 8-11.

ضمیمه

اجرای الگوریتم آبرمکعب لاتین در نرم‌افزار R با استفاده از بسته lhs برای زیرحوضه نمونه

```
Library(lhs)
```

```
X <- optimumLHS(60,2,30,.05); Y <- X
```

```
Y[,1] <- qnorm(X[,1],-0.005,0.025); Y[,2] <- qnorm(X[,2],-0.006,0.038)
```



Gorgan University of Agricultural
Sciences and Natural Resources

J. of Conservation and Utilization of Natural Resources, Vol. 6 (1), 2017
<http://ejang.gau.ac.ir>

Latin hypercube sampling in natural resources using R

***M. Hosseinalizade¹, H. Ahmadi², S. Feiznia² and F. Rivaz³**

¹Assistant Prof., Dept. of Watershed and Arid Zone Management, Gorgan University of Agricultural Sciences and Natural Resources, ²Full Professor, Dept. of Reclamation of Arid and Mountainous Regions, Faculty of Natural Resources, University of Tehran, ³Assistant Prof., Dept. of Statistics, Faculty of Mathematical Sciences, Shahid Beheshti University, G.C., Tehran, Iran

Received: 06/25/2013; Accepted: 03/05/2015

Abstract

Background and Objectives: In sampling strategies numerous goals like prediction in space and time is of importance. A great number of limitations including time, cost and available resources make sampling so important for all experts. Therefore, experts are searching for sampling optimization techniques and considering co-variables is one of the methods of such a problem solving. So, in cases of restricted primary variables, choosing the best auxiliary covariates is very important. In contrast to geographical space sampling that may not be a good representative of interested variable, feature space sampling is of great interest. Feature space sampling obtains samples with good coverage in both spaces. Latin Hypercube Sampling (LHS) is a stratified random procedure that provides an efficient way of sampling variables from their multivariate distributions. It provides a full coverage of the range of each variable by maximally stratifying the marginal distribution.

Materials and Methods: This study presents the conditioned Latin hypercube as a sampling strategy of an area with prior information (profile and plan curvature) as exhaustive ancillary data of soil erodibility in a paired sub-catchment of Golestan province. In reality, soil sample locations are chosen so that each variable by maximally stratify the marginal distribution. For both sub-catchments, 60 samples were chosen using sequential-based method in each sub-catchment by LHS package in R. Covariables point map by 20m intervals considered as input distribution to run LHS a maximum likelihood applied for estimation of distribution parameters.

Results: Results showed that the average of spatial distances for LHS for control and sample sub-catchments is 639 and 693 m, respectively. The highest frequency of sampling intervals for control and sample sub-catchments belongs to 400-600 and 600-800 m elevation categories, respectively.

Conclusion: LHS provides a full coverage of the range of each variable by maximally stratifying the marginal distribution and consider their extreme values. In conventional method of sampling that is based on minimum mapping unit, extreme values overlooked. Therefore, presenting a design-based sampling method based on marginal distribution with emphasis on feature space sampling is the main results of this research.

Keywords: Latin hypercube sampling, Soil erodibility, Paired sub-catchment, R software

* Corresponding author: alizadeh_m2001@yahoo.com