

تعیین الگوی نمونه برداری مکانی با استفاده از روش‌های مختلف (مطالعه‌ی موردی: شهرستان تفت)

روح‌اله تقی‌زاده مهرجردی*

استادیار دانشکده کشاورزی و منابع طبیعی دانشگاه اردکان

تاریخچه مقاله	چکیده
دریافت: ۱۳۹۲/۱۱/۳۰ پذیرش نهایی: ۱۳۹۴/۰۱/۲۱ کلمات کلیدی: داده‌های محیطی، نقشه برداری رقومی خاک، مربع لاتین، فازی کی مینز، چاردرخت مرکزی، چاردرخت تصادفی	طراحی الگوی نمونه برداری مکانی از مراحل اصلی نقشه- برداری رقومی خاک به شمار می‌رود. در این تحقیق از تکنیک‌های مربع لاتین، تصادفی، شبکه منظم، فازی کی مینز، چاردرخت مرکزی و چاردرخت تصادفی جهت تعیین ۱۵۱ نقطه‌ی نمونه برداری در منطقه‌ای به وسعت ۳۵۰۰ هکتار در شهرستان تفت استفاده گردید. نتایج آماری داده‌های محیطی از قبیل شاخص خیسی، شیب، جهت شیب، انحنای شیب، ارتفاع و شاخص گیاهی نرمال شده در نقاط نمونه برداری شده توسط تکنیک‌ها با جامعه اصلی مقایسه گردید. نتایج نشان داد که پارامترهای آماری میانگین و واریانس نمونه‌ها نسبت به جامعه در روش چاردرخت مرکزی بیشترین شباهت را دارند؛ به طوری که اختلاف میانگین شاخص خیسی در تکنیک‌های مربع لاتین، تصادفی، فازی کی مینز، شبکه منظم، چاردرخت مرکزی و چاردرخت تصادفی با جامعه به ترتیب ۰/۸ درصد، ۲/۲ درصد، ۱/۷ درصد، ۱/۴ درصد، ۰/۱۵ درصد، ۵/۶ درصد و ۱/۰۹ درصد می‌باشد. در حقیقت الگوی نمونه برداری چاردرخت مرکزی طوری محل نقاط را از جامعه انتخاب کرده است که توزیع نمونه گیری در هیستوگرام یکنواخت باشد. در حالی که در روش‌های دیگر شاهد نمونه برداری بیشتر یا کمتر در بعضی از قسمت‌های هیستوگرام‌ها و یا به نوعی اریب هستیم. درک و تلقی عمومی از نمونه برداری خوب آن است که می‌باید پراکنش جغرافیایی خوبی از نقاط نمونه برداری وجود داشته باشد. در حالی که نتایج نشان داد که همیشه نمونه برداری با توزیع جغرافیایی مناسب، تمام ویژگی‌های منطقه را مشخص نمی‌کند.

* عهده دار مکاتبات

E-mai: rtaghizadeh@ardakan.ac.ir

مقدمه

فرآیند طراحی الگوی نمونه‌برداری در نقشه‌برداری سنتی دارای مشکلات متعددی می‌باشد (۹). به عنوان مثال، روش‌های معمول شناسایی خاک نیازمند نمونه برداری‌های متراکم و تجزیه‌های آزمایشگاهی است که این گونه عملیات نیازمند صرف وقت و هزینه‌های فراوان می‌باشد؛ علاوه بر این، هیچ معیار آماری برای مشخص کردن محل نمونه‌های خاک در نقشه‌برداری سنتی خاک وجود ندارد و این موضوع منجر به عدم پراکنش مناسب در نمونه برداری می‌شود. این درحالی است که در نقشه‌برداری رقومی خاک، روش‌های آماری مختلفی از جمله واریانس چاردرخت^۱ (۱۵)، مربع لاتین^۲ (۱۴) و فازی کی-مینز^۳ (۱۳) جهت تعیین الگوی نمونه-برداری معرفی شدند. در نقشه‌برداری رقومی خاک، مکان‌های نمونه‌برداری خاک برای ایجاد روابط کمی بین خاک و داده‌های محیطی تأثیر بنیادین دارد و در نتیجه کیفیت نقشه‌های حاصله وابستگی شدیدی به محل مشاهدات خاک دارد. در حقیقت، صحت و دقت نقشه-برداری رقومی خاک در مرحله اول تابع تعداد نمونه و در مرحله دوم به مکان فضائی نمونه در سطح مورد مطالعه بستگی دارد (۱۶).

با توجه به اهمیت طراحی الگوی نمونه‌برداری در نقشه‌برداری رقومی خاک، محققان از تکنیک‌های مختلف نمونه‌برداری در مطالعات استفاده کرده‌اند. ساده‌ترین شکل نمونه‌برداری، شبکه منظم نمونه‌برداری می‌باشد. اما سوال پیش رو در این روش، فاصله‌ی نمونه‌ها می‌باشد. مک‌براتنی و وبستر^۴ (۱۰) استفاده از آنالیز واریوگرام را برای تعیین فواصل شبکه پیشنهاد دادند. به طور مشابه، وان گروگین و همکاران^۵ (۲۲) از حداقل

کردن واریانس کریجینگ برای مشخص کردن محل نمونه‌ها استفاده کردند. اما مشکل اساسی این تکنیک‌ها این است که نیاز به دانستن واریوگرام پارامتر مورد نظر در منطقه دارند. لذا محققان، اقدام به استفاده از روش‌هایی نمودند که تنها بر اساس اطلاع از موقعیت جغرافیایی منطقه، سعی در انتخاب نقاطی می‌نمایند که حداقل فاصله مکانی را از یکدیگر داشته باشند. به عنوان مثال، بروس و همکاران^۶ (۲ و ۳) الگوریتم‌های کی‌مینز را برای حداقل کردن متوسط کوتاهترین فواصل استفاده کردند. میناسنی و مک‌براتنی (۱۳) نیز از الگوریتم فازی کی‌مینز برای تعیین محل جغرافیایی نمونه‌ها بهره گرفتند. در سال‌های اخیر، روش‌هایی توسعه یافته‌اند که از داده‌های محیطی بهره می‌گیرند. این روش‌ها سعی می‌کنند که نمونه‌ها در فضاهایی انتخاب شوند که تمام دامنه تغییرات هر متغیر کمکی را پوشش دهند. به عنوان مثال، هاولینک و همکاران^۷ (۷) نمونه‌برداری با حداقل کردن واریانس کریجینگ جهانی را طراحی کردند. سیمباهان و دربرمن^۸ (۲۰) سه معیار متفاوت بهینه‌سازی را مقایسه کردند: حداقل کردن متوسط کوتاهترین فواصل، یک توزیع یکنواخت از جفت نقاط برای تخمین واریوگرام و ترکیبی از هر دو. همه این روش‌ها شامل الگوریتم‌های شبیه‌سازی شده است. هنگلو همکاران^۹ (۵) نمونه‌برداری را در طول اجزاء اصلی داده‌های محیطی پیشنهاد کردند. تعداد نمونه‌های گرفته شده از هر بخش، متناسب با واریانس کل تشریح شده به وسیله آن بخش می‌باشد. لارک^{۱۰} (۸) مجموعه‌های فازی فضاهای شبکه‌ای را، در صورت وجود عدم قطعیت برای واریوگرام‌ها معرفی کرد و این مدل پیشنهادی، سعی در کاهش دادن واریانس متغیرهای وابسته دارد. در این روش، منطقه‌ی

6- Brus *et al.*7- Heuvelink *et al.*

8- Simbahan and Dobermann

9- Hengl *et al.*

10- Lark

1- Variance Quadtree

2- Latin Hypercube

3- Fuzzy K-means

4- McBratney and Webster

5- Van Groenigen *et al.*

روش‌های نمونه‌برداری سنتی (تصادفی و شبکه منظم) و جدید (مربع لاتین، فازی کی‌مینز، الگوریتم چاردرخت تصادفی و الگوریتم چاردرخت مرکزی) بر اساس واریانس داده‌های محیطی مقایسه شده و سپس بهترین الگو به عنوان اولین گام در مطالعات نقشه‌برداری رقومی خاک معرفی می‌گردد.

مواد و روش‌ها

معرفی محدوده‌ی مطالعاتی

برای اجرای این تحقیق، قسمتی از اراضی حوضه‌ی جان‌برازان (ده‌بالا) واقع در جنوب غربی شهرستان تفت به مساحت تقریبی ۳۵۰۰ هکتار انتخاب شد. این حوزه در محدوده جغرافیایی با عرض شمالی ۳۳° ۳۱' تا ۳۹° ۳۱' و طول شرقی ۵۲° ۰۳' تا ۵۴° ۰۹' قرار دارد. بلندترین نقطه‌ی حوضه با ارتفاع ۴۰۵۰ متر از سطح دریا و پست‌ترین نقطه‌ی آن دارای ارتفاع ۲۱۷۱ متر می‌باشد. ساختار زمین‌شناسی منطقه از واحدهای گوناگونی تشکیل گردیده که در ادامه شرح داده شده است: رسوبات دوران کواترنری که اراضی باغی و زراعی در آن قرار گرفته است؛ سنگ‌های گرانیتی دوران ژوراسیک که مساحت زیادی از حوزه را در بر گرفته است، سازندهای آهکی دوران مزوزوئیک که بیشتر در اطراف مرز حوزه مشاهده می‌شوند و سازند شتری که دوران ائوسن که کمترین مساحت را دارا می‌باشند. شکل (۱) موقعیت منطقه‌ی مورد مطالعه بر روی مدل رقومی ارتفاع در حوزه ده‌بالا را نشان می‌دهد.

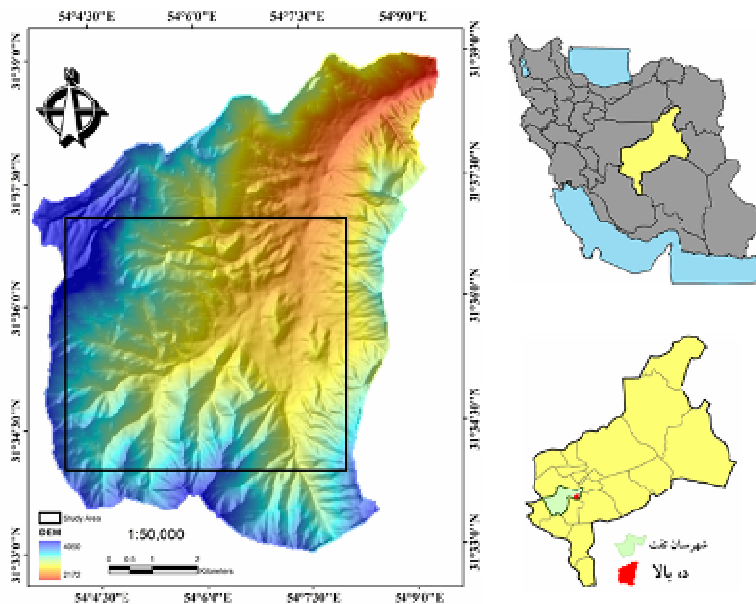
مورد مطالعه به نواحی یکنواخت بر اساس داده‌های محیطی تقسیم‌بندی می‌گردند و سپس در هر یک از نواحی، تعدادی نمونه به صورت تصادفی انتخاب می‌شوند. به طور مشابه، مک کنزی و رایان^۱ (۱۲) پارامترهای سرزمین، داده‌های اقلیمی و زمین‌شناسی را برای کلاس-بندی منطقه استفاده کردند و سپس نمونه‌ها را به طور تصادفی درون هر کلاس انتخاب نمودند. گسler و همکاران^۲ (۴) محل نمونه‌های خاک را به طور تصادفی در دامنه شاخص توپوگرافی مرکب انتخاب کردند. میناسنی و مک براتنی (۱۴) طرح نمونه‌برداری مربع لاتین را برای نقشه‌برداری رقومی خاک پیشنهاد کردند؛ البته در مناطقی که تنها اطلاعات موجود، داده‌های محیطی منطقه هستند. مربع لاتین یک طرح نمونه‌برداری موتو کارلوی محدود شده است. این روش، یک طرح تصادفی طبقه‌بندی شده‌ای است که باعث نمونه‌برداری موثری به کمک توزیع چند متغیر می‌شود. میناسنی و همکاران (۱۵) همچنین الگوریتم چاردرخت را برای تعیین محل نمونه‌ها بر اساس داده‌های محیطی معرفی کردند. واژه‌ی چاردرخت به مفهوم درختی می‌باشد که هر گره آن دارای چهار شاخه می‌باشد. در حقیقت، چاردرخت یک روش تجزیه کننده سلسله مراتبی است که سعی در تقسیم کردن تصویر برای رسیدن به نواحی همگن دارد (۱۹). مک براتنی و همکاران (۱۱) از تکنیک الگوریتم چاردرخت برای تقسیم‌بندی منطقه به نواحی مشابه استفاده و سپس در نواحی جدا شده تعدادی نمونه انتخاب کردند.

با توجه به مطالب فوق، به نظر می‌رسد که از مهم‌ترین مراحل نقشه‌برداری رقومی خاک، تعیین الگوی نمونه‌برداری می‌باشد؛ زیرا دقت و صحت تمام داده‌ها و اطلاعات خاک وابسته به نمونه‌های خاک برگرفته از نمونه‌برداری می‌باشد؛ بنابراین در تحقیق حاضر در ابتدا

1- McKenzie and Ryan

2- Gessler *et al.*

تقی زاده مهرجردی: تعیین الگوی نمونه برداری مکانی با استفاده ...



شکل (۱) موقعیت منطقه مورد مطالعه بر روی مدل رقومی ارتفاع استان یزد

Figure (1) Location of study area dropped over digital elevation model in Yazd province

تهیه داده‌های محیطی

در تحقیق حاضر از مدل رقومی وب سایت مدل رقومی ارتفاع سازمان زمین شناسی آمریکا (رزولیشن ۹۰ متر) استفاده گردید. بعد از آماده کردن مدل ارتفاعی رقومی، از آن برای استخراج پارامترهای سرزمین استفاده گردید. برای آماده سازی مدل رقومی ارتفاع در ابتدا مقادیر پرت پیکسل‌ها (چاله‌های مصنوعی) اصلاح شده^۱ و سپس فیلتر از نوع ساده^۲ بروی آن اعمال گردید. همچنین با دستور نمونه گیری مجدد^۳ اندازه پیکسل‌ها به ۳۰ متر تغییر یافت. پارامترهای سرزمین از قبیل شیب، جهت، انحنای طولی شیب، انحنای سطحی، ارتفاع، ارتفاع بالای شبکه زهکشی، مساحت حوزه‌های اصلاح شده، موقعیت میانی شیب، عمق دره، شاخص خیسی، شاخص همواری دره با درجه تفکیک بالا، شاخص بالای پشته با درجه تفکیک بالا و شیب حوزه در محیط سامانه‌ی جغرافیایی ساگا محاسبه و استخراج گردید (جدول ۱). روش

استخراج تمام پارامترهای مزبور در روش ارائه شده توسط هنگل و همکاران (۶) تشریح گردیده است. همچنین در این مطالعه از تصاویر ماهواره‌ای سنجنده ETM^+ استفاده شد. تصاویر این سنجنده مربوط به تابستان ۱۳۹۲ می‌باشد. این تصویر از ۶ بانده (باند ۱: ۰/۴۵-۰/۵۲ میکرومتر، باند ۲: ۰/۵۲-۰/۶۰ میکرومتر، باند ۳: ۰/۶۳-۰/۶۹ میکرومتر، باند ۴: ۰/۷۶-۰/۹۰ میکرومتر، باند ۵: ۱/۵۵-۱/۷۵ میکرومتر، باند ۷: ۲/۰۸-۲/۳۵ میکرومتر) تشکیل شده است. برای کاهش دادن تعداد داده‌های خام (تعداد باندها) تصویر ماهواره‌ای از تجزیه به مولفه‌های اصلی بر اساس ماتریس همبستگی استفاده گردید. یکسری شاخص‌ها از قبیل شاخص گیاهی نرمال شده (نشان دهنده‌ی وضعیت پوشش گیاهی)، شاخص رس، شاخص کرنات، شاخص گچ، شاخص شوری و شاخص روشنائی (نشان دهنده‌ی ترکیبات زمین شناسی مواد سطحی) محاسبه گردید (جدول ۱). شکل (۲) بعضی

- 1- Fill Sink
- 2- Simple Filter
- 3- Resampling

جدول (۱) داده‌های محیطی مورد استفاده در مدل‌های نمونه‌برداری

Table (1) Auxiliary variables used in sampling models

Defination	تعریف	پارامترهای محاسبه شده
Elevation	ارتفاع بالای سطح دریای آزاد (متر)	ارتفاع
Altitude above channel network	ارتفاع بالای آبراهه	ارتفاع بالای شبکه زهکشی
Modified catchment area	تجمعات جریان آب و پارامترهای مربوطه	مساحت حوزه‌های اصلاح شده
Mid-Slop position	مقدار تشابه نسبت به قسمت پایین و بالا	موقعیت شیب میانی
Multi-Resolution valley bottom flatness index	میزان همواری و پائین بودن را محاسبه می‌کند	شاخص همواری دره با درجه تفکیک بالا
Multi-Resolution ridge flatness index	میزان همواری و بالا بودن را محاسبه می‌کند	شاخص همواری قله با درجه تفکیک بالا
Valley depth	شاخص نسبت پهنای دره به عمق دره	عمق دره
Wetness Index	معیاری از تجمع رطوبت	شاخص خبسی
Catchment slope	میزان گرادیان بالای مسیر حرکت	شیب حوزه
Slope	زاویه بین سطح تماس و سطح افقی در یک نقطه در سطح زمین	شیب
Aspect	بیشترین سرعت تغییر در ارتفاع در هر سلول از مدل ارتفاعی	جهت
Curvature	انحنای در صفحه عمودی خط جریان است.	انحنای شیب
Plan Curvature	بیان‌گر اندازه همگرایی و واگرایی جریان می‌باشد.	انحنای سطحی
Principal component analysis	PC1, PC2, PC3	تجزیه به مولفه‌های اصلی ۶ باند
Normalized differential vegetation index	$(B4 - B3) / (B4 + B3)$	شاخص گیاهی نرمال شده
Clay Index	$B5 / B7$	شاخص رس
Gypsum Index	$(B5 - B4) / (B5 + B4)$	شاخص گچ
Salinity Index	$(B3 - B4) / (B2 + B4)$	شاخص شوری
Brightness Index	$((B3)2 + (B4)2)0.5$	شاخص روشنایی
Carbonate Index	$B3 / B2$	شاخص کربناته

اختصاص داده بودند، به عنوان ورودی در نظر گرفته شدند.

$$E_{ij} = -k \sum (P_{ij} \times \ln P_{ij}) \quad (1)$$

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

طراحی الگوی نمونه‌برداری

مربع لاتین: روش نمونه‌برداری مربع لاتین توسعه‌ای جدید در فناوری نمونه‌برداری است که برای شبیه‌سازی دقیق توزیع احتمال داده‌های محیطی از طریق نمونه‌برداری از طریق تعداد کمتری سعی و خطا در مقایسه با

از متغیرهای محیطی را در منطقه‌ی مورد مطالعه نشان می‌دهد.

انتخاب داده‌های کمکی

برای انتخاب ورودی‌های مدل‌ها از تکنیک آنتروبی شانن مطابق با معادله‌ی (۱) استفاده گردید. ایده‌ی روش فوق این است که هر چه پراکندگی در مقادیر یک شاخص، بیشتر باشد، آن شاخص نسبت به دیگر شاخص‌ها از اهمیت بیشتری برخوردار است (۱۶). برای رتبه‌بندی کردن ورودی‌ها از نرم افزار اکسل استفاده شد. در مطالعه‌ی حاضر شش متغیری که بیشترین وزن را به خود

تقی زاده مهرجردی: تعیین الگوی نمونه برداری مکانی با استفاده ...

که در این معادله A ماتریس نرمال شده فاصله است که در اینجا از فاصله ماهالونوبیس استفاده گردید. کمینه-سازی تابع هدف J راه حلی را برای ماتریس توابع عضویت M و مرکز خوشه C به صورت معادله‌ی (۴) و (۵) فراهم می‌کند.

$$m_{ij} = \frac{d_{ij}^{2/(\varnothing-1)}}{\sum_{j=1}^k d_{ij}^{2/(\varnothing-1)}} \quad i=1,2,\dots,n; j=1,\dots,k \quad (4)$$

$$C_j = \frac{\sum_{i=1}^n m_{ij}^{\varnothing} X_i}{\sum_{i=1}^n m_{ij}^{\varnothing}} \quad j=1,\dots,k \quad (5)$$

نمای فازی به زیرساخت‌های موجود در مجموعه داده‌ها و در نتیجه به تعداد مطلوب کلاس‌ها بستگی دارد. در واقع نمای فازی میزان همپوشی کلاس‌ها را تعیین می‌کند. بنابراین توابع اعتبارسنجی خوشه‌بندی، تا حدودی معیاری برای سنجش فازی بودن نیز هستند؛ با این وجود با بسط توابع عضویت فازی باید مقدار مناسبی از \varnothing را انتخاب کرد تا نمای فازی، زیرساخت‌های مجموعه داده‌ها را نشان دهد. به منظور تعیین بهترین تعداد کلاس (خوشه) از شاخص‌های آنتروپی خوشه‌بندی اصلاح شده و شاخص خوشه‌بندی فازی استفاده گردید (۱۳). تعداد بهینه خوشه‌ها، با مقدار حداقل این توابع انطباق دارد؛ سپس، در هر خوشه (بر حسب درجات عضویت) تعدادی نقاط نمونه‌برداری (جمعاً ۱۵۱ نقطه) انتخاب گردید. برای انجام این روش از نرم‌افزار MATLAB استفاده گردید (۱۳).

چاردرخت: الگوریتم واریانس چاردرخت روشی سلسله مراتبی جهت تقسیم کردن منطقه‌ی مورد مطالعه به واحدهای همگن می‌باشد. در این روش، ابتدا، منطقه مورد مطالعه به چهار قسمت مساوی تقسیم می‌شوند. در مرحله‌ی بعد، دوباره هر یک از ناحیه‌های ایجاد شده در مرحله قبل به چهار قسمت مساوی دیگر تقسیم می‌شوند. لازم به ذکر می‌باشد که از مرحله دوم به بعد، تقسیم‌بندی چهارتایی فقط در قسمت‌هایی از منطقه مورد مطالعه انجام می‌گیرد که داده‌های محیطی در آنجا دارای بیشترین واریانس باشند. الگوریتم این فرآیند را آنقدر

روش مونت کارلو طراحی شده است. نکته‌ی کلیدی در نمونه‌برداری به روش مربع لاتین دسته‌بندی توزیع احتمال داده‌های محیطی است. توزیع احتمال تجمعی داده‌های محیطی از مقیاس ۰ تا ۱ به فواصل مساوی تقسیم‌بندی می‌شود. سپس از هر فاصله یا دسته توزیع احتمال تجمعی یک نمونه گرفته می‌شود. در این روش از مقادیری در هر فاصله یک نمونه گرفته می‌شود و در نهایت توزیع احتمال ورودی شبیه‌سازی می‌شود. برای انجام این روش از نرم‌افزار MATLAB استفاده گردید (۱۴).

فازی کی‌مینز: خوشه‌بندی، یکی از شاخه‌های یادگیری بدون نظارت است و فرآیند خودکاری است که طی آن، نمونه‌ها به دسته‌هایی که اعضای آن مشابه یکدیگر هستند، تقسیم می‌شوند و به این دسته‌ها، خوشه گفته می‌شود. الگوریتم کی‌مینز یکی از روش‌های خوشه‌بندی ساده و سریع محسوب می‌شود. این الگوریتم با حرکت کردن بر روی مجموعه داده‌های محیطی قصد دارد تا این داده‌ها را به k خوشه دسته‌بندی کند. الگوریتم با انتخاب k عدد داده‌ی انتخابی برای استفاده به عنوان خوشه‌ی اصلی به روش‌های مختلف (شانسی یا بر اساس اطلاعات اولیه‌ی در دسترس) شروع می‌شود و میانه هر شاخه را محاسبه می‌کند: در مرحله‌ی اول، با حرکت بر روی مجموعه داده‌ها و در نظر گرفتن میانه خوشه‌ها، داده جدید به نزدیک‌ترین خوشه اضافه می‌شود. در مرحله‌ی دوم، با اضافه شدن داده جدید برای هر خوشه، میانه جدید محاسبه می‌شود. این کار تا زمانی که تغییری در مرکز خوشه ایجاد نشود، ادامه می‌یابد. در این روش فاصله داخل خوشه‌ها کمینه می‌شود. تابع هدف به صورت معادله‌ی (۲) می‌باشد:

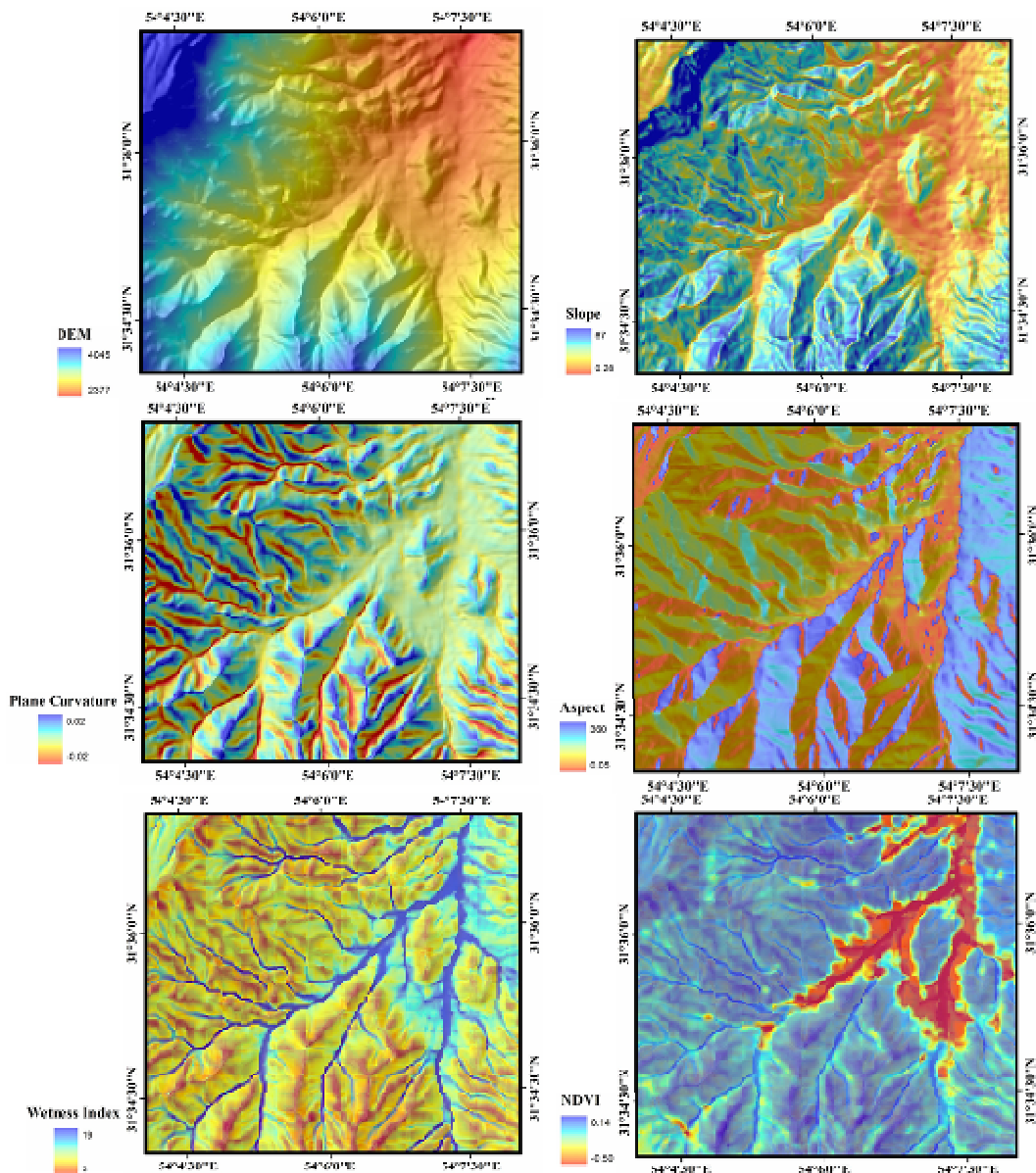
$$J(M, C) = \sum_{i=1}^n \sum_{j=1}^k m_{ij}^{\varnothing} d_{ij}^2 \quad (2)$$

که در این معادله \varnothing نمای فازی معادله است که درجه فازی شدن را نشان می‌دهد و d ماتریس فاصله می‌باشد و به صورت معادله‌ی (۳) محاسبه می‌شود:

$$d_{ij}^2 = (X_i - C_j)^T A (X_i - C_j) \quad (3)$$

دو رویکرد مرکز سلول و تصادفی بهره گرفته شد؛ بدین ترتیب که در رویکرد اول، موقعیت جغرافیایی در مرکز پیکسل‌های نهایی منتخب توسط الگوریتم چاردرخت به عنوان محل نمونه‌برداری انتخاب شد و در رویکرد دوم (تصادفی) موقعیت جغرافیایی محل نمونه‌برداری به طور تصادفی از داخل پیکسل انتخاب گردید.

ادامه می‌دهد تا واریانس در سلول‌های ایجاد شده حداقل شود؛ البته می‌توان شرط اتمام الگوریتم چاردرخت را تکرار معینی نیز قرار داد. در مطالعه حاضر با توجه به این که فقط تعداد ۱۵۱ نمونه لازم می‌باشد، از شرط اتمام ۵۰ تکرار استفاده گردید. پس از اتمام فرآیند تقسیم‌سازی منطقه به نواحی چهارگانه، در هر یک از سلول‌های ایجاد شده نمونه‌ای برداشت شد. در تحقیق حاضر از



شکل (۲) برخی از پارامترهای محیطی در منطقه‌ی مورد مطالعه
Figure (2) Selected auxiliary variables in the study area

تقی زاده مهرجردی: تعیین الگوی نمونه برداری مکانی با استفاده ...

با توجه به جدول ۲، شش پارامتری که دارای وزن بیشتری هستند از قبیل نقشه ارتفاع، شیب، جهت، شاخص خیسی، انحنا، شیب و شاخص گیاهی نرمال شده وارد فرآیند مدل سازی مربع لاتین، الگوریتم چاردرخت و فازی کی مینز گردیدند. همان طور که در جدول ۲ ملاحظه می گردد، مجموع اوزان یک نمی شود که به این دلیل است که پارامترهای دیگر مقدار وزن کمتری را به خود اختصاص دادند و در نتیجه وارد مدل نشدند. پس از مشخص شدن پارامترهای ورودی، مدل مربع لاتین اجرا گردید. تکامل تابع هدف که واریانس می باشد را با ده هزار تکرار در شکل ۳ ارائه شده است. این الگوریتم، شروع به نمونه برداری تصادفی می کند تا بتواند نقاطی که مربع لاتین را تشکیل می دهند، پیدا کند. مقدار تابع هدف^۱ با افزایش تکرار کاهش می یابد و در شرایط تکرارهای کم تر از ۱۰۰۰ قدری آشفتگی از خود نشان می دهد. برای این تحقیق از تعداد تکرار ده هزار برای یافتن جواب بهینه استفاده گردید. محل هایی که توسط مربع لاتین انتخاب شده در شکل (۹) ارائه شده است.

در این روش داده ها در گروه هایی که اعضای هر گروه از جنبه های خاصی بیشترین شباهت را با هم داشته و با اعضای خوشه های دیگر هیچ شباهتی ندارند، قرار داده می شوند. برای انتخاب تابع فاصله برای خوشه بندی در مطالعه حاضر از نرم ماها لانونیسیس استفاده شد. پس از انتخاب تابع فاصله مهم ترین پارامتر در طبقه بندی فازی تعیین مقدار نمای فازی بودن است. در این مطالعه خوشه بندی بر اساس تعداد کلاس های فازی در گستره ۳-۷ و نمای فازی بودن در دامنه ۱-۲ با گام ۰/۰۵ در نظر گرفته شد و این الگوریتم بروی داده های محیطی (جدول ۲) اجرا گردید. از مهم ترین شرایط برای قابل

برای انجام این روش از نرم افزار MATLAB استفاده گردید (۱۵).

تصادفی: در نمونه برداری تصادفی، هر پیکسل در کل منطقه ی مورد مطالعه، شانس یکسانی برای برگزیده شدن دارد. جهت انتخاب ۱۵۱ نمونه تصادفی از بین ۴۴۰۰۰ پیکسل (۳۰ متری) از نرم افزار مینی تب استفاده گردید. لازم به ذکر می باشد که بر اساس مطالعات نیمه-تفضیلی دقیقی این تعداد نمونه انتخاب شده است (۱۷).

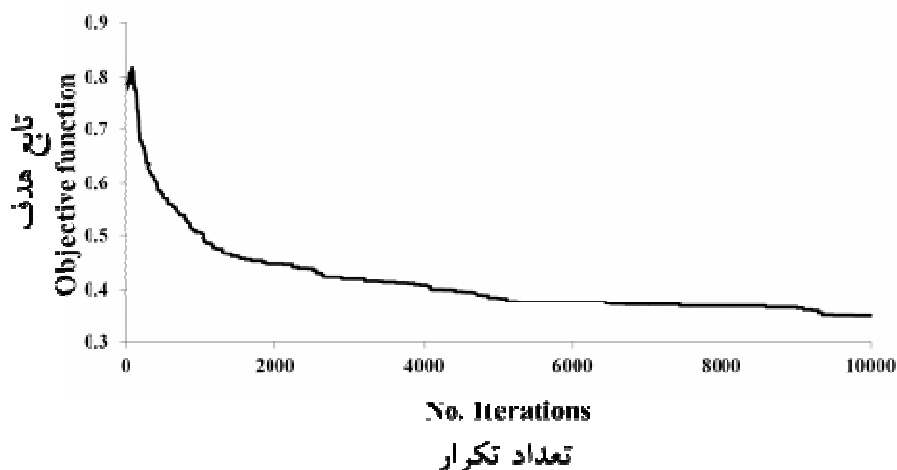
شبکه ای منظم: در نمونه برداری شبکه ای، تعداد ۱۵۱ نمونه به فاصله یکسان (۵۰۰ متر) و به صورت شبکه ای در منطقه توزیع شدند. جهت انجام این روش از نرم افزار ArcGIS استفاده گردید.

ارزیابی روش های نمونه برداری

در مطالعه حاضر جهت ارزیابی روش های نمونه برداری از پارامترهای آماری از قبیل انحراف معیار و ضریب تغییرات استفاده گردید؛ بدین منظر، در ۱۵۱ نقطه انتخاب شده توسط روش های مختلف نمونه برداری مقادیر عددی داده های محیطی استخراج شد. سپس واریانس و ضریب تغییرات محاسبه گردید. با توجه به این که هدف نمونه برداری، پوشش دادن حداکثر تغییرات است، در نتیجه بهترین روش نمونه برداری، روشی خواهد بود که بالاترین انحراف معیار و ضریب تغییرات را در داده های محیطی داشته باشد. همچنین، الگوی نمونه برداری مناسب تر خواهد بود که توزیع هیستوگرام داده های محیطی در محل های نمونه برداری شباهت بیشتری با توزیع هیستوگرامی داده های محیطی در کل منطقه داشته باشد (۱۴).

نتایج و بحث

در تحقیق حاضر از انواع روش های طراحی الگوی نمونه برداری برای انتخاب ۱۵۱ محل نمونه برداری از بین ۴۴۰۰۰ فضای ممکن استفاده گردید. ورودی های این مدل ها بر اساس تکنیک آنتروپی شانن انتخاب گردید (جدول ۲).



شکل (۳) تکامل معادله هدف با تعداد تکرار ده هزار

Figure (3) Evolution of objective function with 10000 of iterations

جدول (۲) وزن‌های داده‌های محیطی انتخاب شده برای مدل‌های مربع لاتین، الگوریتم چاردرخت و فازی کی‌میز
 Table (2) Weight of selected auxiliary variables for sampling models (i.e. conational hyper cube method, variance quadtree algorithm and fuzzy k-means)

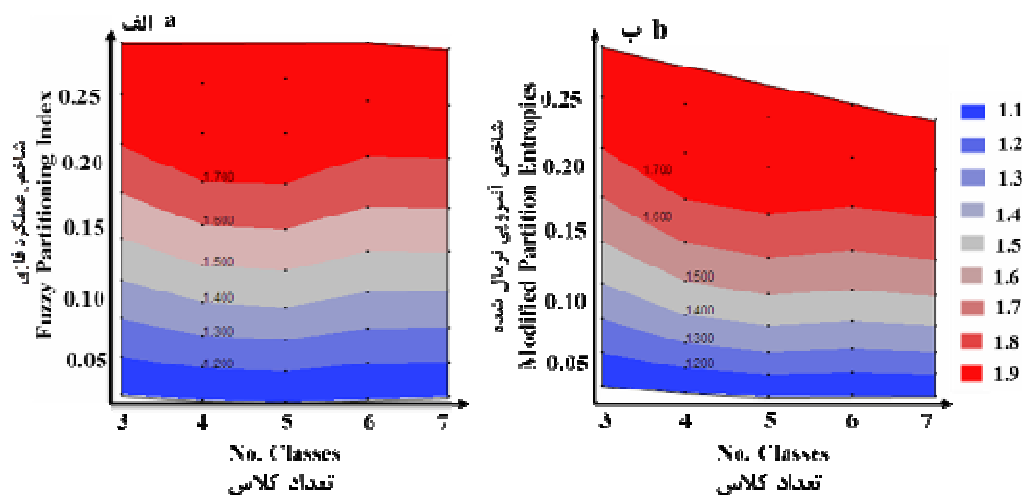
NDVI	جهت Aspect	انحنای شیب Curvature	شاخص خیزی Wetness Index	شیب Slope	ارتفاع Elevation	پارامتر Parameter
0.03	0.05	0.11	0.11	0.28	0.41	وزن Wieght

دارد. در تعداد کلاس پنجم کم‌ترین این مقدار ملاحظه می‌شود. اما شاخص آنتروپی نرمالیزه شده در کلاس هفتم حداقل مقدار خود را دارد. در چنین شرایطی از شاخصی استفاده می‌کنیم که حداکثر تعداد کلاس را مشخص کرده است (۱۳). بنابراین می‌توان منطقه مورد مطالعه را به هفت گروه همگن تقسیم کرد (شکل ۵). سپس در هر کلاس نسبت به وسعتی که در منطقه داشت تعدادی محل نمونه‌برداری انتخاب گردید (شکل ۹). در این مرحله پیکسل‌هایی که دارای بیشترین درجه عضویت می‌باشند انتخاب شدند.

جهت نمونه‌برداری با استفاده از الگوریتم چاردرخت، از ورودی‌های جدول (۲) استفاده گردید. در این طرح، در ابتدا منطقه به چهار قسمت مساوی تقسیم شد (شکل ۶).

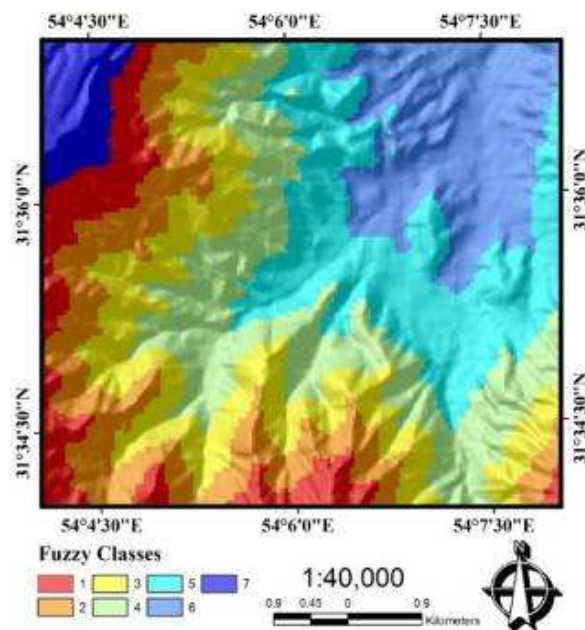
پذیرش بودن یک خوشه‌بندی آن است که میانگین مقدار عضویت برای هر نمونه در کلاس‌های عادی و کلاس نامعمول در حد امکان نزدیک به هم بوده و در حالت آرمانی برابر باشد، در این حالت شانس قرار گرفتن یک داده در همه کلاس‌ها برابر و آن چه در تعلق آن اثر دارد، فاصله آماری آن از مراکز کلاس‌ها است. بنابراین در مطالعه حاضر تمامی حالت‌های خوشه‌بندی انجام گرفت و انتخاب تعداد بهینه کلاس‌ها با استفاده از توابع شاخص عملکرد فازی و آنتروپی نرمال شده در هر نمای فازی بودن محاسبه و نسبت به تغییرات مقادیر شاخص‌ها بر حسب تعداد کلاس‌ها ترسیم شد. همان‌طور که در شکل ۴ ملاحظه می‌شود، با افزایش تعداد کلاس‌ها تابع شاخص عملکرد فازی روند کاهشی

تقی زاده مهرجردی: تعیین الگوی نمونه برداری مکانی با استفاده ...



شکل (۴) تغییرات شاخص خوشه بندی فازی (الف) و شاخص آنتروپی خوشه بندی اصلاح شده (ب) با درجات فازی شدن متفاوت برای کلاس های سه تا هفت

Figure (4) Variation of fuzzy partitioning index (a) and modified partition entropies (b) using different fuzzification values for three to seven classes.

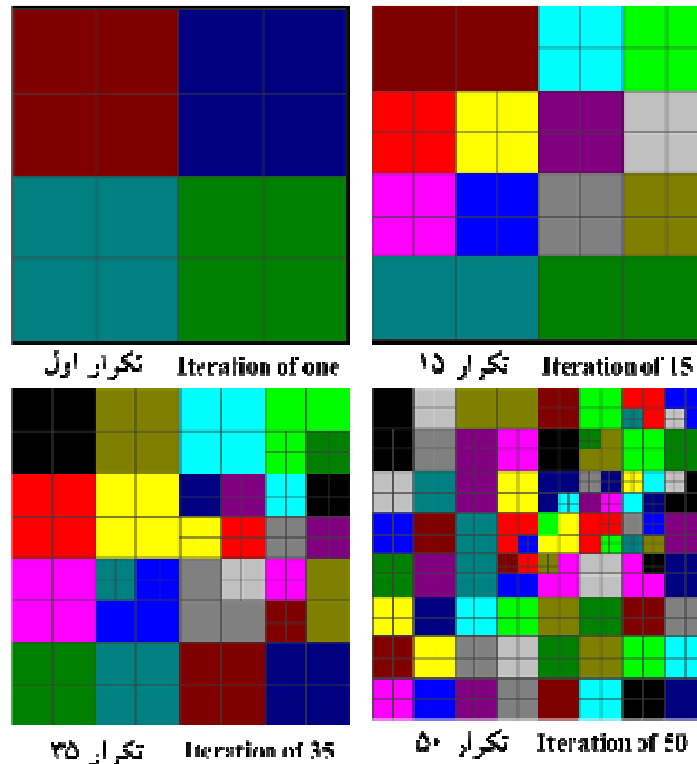


شکل (۵) نقشه طبقه بندی داده های محیطی بر اساس روش فازی کی مین

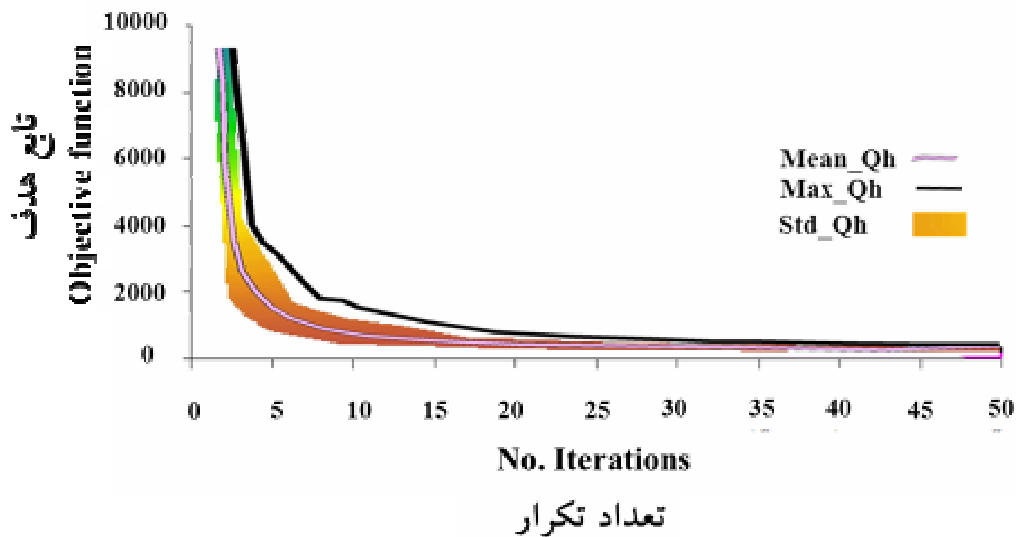
Figure (5) Classification map of auxiliary variables using fuzzy k-means

مناطقی که دارای تغییرات زیادی نمی باشند، تقسیم بندی برای آن ها صورت نگرفته است. همچنین شکل ۷ کاهش تدریجی مقدار ماکزیمم و میانگین واریانس را با افزایش تکرار نشان می دهد. بیشترین کاهش واریانس در تکرار ۱۰ مشاهده می شود. بعد از ۴۵ تکرار تقریباً تغییرات

سپس مدل منطقه را به سلول های چهارتایی بیشتر تقسیم می کند. همان طور که در شکل مشاهده می شود این تقسیم بندی ها در مناطقی صورت گرفته است که بیشترین واریانس وجود دارد. به عنوان مثال، مناطق مرکزی دارای واریانس بیشتری می باشند. به طور مشابه



شکل (۶) مثالی از واریانس چاردرخت در تکرارهای یک، ۱۵، ۳۵ و ۵۰
 Figure (6) Example of variance quadtree for iteration 1, 15, 35 and 50



شکل (۷) کاهش تدریجی ماکزیمم، میانگین و انحراف معیار واریانس داخل سلولها (Q_h)
 Figure (7) Gradual reduction of maximum, mean and standard deviation within-strata semi-variance

عنوان مثال در مورد شاخص جهت شیب، مقدار میانگین داده‌های محیطی انتخاب شده توسط تکنیک مربع لاتین، چاردرخت تصادفی، فازی کی‌مینز، تصادفی و شبکه منظم به ترتیب ۲/۸۱، ۲/۸۹، ۲/۸۳، ۲/۸۳ و ۲/۶۹ می‌باشند. برای بهتر مشخص شدن تفاوت آماری روش‌های مختلف نمونه‌برداری، میانگین و واریانس متغیرهای محیطی نمونه‌ها نسبت به جامعه محاسبه گردیدند. به عنوان مثال اختلاف میانگین شاخص خیسی در تکنیک‌های مربع لاتین، تصادفی، فازی-کی‌مینز، شبکه منظم، چاردرخت مرکزی و چاردرخت تصادفی با جامعه به ترتیب ۰/۸ درصد، ۲/۲ درصد، ۱/۷ درصد، ۱/۴ درصد، ۰/۱۵ درصد، ۵/۶ درصد و ۱/۰۹ درصد می‌باشد. همان‌طور که ملاحظه می‌گردد، کمترین اختلاف مربوط به تکنیک چاردرخت مرکزی می‌باشد. در حقیقت الگوی نمونه‌برداری چاردرخت مرکزی طوری نقاط نمونه را از جامعه انتخاب کرده است که توزیع نمونه‌گیری در هیستوگرام یکنواخت باشد اما در روش‌های دیگر شاهد نمونه‌برداری بیشتر یا کمتر در بعضی از قسمت‌های هیستوگرام‌ها و یا به نوعی اریب هستیم. درک و تلقی عمومی از نمونه‌برداری خوب آن است که می‌بایستی پراکنش جغرافیایی خوبی از نقاط نمونه‌برداری وجود داشته باشد؛ اما روش فازی کی‌مینز و شبکه منظم با این که توزیع جغرافیایی مناسبی را در منطقه مورد مطالعه نشان می‌دهد، اما نتوانسته پوشش مناسبی بر روی متغیرهای کمکی ایجاد کند. توانایی بالای چاردرخت مرکزی در انتخاب محل‌های نمونه‌برداری مناسب، ما را در مدل‌سازی مناسب و در نهایت تهیه نقشه رقومی خاک کمک خواهد کرد. این نتایج نشان داد که همیشه نمونه‌برداری با توزیع جغرافیایی مناسب تمام ویژگی‌های منطقه را مشخص نمی‌کند. میناسنی و همکاران (۱۵) نیز از این تکنیک جهت طراحی الگوی نمونه‌برداری در منطقه‌ی هانتز ولی^۱ استفاده کردند و نتایج آماری متغیرهای محیطی در محل‌های منتخب را با جامعه مقایسه کردند.

واریانس ثابت شده است که نشان‌دهنده واریانس یکنواخت در تمام سلول‌ها می‌باشد. از این نمودار می‌توان تعداد نواحی تقسیم‌شده را بهینه کرد. بعد از بهینه‌سازی سلول‌ها، تعداد نمونه مورد نیاز از مراکز سلول (چاردرخت منظم) و یا به شکل تصادفی (چاردرخت تصادفی) انتخاب شدند (شکل ۹).

برای بررسی و انتخاب روش نمونه‌برداری مناسب، توزیع هیستوگرامی داده‌های محیطی استفاده شده توسط مدل‌ها (جدول ۲) در محل‌های منتخب شده رسم گردید که نتایج در شکل (۸) ارائه شده است. همان‌طور که مشاهده می‌شود تفاوت قابل توجهی بین روش‌های مختلف نمونه‌برداری در توزیع هیستوگرامی وجود ندارد؛ اما به نظر می‌رسد که توزیع هیستوگرامی متغیرهای انتخاب شده توسط چاردرخت مرکزی بیشترین تبعیت را از توزیع جامعه دارند. برای بررسی بیشتر می‌توان جدول ۳ را ملاحظه کرد. نتایج نشان داد که پارامترهای آماری متغیرهای محیطی در محل‌هایی که توسط تکنیک چاردرخت مرکزی انتخاب شده‌اند، بیشترین شباهت را با جامعه مورد نظر دارند. به عنوان مثال، در مورد پارامتر کمکی شاخص خیسی مقدار میانگین متغیرهای محیطی انتخاب شده توسط این تکنیک (۶/۴۰) و کل فضای نمونه‌برداری (۶/۳۹) بیشترین شباهت را با همدیگر دارند. همچنین ضریب تغییرات شاخص خیسی در نمونه‌های چاردرخت مرکزی و جامعه به ترتیب ۳۶/۸۷ و ۳۶/۹۹ می‌باشد. این درحالی است که میانگین و ضریب تغییرات شاخص خیسی در نمونه‌های تصادفی به ترتیب ۶/۲۵ و ۳۴/۵۸ به دست آمده است، که نشان دهنده فاصله‌ی پارامترهای آماری نمونه با جامعه دارد. به‌طور مشابه مقدار میانگین شاخص جهت شیب در تکنیک چاردرخت مرکزی (۲/۷۶) بیشترین شباهت را با جامعه اصلی (۲/۷۸) دارد. همچنین با دقت بیشتر در جدول (۲) متوجه می‌شویم که بعد از مدل چاردرخت مرکزی، روش مربع لاتین نتوانسته است تا بهترین پوشش را از داده‌های محیطی در مقایسه با کل منطقه‌ی مورد مطالعه از خود نشان دهد. به

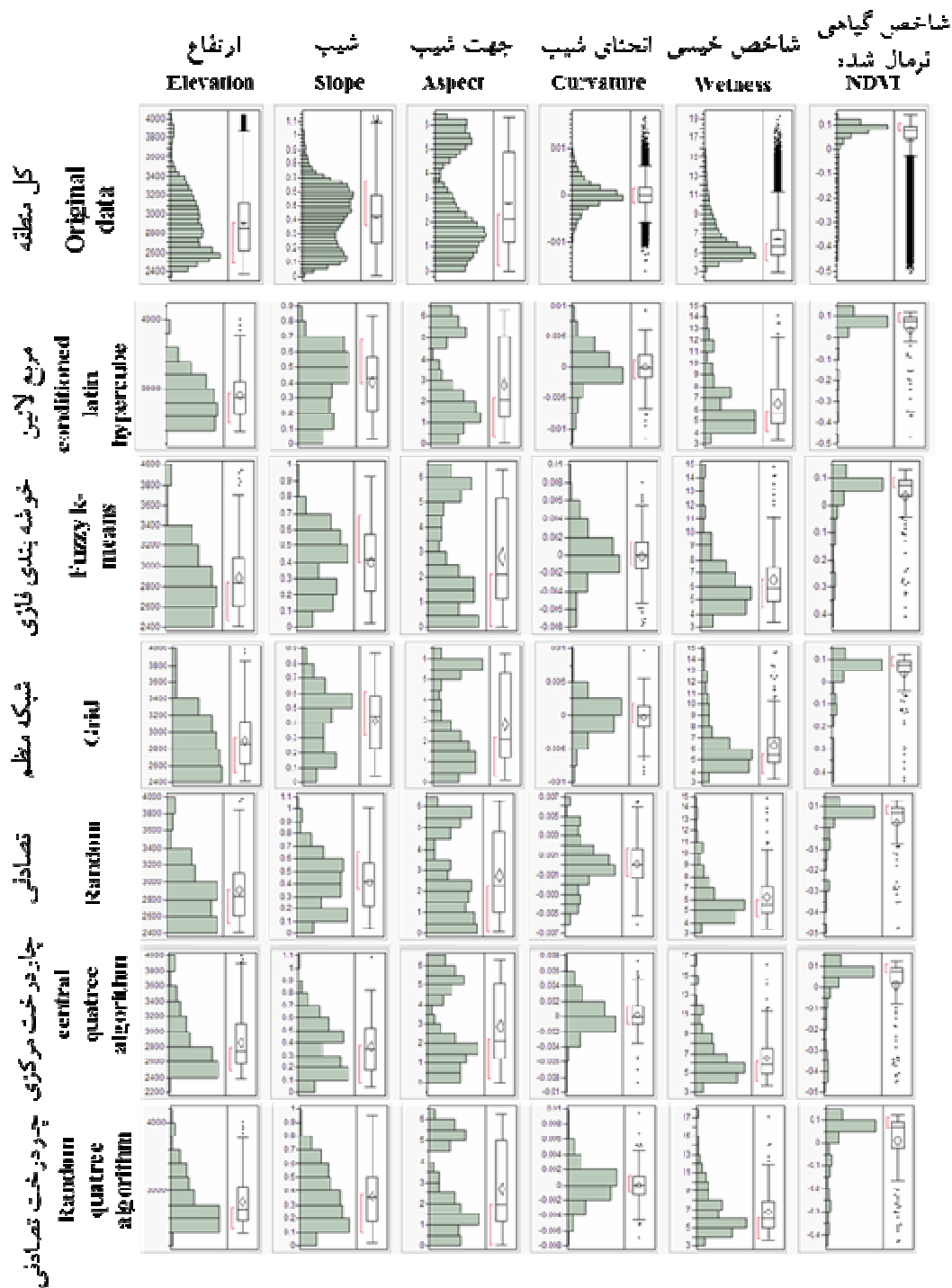
1- Hunter Valley

جدول (۳) توزیع آماری داده‌های محیطی با توجه به روش‌های مختلف نمونه‌برداری

ضریب تغییرات Coefficient of variation	واریانس Variance	حداکثر Max	حداقل Min	انحراف معیار Standard Deviation	میانگین Mean	میانه Median		
70.02	3.80	6.28	0.00	1.49	2.78	2.11	داده اصلی ^۱	جهت شیب Aspect
69.81	3.83	6.27	0.02	1.95	2.81	2.12	مربع لاتین ^۲	
71.53	3.74	6.24	0.05	1.99	2.89	2.24	تصادفی ^۳	
73.36	3.32	6.26	0.01	2.07	2.83	2.09	فازی کی مینز ^۴	
71.69	3.04	6.20	0.10	2.01	2.83	2.08	شبکه منظم ^۵	
69.95	3.81	6.26	0.00	1.91	2.76	2.10	چاردرخت مرکزی ^۶	
72.73	3.84	6.21	0.02	1.96	2.69	1.98	چاردرخت تصادفی ^۷	
11.64	114493.71	4045.19	2378.50	338.33	2905.48	2857.42	داده اصلی	ارتفاع Elevation
11.43	114803.20	3999.55	2397.02	326.26	2904.93	2871.03	مربع لاتین	
11.60	132407.49	3975.16	2405.11	366.61	2908.13	2897.56	تصادفی	
11.49	110255.63	3940.29	2403.97	332.00	2888.46	2840.16	فازی کی مینز	
11.11	113050.53	3987.55	2408.59	339.22	2901.95	2883.89	شبکه منظم	
11.68	114549.98	3987.56	2395.18	336.87	2904.38	2852.73	چاردرخت مرکزی	
12.79	133065.71	4003.98	2393.80	364.78	2850.17	2740.27	چاردرخت تصادفی	
50.60	0.04	1.14	0.03	0.20	0.41	0.42	داده اصلی	شیب Slope
49.90	0.04	0.83	0.02	0.20	0.41	0.42	مربع لاتین	
48.92	0.04	1.00	0.05	0.22	0.41	0.41	تصادفی	
50.05	0.04	0.92	0.02	0.20	0.40	0.41	فازی کی مینز	
49.32	0.04	1.00	0.04	0.20	0.41	0.44	شبکه منظم	
50.25	0.04	0.94	0.03	0.20	0.41	0.42	چاردرخت مرکزی	
50.67	0.04	1.00	0.02	0.20	0.35	0.35	چاردرخت تصادفی	
36.99	5.59	19.44	2.91	2.36	6.39	5.57	داده اصلی	شاخص خبسی Wetness Index
36.57	5.62	14.06	3.30	2.33	6.44	5.73	مربع لاتین	
34.58	4.68	14.90	3.38	2.16	6.25	5.98	تصادفی	
36.52	5.65	14.79	3.34	2.37	6.50	5.88	فازی کی مینز	
34.58	5.75	14.58	3.30	2.39	6.48	5.31	شبکه منظم	
36.87	5.62	16.07	3.62	2.34	6.40	5.57	چاردرخت مرکزی	
35.10	5.63	17.04	3.56	2.37	6.46	5.98	چاردرخت تصادفی	
3425.28	0.00	0.01	-0.01	0.002	0.00	0.00	داده اصلی	انحنای شیب Curvature
3762.49	0.00	0.009	-0.01	0.002	0.00	-0.00	مربع لاتین	
2428.00	0.00	0.006	-0.005	0.002	0.00	0.00	تصادفی	
2315.12	0.00	0.008	-0.00	0.003	0.00	-0.00	فازی کی مینز	
1338.55	0.00	0.01	-0.00	0.002	0.00	0.00	شبکه منظم	
3444.48	0.00	0.007	-0.008	0.002	0.00	-0.00	چاردرخت مرکزی	
4191.08	0.00	0.009	-0.006	0.002	0.00	0.00	چاردرخت تصادفی	
373.70	0.01	0.13	-0.47	0.11	0.03	0.07	داده اصلی	NDVI
351.87	0.01	0.11	-4.82	0.11	0.03	0.07	مربع لاتین	
313.65	0.01	0.12	-0.47	0.12	0.02	0.07	تصادفی	
343.23	0.01	7.68	-0.40	0.11	0.03	0.07	فازی کی مینز	
317.42	0.01	0.12	-0.44	0.11	0.03	0.07	شبکه منظم	
363.29	0.01	0.12	-0.45	0.11	0.03	0.07	چاردرخت مرکزی	
361.73	0.01	0.12	-0.43	0.13	0.03	0.06	چاردرخت تصادفی	

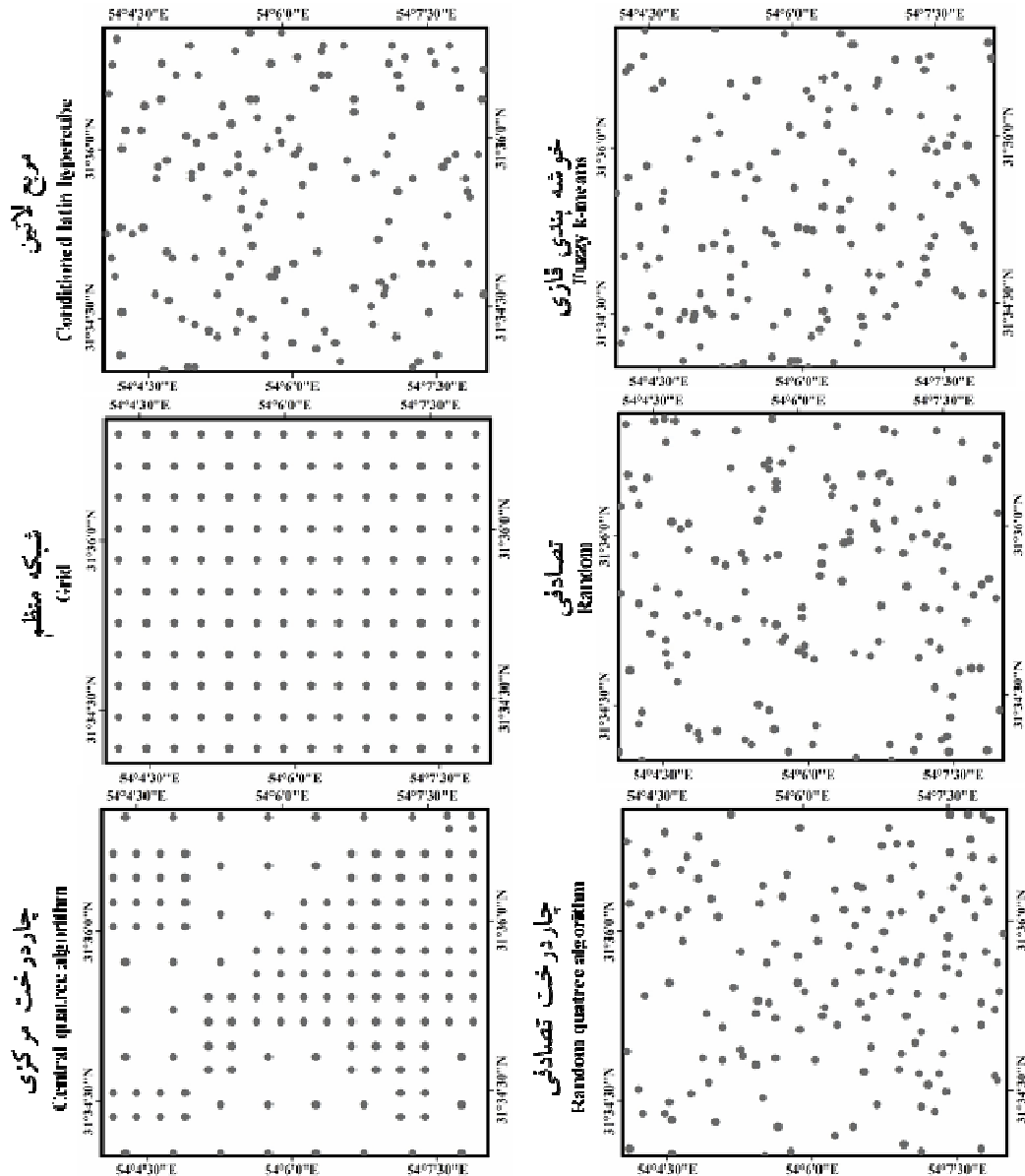
^۱: Original data; ^۲: conditional hyper cube; ^۳: random; ^۴: Fuzzy k-means; ^۵: grid; ^۶: central quatree algorithm; ^۷: random quatree algorithm

تقی زاده مهرجردی: تعیین الگوی نمونه برداری مکانی با استفاده ...



شکل (۸) هیستوگرام داده‌های محیطی با توجه به روش‌های مختلف نمونه برداری

Figure (8) Histogram of auxiliary variables according to the different sampling methods



شکل (۹) توزیع جغرافیایی محل‌های نمونه‌برداری با توجه به تکنیک‌های مختلف

Figure (9) Spatial distribution of sampling points according to the different sampling methods

ولی به وسعت ۱۴ کیلومتر مربع در کشور استرالیا اقدام به مقایسه‌ی چهار روش نمونه‌برداری مربع لاتین، کی‌میتز، مولفه‌های اصلی و تصادفی نمودند. ایشان از لایه رقومی ارتفاع، باندهای سوم و چهارم تصویر ماهواره‌ای ETM⁺ و واحدهای یکنواخت خاک/زمین‌نما به عنوان ورودی-های مدل مربع لاتین استفاده کردند. نتایج ایشان نشان داد

آن‌ها نشان دادند که تکنیک چاردرخت مرکزی به خوبی محل‌های نمونه‌برداری را در منطقه مورد مطالعه توزیع کرده است و دقت آن از روش‌هایی مانند تصادفی و فازی کی‌میتز بیشتر است. نتایج تحقیق حاضر همچنین نشان داد که تکنیک مربع لاتین نیز دارای دقت مناسب می‌باشد. میناسنی و مک‌برانتی (۱۴) نیز در منطقه هانتر

تقی زاده مهرجردی: تعیین الگوی نمونه برداری مکانی با استفاده ...

تصاویر ماهواره) در دقت پراکنش مکانی نمونه‌ها مورد بررسی قرار گیرد.

سپاس‌گزاری

بدین وسیله، از دانشگاه اردکان، جهت تأمین بودجه و امکانات لازم برای این تحقیق، حاضر سپاس‌گزاری می‌شود.

که روش مربع لاتین به خوبی هیستوگرام داده‌های محیطی در منطقه را در بر گرفته است و نمونه‌برداری کی‌میز تنها توزیع جغرافیایی مناسبی دارد. همچنین برونگارد و بوتینگر^۱ (۱) نیز شباهت توزیع فراوانی روش مربع لاتین به متغیرهای جامعه را دلیل برتری نسبت به سایر روش‌های نمونه‌برداری بیان کرده‌اند. توماس^۲ و همکاران (۲۱) نیز جهت مشخص کردن محل‌های نمونه‌برداری در منطقه‌ای به وسعت ۲۳۰۰۰ هکتار در کشور فیلیپین، اقدام به مقایسه روش‌های تصادفی، شبکه‌ای و مربع لاتین کردند. نتایج ایشان حاکی از برتری مربع لاتین نسبت به سایر روش‌ها بود. رودر^۳ و همکاران (۱۸) نیز به نتایج مشابه‌ای در مورد مربع لاتین دست یافتند.

نتیجه‌گیری

اجرای تکنیک‌های نقشه‌برداری رقومی و رسیدن به نتایج دقیق، نیاز به اطلاعات ورودی دقیق با تراکم مناسب از منطقه‌ی مورد مطالعه دارد. اطلاعات ورودی شامل اطلاعات نقطه‌ای (خاک‌رخ‌ها)، چندوجهی (نقشه‌های چندوجهی از جمله زمین‌شناسی، ژئومرفولوژی و کاربری اراضی) و سلولی (یا پیکسلی شامل تصاویر ماهواره‌ای) می‌باشد که هر یک به عنوان متغیر پیش‌بینی کننده و به عبارتی نماینده‌ای از عامل خاک‌سازی مربوطه وارد مدل می‌گردد. نتایج نشان داد که تکنیک چاردرخت مرکزی به خوبی قادر به تعیین توزیع محل‌های نمونه‌برداری در منطقه‌ی مورد مطالعه می‌باشد، به طوری که فراوانی محل‌های منتخب به فراوانی کل متغیرهای کمکی بسیار شبیه می‌باشد. همچنین همیشه توزیع جغرافیایی مناسب نشان‌دهنده‌ی الگوی نمونه‌برداری مناسب نمی‌باشد، بنابراین پیشنهاد می‌شود که در مطالعات آتی نقشه‌برداری خاک از این تکنیک استفاده شود. همچنین پیشنهاد می‌گردد که در مطالعات آتی اثر اندازه پیکسل داده‌های محیطی (مدل رقومی ارتفاع و

1- Brungard and Boettinger

2- Thomas *et al.*

3- Roudier

منابع

1. Brungard, C.W., and Boettinger, J.L. 2010. Conditioned Latin Hypercube Sampling: Optimal Sample Size for Digital Soil Mapping of Arid Rangelands in Utah, USA. In Boettinger, J.L., Howell, D.W., Moore, A.C., Hartemink, A.E., and Kienast-Brown, S. (eds), Digital Soil Mapping: Bridging Research, Environmental Application, and Operation. Springer, Amsterdam, Netherlands. pp: 67-75.
2. Brus, D.J., de-Gruijter, J.J., and van-Groenigen, J.W. 2006. Designing spatial coverage samples using the k-means clustering algorithm. In Lagacherie, P., McBratney, A.B., Voltz, M. (eds), Digital Soil Mapping: An Introductory Perspective. Elsevier, Amsterdam, Netherland. pp: 183-192.
3. Brus, D.J., Spatjens, L.E.E.M., and de-Gruijter, J.J. 1999. A sampling scheme for estimating the mean extractable phosphorus concentration of fields for environmental regulation. *Geoderma*, 89: 129–148.
4. Gessler, P.E., Moore, I.D., McKenzie, N.J., and Ryan, P.J. 1995. Soil-landscape modeling and spatial prediction of soil attributes. *International Journal of Geographical Information Science*, 9: 421–432.
5. Hengl T., Rossiter, D.G., and Stein, A. 2003. Soil sampling strategies for spatial prediction by correlation with auxiliary maps. *Geoderma*, 120: 75–93.
6. Hengl, T., Huvelink, G.B.M., and Stein, A. 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma*, 120: 75– 93.
7. Heuvelink, G., Brus, D., and de-Gruijter, J. 2007. Optimisation of sample configurations for digital soil mapping with universal kriging. In Lagacherie, P., McBratney, A.B., Voltz, M. (eds), Digital Soil Mapping: An Introductory Perspective. Elsevier, Amsterdam, Netherland. pp: 137-152.
8. Lark, R.M. 2000. Designing sampling grids from imprecise information on soil variability, an approach based on fuzzy kriging variance. *Geoderma*, 98: 35–59.
9. McBratney, A.B., Mendonça-Santos, M.L., and Minasny, B. 2003. On digital soil mapping. *Geoderma*, 117: 3-52.
10. McBratney, A.B. and Webster, R. 1981. The design of optimal sampling schemes for local estimation and mapping of regionalized variables. *Computer and Geography*, 7: 331–334.
11. 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, UK. pp. 101–110.
12. McKenzie, N.J. and Ryan, P.J. 1999. Spatial prediction of soil attributes using terrain analysis. *Geoderma*, 89: 67–94.

13. Minasny, B. and McBratney, A.B. 2002. FuzME version 3.0, Australian Centre for Precision Agriculture. McMillan Building A05, The University of Sydney, NSW 2006. <http://www.usyd.edu.au/su/agric/acpa>.
14. Minasny, B. and McBratney, A.B. 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computer and Geography*, 32: 1378-1388.
15. Minasny, B., McBratney, A.B., and Wavoort, D. 2007. The variance quadtree algorithm: Use for spatial sampling design. *Computer and Geography*, 33: 383–392.
16. Minasny, B., McBratney A.B., and Hartemink, A.E. 2010. Global pedodiversity, taxonomic distance, and the World Reference Base. *Geoderma*, 155: 132-139.
17. Rossiter, D.G. 2000. Methodology for soil resource inventories, 2nd revised version, Soil Science Division, International institute for Aerospace Survey and Earth Science (ITC). 132 pp.
18. Roudier, P., Hewitt, A.E., and Beaudette, D.E. 2012. A conditioned Latin hypercube sampling algorithm incorporating operational constraints. In Minasny, B., Malone, B.P., and McBratney, A. (eds), *Digital Soil Mapping: Digital Soil Assessments and Beyond*. Taylor and Francis Group, London, UK. pp: 227-231
19. Samet, H. 1990. *The Design and Analysis of Spatial Data Structures*. Addison Wesley Publishing Company, Inc., Reading. 493pp.
20. Simbahan, G.C. and Dobermann, A. 2006. Sampling optimization based on secondary information and its utilization in soil carbon mapping. *Geoderma*, 133: 345–362.
21. Thomas, M., Odgers, N.P., Ringrose-Voase, A., Grealish, G., Glover, M., and Dowling, T. 2012. Soil survey design for management-scale digital soil mapping in a mountainous southern Philippine catchment. In Minasny, B., Malone, B.P., and McBratney, A. (eds), *Digital Soil Mapping: Digital Soil Assessments and Beyond*. Taylor and Francis Group, London, UK. pp: 233-238.
22. Van Groenigen, J.W., Siderius, W., and Stein, A. 1999. Constrained optimisation of soil sampling for minimisation of the kriging variance. *Geoderma*, 87: 239–259.