

تعیین تراکم نقاط و روش مناسب نمونه‌برداری برای ارزیابی

شوری خاک پیش از نمونه‌برداری

یوسف هاشمی نژاد، مهدی همایی¹ و علی اکبر نوروزی

دانشجوی دکتری دانشگاه تربیت مدرس؛ y.hashemi@modares.ac.ir

استاد دانشگاه تربیت مدرس؛ mhomaee@modares.ac.ir

استادیار پژوهشکده حفاظت خاک و آبخیزداری؛ noroozi@itc.nl

دریافت: 93/8/20 و پذیرش: 94/7/11

چکیده

ارزیابی دقت نمونه‌برداری پیش از انجام نمونه‌برداری از اهمیت زیادی برخوردار است. از آن‌جا که انتظار می‌رود بین متغیرهای محیطی کمکی و ویژگی‌های خاک روابطی وجود داشته باشد، نقاط نمونه‌برداری مناسب نقاطی است که توزیع تجمعی احتمال متغیرهای محیطی کمکی را به گونه‌ای مطلوب شبیه‌سازی کند. از آن‌جا که در این پژوهش برای تعیین نقاط نمونه‌برداری از باندهای تصویر ماهواره لندست و مدل رقومی ارتفاع به عنوان ورودی‌ها استفاده شد، سه روش نمونه‌برداری تصادفی، شبکه منظم و فرامکعب لاتین در سه تراکم 100، 50 و 25 نقطه برای شبیه‌سازی توزیع احتمال تجمعی باندهای تصویر ماهواره لندست و مدل رقومی ارتفاع در محدوده‌ای به وسعت حدود 52 هزار هکتار استفاده شد. سپس، توزیع احتمال شبیه‌سازی شده با توزیع جامعه اصلی و شوری خاک مورد مقایسه قرار گرفت. نتایج نشان داد که در تراکم نمونه‌برداری 100 و 50 نقطه، تفاوت دقت روش‌های نمونه‌برداری چشمگیر نبوده و با کاهش تراکم نمونه‌برداری میانگین مربعات خطای باقیمانده افزایش می‌یابد. حال آن‌که در تراکم 25 نقطه نمونه‌برداری، مقدار میانگین مربعات خطای باقیمانده در روش‌های فرامکعب لاتین، شبکه منظم و کاملاً تصادفی به ترتیب 4/783 و 17/15 درصد بود. به این ترتیب نتایج نشان داد که روش فرامکعب لاتین در تراکم نمونه‌برداری کم قادر به شبیه‌سازی دقیق‌تر توزیع احتمال ورودی است. علت این تفاوت آن است که در روش‌های کاملاً تصادفی و شبکه منظم احتمال نمونه‌برداری از نقاط با فراوانی کم، بسیار اندک است. مقدار ریشه میانگین مربعات خطا در برآورد شوری خاک جامعه اصلی در تراکم 100 و 25 نقطه به روش فرامکعب لاتین به ترتیب 4/22 و 8/66 دسی زمینس بر متر بود. هرچند مقداری بیش از خطای متغیرهای اولیه یعنی باندهای تصویر و مدل رقومی ارتفاع است، لیکن با دقتی مناسب توزیع احتمال شوری خاک در جامعه اصلی را شبیه‌سازی کرده است. در نتیجه مناسب‌ترین روش و تراکم نمونه‌برداری در این پژوهش روش فرامکعب لاتین با تراکم 25 نقطه بود.

واژه‌های کلیدی: تابع توزیع احتمال، شبکه نمونه‌برداری منظم، فرامکعب لاتین

¹ نویسنده مسئول، آدرس: اتوبان تهران - کرج، کیلومتر 17، بعد از پیکان شهر، بلوار پژوهش، دانشکده کشاورزی، گروه خاکشناسی

مقدمه

هدف از نمونه‌برداری خاک جمع‌آوری اطلاعات در مورد یک خاک است. خاک‌ها با انواع مختلفی از تغییرپذیری مواجه هستند و برای نشان دادن کامل یک جامعه خاک باید تنوع تغییرپذیری را در نظر داشت. به ندرت نمونه، تمام حجم خاک مورد بررسی را تشکیل می‌دهد (پترسون و کلاین، 1986). در نتیجه برای پوشش دادن کامل انواع تغییرپذیری خاک و همچنین صرفه‌جویی در وقت و هزینه به‌ناچار باید راهبرد مناسب نمونه‌برداری را انتخاب نمود.

راهبرد نمونه برداری شامل گردآوری تعدادی مشاهده که نمونه نامیده می‌شوند، از یک جمعیت است که در واقع مجموع حالات محتمل مشاهدات است (هنسون و گرتن، 1996). هدف نمونه‌برداری تخمین پارامترها با دقت قابل قبول و کمترین هزینه است (پترسون و کلاین، 1986). هر پارامتری در طبیعت دارای یک توزیع احتمال تصادفی و معمولاً زنگوله‌ای شکل است. بالاترین دقت در تخمین پارامترها، زمانی حادث می‌شود که توزیع نمونه‌ها نزدیک به توزیع جامعه باشد. بدیهی است با افزایش تعداد مشاهدات، توزیع نمونه و جامعه به هم نزدیک می‌شوند ولی هم‌زمان هزینه نمونه‌برداری افزایش می‌یابد. بنابراین لازم است پیش از نمونه‌برداری مناسب‌ترین تراکم و موقعیت نقاط نمونه‌برداری مشخص شود. در یک شبیه‌سازی نمونه‌برداری از طریق درآوردن تدریجی نمونه در هر بار سعی و خطا از توزیع احتمال انجام می‌شود. با انجام سعی و خطا به میزان کافی، توزیع احتمال مقادیر نمونه‌برداری شده به توزیع احتمال جامعه نزدیک می‌شود. آماره‌های توزیع نمونه‌برداری شده (میانگین، انحراف معیار و مقادیر بالا) تخمینی از توزیع احتمال ورودی را فراهم می‌آورد. در شبیه‌سازی به دلیل آریبی¹ ممکن است مقداری عدم صحت² ایجاد شود. این در حالی است که انتخاب نامناسب نقاط نمونه‌برداری می‌تواند خود باعث ایجاد آریبی شود (بیئتس، 1981؛ کمپتورن و آلماراس، 1986). این آریبی ممکن است در اثر انتخاب یک نمونه با قصد قبلی، انتخاب نمونه‌ها بر اساس ویژگی‌هایی که دارای همبستگی با محدوده مورد نمونه‌برداری هستند، اشتباه عمدی³ یا غیر عمدی در انتخاب نمونه، جایگزینی یک محل نمونه‌برداری با محلی دیگر و یا نقص در پوشش تمام محدوده نمونه‌برداری حادث شده باشد (هنسون و گرتن، 1996).

اصول اولیه نمونه‌برداری خاک به وسیله کلاین (1944) توضیح داده شده که از آن زمان تغییر زیادی نداشته است. در عین حال توجه فزاینده‌ای به روش‌های بررسی خطای نمونه‌برداری انجام شده است (پترسن و کالوین، 1986). حاج رسولیها و همکاران (1980) برای ارزیابی روش نمونه‌برداری مورد استفاده در کشت و صنعت نیشکر هفت تپه نشان دادند که ساختار واریانس شوری خاک در دو مکان از سه مکان مورد بررسی دارای وابستگی مکانی است و در نتیجه از روش‌های زمین‌آماری برای ترسیم نقشه شوری خاک استفاده شد. در عین حال، ارزیابی روش نمونه‌برداری پیش از انجام نمونه‌برداری موضوعی است که در مطالعات خاکشناسی کمتر به آن توجه شده است. به هنگام ارزیابی روش نمونه‌برداری، عامل مهم، آزمون تعداد سعی و خطای مورد نیاز برای خلق مجدد و دقیق توزیع احتمال جامعه از طریق نمونه‌برداری است. چنانچه یک روش نمونه‌برداری نیازمند سعی و خطای بیشتر و زمان اجرای بیشتر برای تخمین توزیع احتمال ورودی باشد، آن روش کارایی کمتری دارد. دو روش مونت کارلو⁴ و فرا مکعب لاتین⁵ از نظر تعداد سعی و خطای مورد نیاز برای تخمین توزیع احتمال ورودی با هم تفاوت دارند. نمونه‌برداری به روش مونت کارلو برای تخمین توزیع احتمال ورودی نیاز به تعداد زیادی نمونه‌برداری دارد، به‌ویژه هنگامی که توزیع احتمال ورودی به شدت چولگی داشته باشد و یا اینکه دارای دنباله‌هایی با احتمال کم باشد. روش نمونه‌برداری مربع لاتین، در عوض، تعداد نمونه مورد نیاز را تحمیل می‌کند و در نتیجه سریع‌تر روی مقادیر آماره‌های واقعی توزیع ورودی همگرا می‌شود (هانچل و کاراف، 2009). برآورد ویژگی‌ها و یا کلاس‌های خاک بر مبنای ایجاد روابطی بین مقادیر مشاهده شده و متغیرهای کمکی محیطی است (مک براتنی و همکاران، 2003). از آنجا که ویژگی‌های خاک می‌توانند در داده‌های طیفی منعکس شوند و یا ممکن است به صورت یک شاخص چند جزئی با خصوصیات طیفی رابطه داشته باشند، تصاویر ماهواره‌ای می‌توانند برای تعیین نقاط نمونه‌برداری داده‌های کمکی مؤثر کنند. استفاده از داده‌های کمکی برای تولید نقشه ویژگی‌های خاک در سال‌های اخیر رو به گسترش است (هنگل و همکاران، 2000). این داده‌های کمکی می‌توانند از تحلیل رقومی عوارض زمینی⁷

⁴ Monte Carlo

⁵ Latin Hypercube

⁶ Hantschel and Kauerauf

⁷ Digital Terrain Analysis

¹ Bias

² Inaccuracy

³ Deliberated bias

شبکه منظم، تصادفی و فرامکعب لاتین در سه تراکم نقطه 50، 100 و 25 نقطه، موقعیت نقاط نمونه‌برداری تعیین شد.

برای تهیه نقشه تغییرات شوری خاک در دو سال متوالی (1392-1393) با استفاده از روش فرامکعب لاتین به ترتیب موقعیت 100 و 25 نقطه تعیین و نمونه‌برداری خاک در سه عمق به فواصل 30 سانتیمتری تا عمق 90 سانتیمتر انجام شد. پس از انتقال نمونه‌ها به آزمایشگاه، اندازه‌گیری هدایت الکتریکی عصاره اشباع خاک به روش استاندارد (ریچاردز، 1954) تعیین شد. برای درونیابی تغییرات شوری خاک در محدوده مورد مطالعه از روش زمین‌آماري کریجینگ پس از بهینه‌سازی استفاده شد. پس از تهیه نقشه تغییرات شوری مقادیر شوری خاک برای تمامی پیکسل‌ها استخراج و توزیع تجمعی احتمال جامعه رسم گردید. سپس مقادیر شوری خاک در نقاط نمونه‌برداری به این توزیع منتقل و با آن مورد مقایسه قرار گرفت.

نمونه‌برداری الگوریتم مونت کارلو

نمونه‌برداری در الگوریتم مونت کارلو بر پایه روش مرسوم استوار است که در آن با استفاده از اعداد تصادفی و یا شبه تصادفی از روی توزیع احتمال تجمعی، نقاط نمونه‌برداری تعیین می‌شود. دامنه وسیعی از الگوریتم‌ها برای تولید نمونه‌های تصادفی از انواع مختلف توزیع احتمال وجود دارد. تکنیک‌های نمونه‌برداری مونت کارلو کاملاً تصادفی هستند. به این معنی که هر نمونه مشخص ممکن است در هر جایی از دامنه توزیع ورودی واقع شود. مطمئناً احتمال استخراج نمونه‌ها از قسمت‌هایی از توزیع که احتمال وقوع بیشتری دارند، بیشتر است. در توزیع تجمعی هر نمونه مونت کارلو از یک عدد تصادفی جدید بین صفر و یک استفاده می‌کند. اگر تعداد سعی و خطا به اندازه کافی باشد، روش نمونه‌برداری مونت کارلو می‌تواند توزیع ورودی را مجدداً شبیه‌سازی کند. در عین حال هنگامی که تعداد سعی و خطا کم باشد، مسئله خوشه‌بندی ایجاد می‌شود (هانچل و کاراف، 2009)

در شکل (2) هر یک از پنج نمونه استخراج شده، در میانه توزیع واقع شده است. مقادیر در دامنه‌های بیرونی‌تر توزیع در نمونه‌برداری واقع نشده‌اند و در نتیجه تأثیر آن‌ها بر نتایج نشان داده نمی‌شود. مسئله خوشه‌بندی به ویژه هنگامی تشدید می‌شود که یک توزیع احتمال دارای دنباله‌هایی با احتمال وقوع کم باشد که تأثیر اصلی بر روی نتایج می‌گذارند. وارد کردن این دنباله‌های با احتمال وقوع کم مهم است. به این منظور این دنباله‌ها نیز باید نمونه‌برداری شوند. لیکن اگر احتمال وقوع خیلی کم

(ویلسون و گالانت¹ 2000) و یا دورسنجی (اوده و مک براتی² 2000) حاصل شوند. برای تعیین شوری خاک نیز شاخص‌های مختلفی معرفی شده است (دعایی و همکاران، 2006؛ النجار و نولر، 2010؛ نوروزی و همکاران، 2012؛ اسکودیرو و همکاران، 2014). این شاخص‌ها یا مستقیماً از طریق باندهای تصاویر ماهواره‌ای به دست می‌آیند و یا به‌طور غیر مستقیم و با توجه به عوارض زمینی از مدل رقومی ارتفاع مشتق می‌شوند.

تقی زاده و همکاران (1392) برای تشخیص موقعیت نقاط نمونه‌برداری خاک با استفاده از متغیرهای کمکی، از تکنیک‌های فرامکعب لاتین و فازی کی مینز³ استفاده کردند. در این مطالعه برای تعیین تعداد و نوع متغیرهای کمکی، از تکنیک آنتروپی شانون استفاده شد و به این ترتیب 10 متغیر ورودی کمکی انتخاب شدند. بر این اساس برای مطالعه 72000 هکتار از اراضی دشت یزد- اردکان، 187 محل نمونه‌برداری انتخاب شد. همچنین تقی‌زاده و همکاران (1393) در مقایسه روش‌های مختلف انتخاب موقعیت نقاط نمونه‌برداری، گزارش کردند که روش فرامکعب لاتین از بالاترین درجه تفکیک عوارض زمینی برخوردار است.

با جمع‌بندی نتایج مطالعات انجام شده می‌توان نتیجه‌گیری کرد که با استفاده از متغیرهای کمکی می‌توان موقعیت نقاط نمونه‌برداری را تعیین نمود اما تاکنون در این مطالعات بهترین روش و تراکم نمونه‌برداری قبل از انجام آن مشخص نشده است لذا این پژوهش با هدف مقایسه سه روش مختلف نمونه‌برداری (تصادفی، شبکه منظم و فرامکعب لاتین) و تراکم‌های مختلف به منظور شبیه‌سازی توزیع احتمال ورودی (شامل باندهای تصویر لندست و مدل رقومی ارتفاع) انجام شد.

مواد و روش‌ها

این پژوهش در بخشی از اراضی دشت سبزوار- داورزن به مساحت 51936 هکتار انجام شد. شکل (1) موقعیت محدوده مورد مطالعه در استان خراسان رضوی، جنوب غرب شهرستان سبزوار را نشان می‌دهد. با توجه به این‌که پیش از نمونه‌برداری در مورد تغییرپذیری مکانی شوری خاک در منطقه اطلاعی در دست نبود، با استفاده از تصویر ماهواره‌ای و مدل رقومی ارتفاع تلاش شد تا نقاط نمونه‌برداری به‌گونه‌ای تعیین شود که بیشترین تغییرپذیری را نشان دهند. در این پژوهش، برای ارزیابی تعداد و روش‌های مختلف نمونه‌برداری با استفاده از سه روش

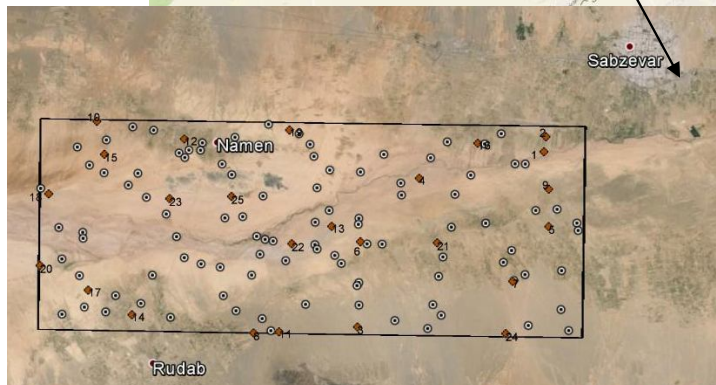
¹ Wilson and Gallant

² Odeh and McBratney

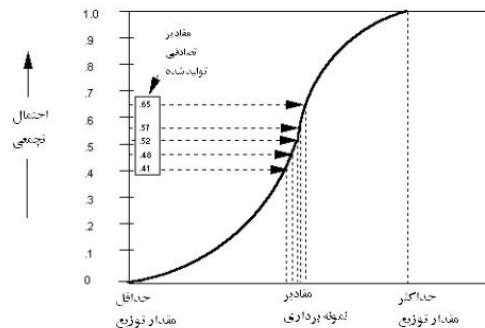
³ Fuzzy K means

تکنیک‌های نمونه برداری طبقه بندی شده همچون نمونه برداری به روش فرا مکعب لاتین شده است.

باشد، نمونه برداری مونت کارلو با تعداد کمی سعی و خطا نمی تواند تعدادی کافی از این دنباله ها را نمونه برداری کند تا توزیع احتمال را شبیه سازی کند. این مشکل منجر به

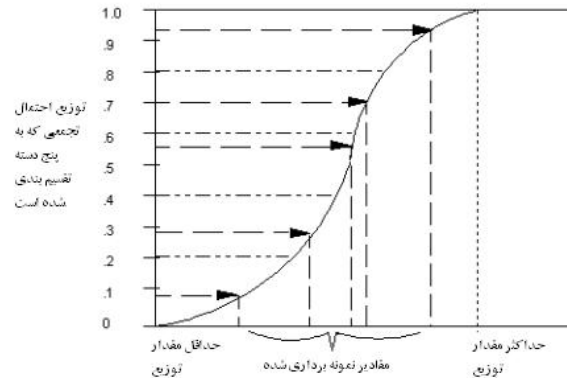


شکل 1- موقعیت محدوده مورد مطالعه واقع در دشت سبزوار - داورزن، خراسان رضوی. نقاط نمونه برداری سال اول مدور و نقاط سال دوم لوزی شکل هستند



شکل 2- پنج سعی و خطای روش مونت کارلو با خوشه بندی (تغییر یافته از هلتن و دیویس¹، 2003)

¹ Helton and Davis



شکل 3- سعی و خطاهای روش نمونه‌برداری مربع لاتین (تغییر یافته از هلتن و دیویس، 2003)

نمونه‌برداری به روش فرا مکعب لاتین

روش فرامکعب لاتین نوعی نمونه‌برداری مونت کارلو است که به منظور بازآفرینی توزیع احتمال متغیر ورودی با استفاده از سعی و خطای کمتر و کارایی بیشتر در مقایسه با روش مونت کارلو توسعه یافته است (مک-کی و همکاران¹، 1979؛ ایمان و کونور²، 1980). همان‌گونه که در شکل (3) نشان داده شده، نکته کلیدی در نمونه‌برداری به روش مربع لاتین دسته‌بندی توزیع احتمال ورودی است. بدین منظور، دسته‌بندی توزیع احتمال تجمعی ورودی بر روی مقیاس صفر تا یک به فواصل مساوی تقسیم‌بندی می‌شود. سپس، از هر فاصله یا دسته توزیع احتمال تجمعی یک نمونه برداشت می‌شود. به نمونه‌برداری اجبار وارد می‌شود تا از مقادیری در هر فاصله یک نمونه گرفته شود و در نتیجه اجبار وارد می‌شود تا توزیع احتمال ورودی را شبیه‌سازی کند.

در نمونه‌برداری از یک توزیع چند متغیره به روش فرامکعب لاتین، اندازه n نمونه از توزیع چند متغیره به نحوی استخراج می‌شود که برای هر متغیر حداکثر حاشیه طبقه‌بندی ایجاد شود (میناسنی و مک‌براتی، 2006). یک نمونه هنگامی بیشترین حاشیه طبقه‌بندی را داراست که تعداد طبقه‌ها معادل اندازه نمونه n باشد و احتمال افتادن در هر طبقه برابر با n^{-1} باشد (مک‌کی و همکاران، 1979).

روش فرا مکعب لاتین یک روش نمونه‌برداری تصادفی دسته‌بندی شده است که راهی مؤثر برای نمونه‌برداری از متغیرها از فضای توزیع چند متغیره آنها فراهم می‌آورد. این روش در دانش خاکشناسی و محیط زیست برای ارزیابی عدم قطعیت در مدل‌های پیش‌یابی (میناسنی و مک‌براتی، 2002) و در زمین‌آمار برای شبیه‌سازی

زمینه‌های تصادفی گوسی (پبسم و هیوولینک³، 1999؛ ژانگ و پیندر⁴، 2004) به کار رفته است.

میناسنی و مک براتی (2006) الگوریتم فرامکعب لاتین مشروط را برای نمونه‌برداری خاک توصیف و کاربرد آن را در نرم افزار متلب⁵ برای مطالعات خاکشناسی تشریح کرده‌اند.

در روش نمونه‌برداری فرامکعب لاتین اطلاعات کمکی شامل اطلاعات باندهای تصویر ماهواره لندست و مدل رقومی ارتفاع (DEM) بودند. برای استخراج مقادیر پارامترها در جامعه اصلی با استفاده از نرم افزار ArcGIS یک شبکه به ابعاد 30 متر در 30 متر بر روی تصویر ایجاد و عدد رقومی⁶ هر باند در وسط هر پیکسل استخراج شد. موقعیت نقاط نمونه‌برداری تصادفی و شبکه منظم نیز در همین محیط نرم‌افزاری و برای روش فرامکعب لاتین از کدهای میناسنی و مک براتی (2006)⁷ استفاده و سپس در محل نقاط مقادیر عدد رقومی تعیین شد. جهت تعیین توزیع تجمعی احتمال مقادیر استخراج شده از نرم‌افزار SPSS استفاده شد. برای مقایسه و تعیین روش و تراکم نمونه‌برداری مناسب، مقادیر پارامترهای یادشده (باندهای تصویر ماهواره لندست و DEM) در محل‌های نمونه‌برداری برای هر روش و تراکم نمونه‌برداری تعیین و با همان پارامترها در جامعه اصلی (کلیه پیکسل‌های تصویر لندست) مورد مقایسه قرار گرفت. ملاک تراکم 50 نقطه،

³ Pebesma and Heuvelink

⁴ Zhang and Pinder

⁵ Matlab

⁶ Digital Number (DN)

⁷ کدها و نرم افزارهای مربوطه در سایت زیر قابل دسترس

است:

<http://www.soil-research.com/software>

¹ McKay et al.

² Iman and Conover

همان‌گونه که در شکل (5) نشان داده شده، با کاهش تراکم نقاط نمونه برداری به 50 نقطه هر چند که باز هم هر سه روش نمونه برداری به یکدیگر و به توزیع احتمال جامعه اصلی نزدیک بودند، لیکن تفاوت بین روش‌ها را می‌توان در توزیع باند 2 و باند 7 به وضوح مشاهده کرد. در عین حال در همه این نمودارها، توزیع تجمعی احتمال به دست آمده از روش نمونه برداری فرامکعب لاتین به توزیع احتمال جامعه اصلی بسیار نزدیک و در واقع روی خط احتمال واقع شده است.

یک نقطه در هر 1000 هکتار و به این ترتیب 100 نقطه 2 برابر این تراکم و 25 نقطه نصف این تراکم بوده است. برای مقایسه میانگین نمونه‌های اخذ شده به روش‌های مختلف با میانگین جامعه اصلی از آزمون نمرات استاندارد Z استفاده شد. به این ترتیب که با استفاده از رابطه (1) نمرات استاندارد محاسبه و در سطوح احتمال 95 و 99 درصد و با استفاده از جداول مرجع با مقادیر جامعه مقایسه شد.

$$Z = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \quad (1)$$

که در آن: \bar{X} میانگین پارامتر در جامعه اصلی (کل تعداد پیکسل‌ها که در این مورد حدود 576000 پیکسل 900 متر مربعی بود)، μ میانگین حاصل از نمونه-ها، s انحراف معیار نمونه و n تعداد نمونه است. همچنین، برای آزمون برابری انحراف معیار جامعه و نمونه، از آزمون χ^2 در سطح احتمال 95 و 99 درصد استفاده شد.

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} \quad (2)$$

که در آن: χ^2 آماره کای اسکور، s انحراف معیار نمونه، n تعداد نمونه و σ انحراف معیار جامعه است. برای مقایسه کمی دقت روش‌ها و تراکم‌های نمونه برداری، میزان انحراف مقادیر توزیع احتمال تجمعی برآورد شده به کمک این روش‌ها با مقادیر جامعه اصلی مقایسه و آماره ریشه میانگین مربعات خطا¹ RMSE تعیین شد:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}} \quad (3)$$

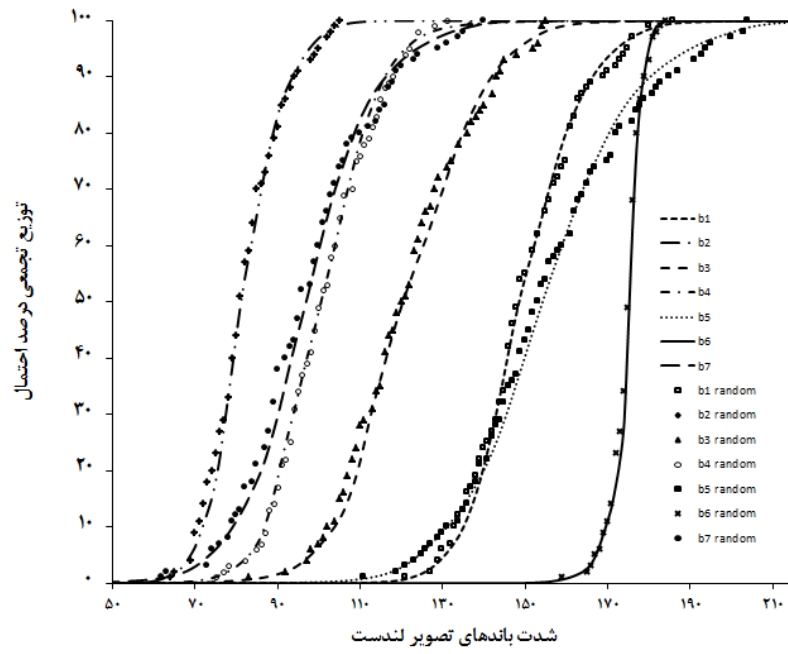
که در آن: P_i و O_i به ترتیب مقادیر پیش‌بینی و مشاهده شده درصد احتمال تجمعی پارامترها (عدد دیجیتال باندهای تصویر و DEM یا شوری خاک) در نمونه و جامعه و n تعداد نقاط نمونه برداری است.

نتایج

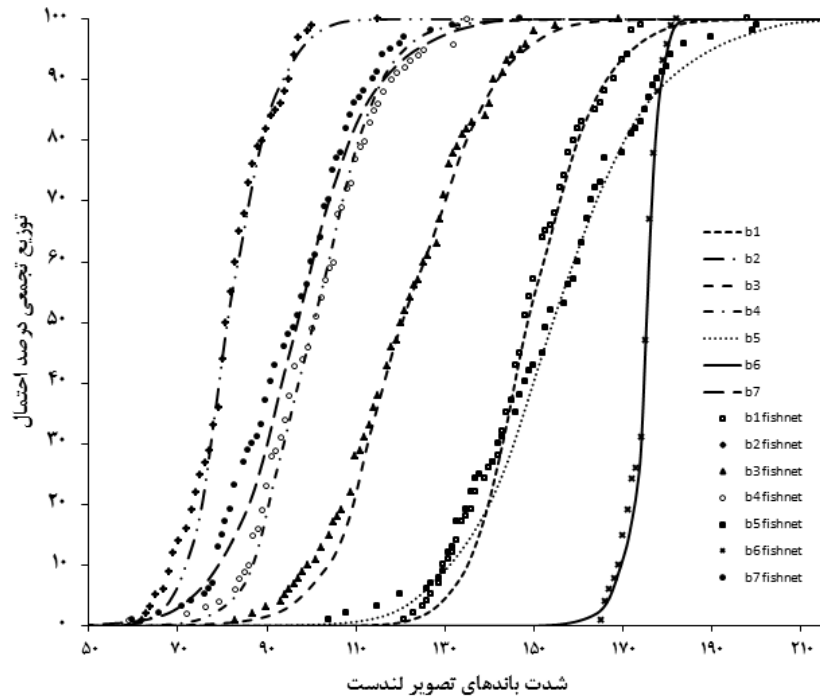
شکل (4) توزیع تجمعی احتمال باندهای هفت گانه تصویر لندست در جامعه اصلی و بازسازی شده به وسیله روش‌های نمونه برداری کاملاً تصادفی (مونت کارلو)، شبکه منظم و روش فرامکعب لاتین برای تراکم 100 نمونه را نشان می‌دهد. همان‌گونه که مشخص است شباهت بین روش‌های نمونه برداری در شبیه‌سازی توزیع احتمال تجمعی به ویژه در بخش میانی توزیع کمابیش مشابه هستند.

¹ Root Mean Square Error

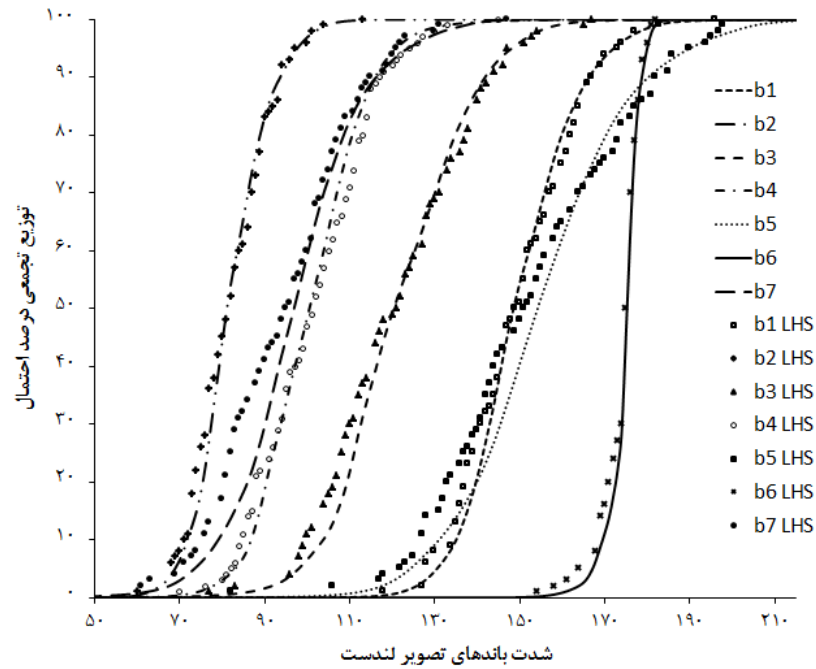
کاملاً تصادفی



شبکه منظم

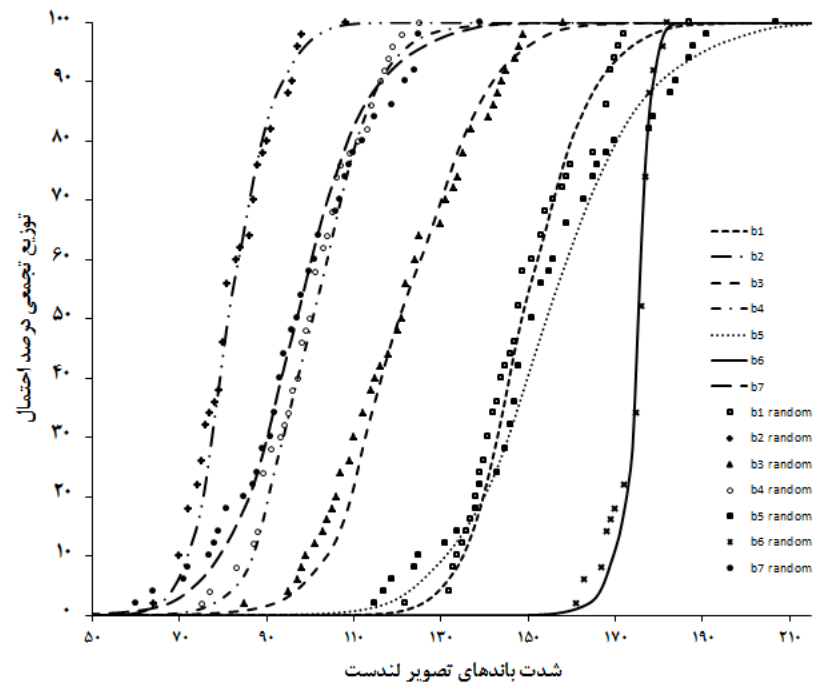


فرامکعب لاتین

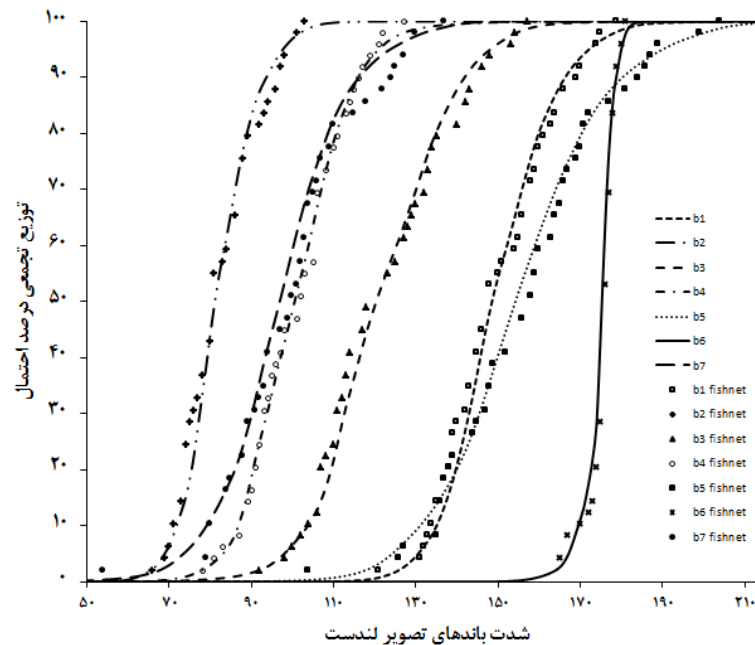


شکل 4- توزیع تجمعی احتمال جامعه اصلی (خطوط) و شبیه سازی شده (نقاط) به وسیله سه روش فرامکعب لاتین (LHS)، کاملاً تصادفی (Random) و شبکه نمونه برداری منظم (Fishnet) برای باندهای تصویر لندست در تراکم 100 نقطه نمونه برداری.

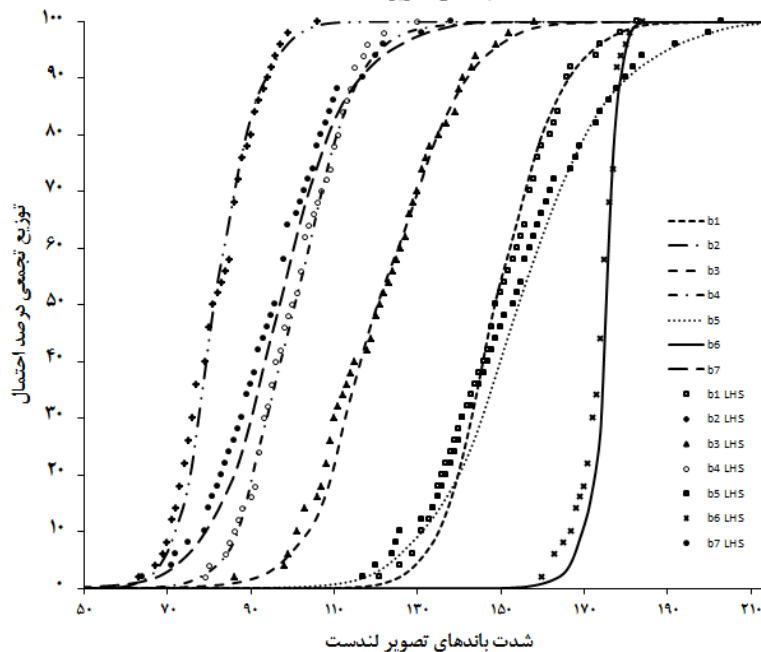
کاملاً تصادفی



شبکه منظم



فرامکعب لاتین

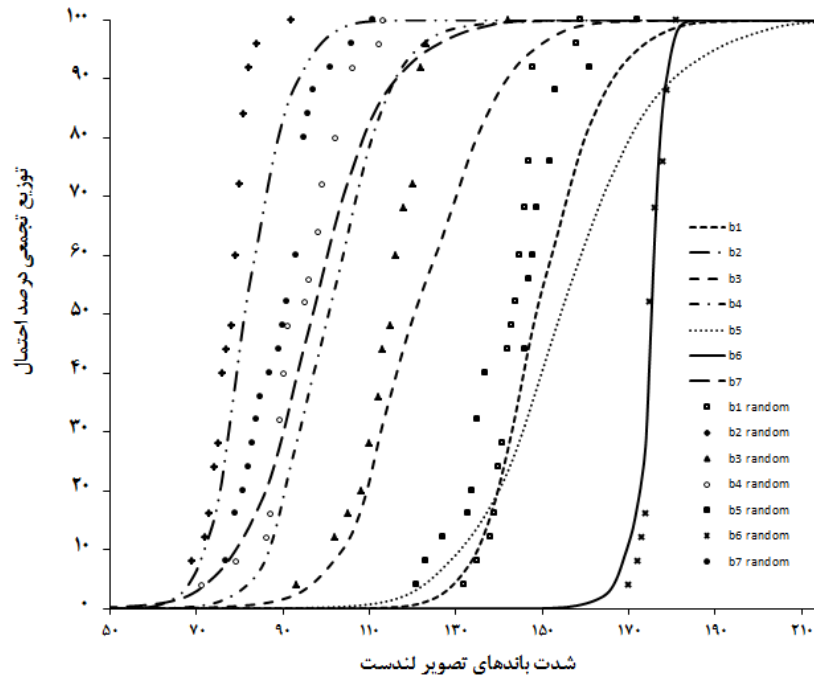


شکل 5- توزیع تجمعی احتمال جامعه اصلی (خطوط) و شبیه سازی شده (نقاط) به وسیله سه روش فرامکعب لاتین (LHS)، کاملاً تصادفی (Random) و شبکه نمونه برداری منظم (Fishnet) برای باندهای تصویر لندست در تراکم 50 نقطه نمونه برداری.

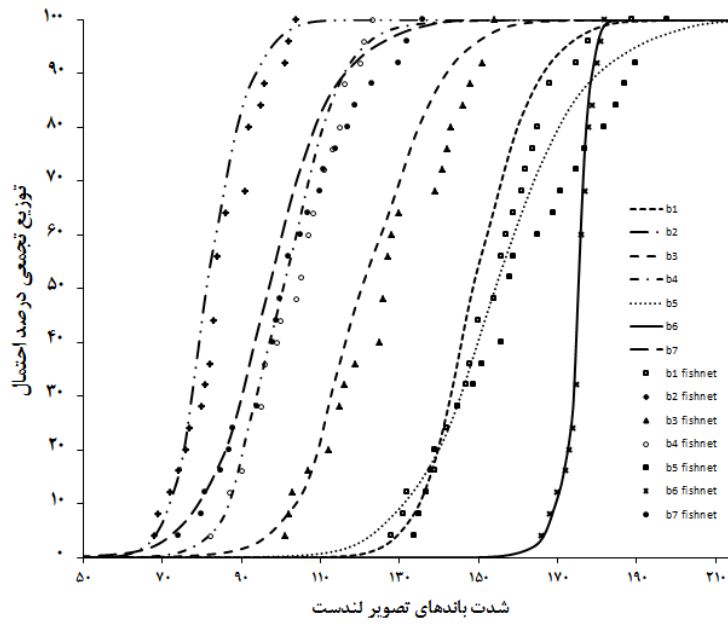
پایین تر از آن است. در حالی که روش فرامکعب لاتین در مورد تمام پارامترها روی توزیع احتمال تجمعی جامعه اصلی واقع شده است.

با کاهش بیشتر تراکم نقاط نمونه برداری به 25 نقطه، تفاوت بین روش‌های نمونه برداری بیشتر مشخص شد. همان‌گونه که در شکل (6) مشخص است، تقریباً در همه موارد توزیع نقاط روش نمونه برداری تصادفی بالاتر از توزیع تجمعی احتمال جامعه اصلی و روش شبکه منظم

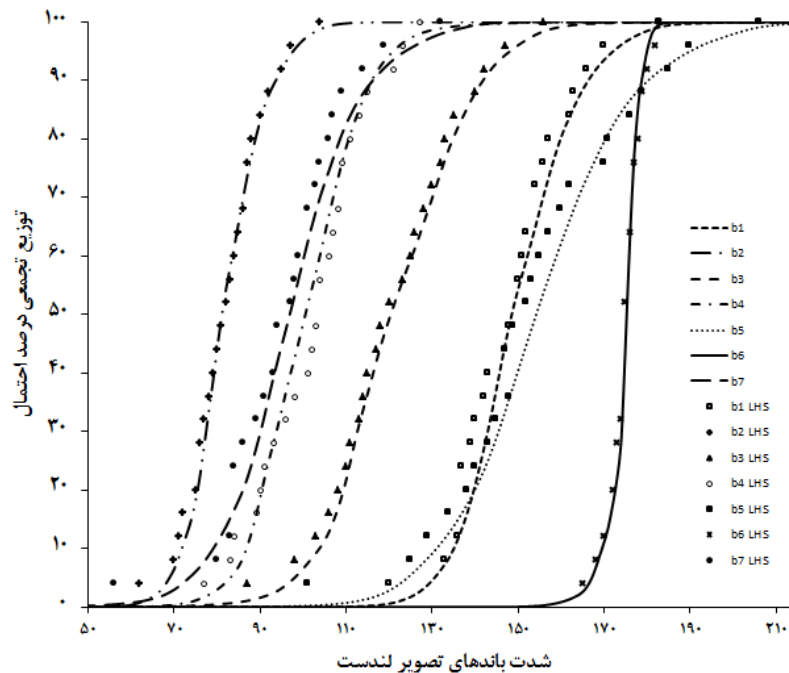
کاملاً تصادفی



شبکه منظم



فرامکعب لاتین

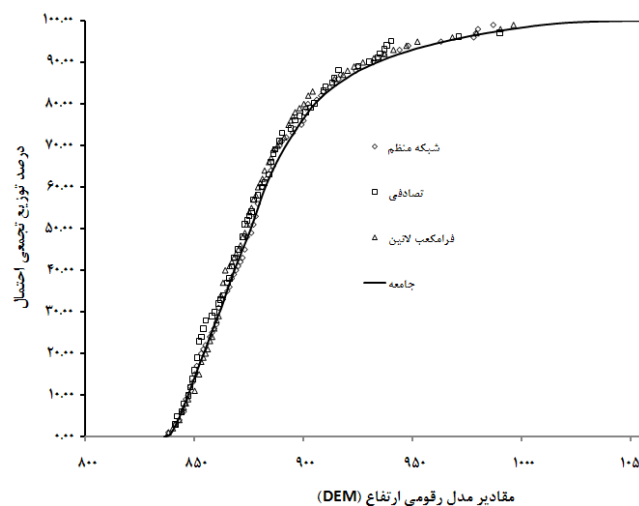


شکل 6- توزیع تجمعی احتمال جامعه اصلی (خطوط) و شبیه سازی شده (نقاط) به وسیله سه روش فرامکعب لاتین (LHS)، کاملاً تصادفی (Random) و شبکه نمونه برداری منظم (Fishnet) برای باندهای تصویر لندست در تراکم 25 نقطه نمونه برداری

داشته در حالی که نزدیک ترین توزیع به جامعه اصلی متعلق به نقاطی بودند که به روش فرامکعب لاتین نمونه برداری شدند. این نتایج مشابه نتایج تقی زاده و همکاران (1393) و نیز میناسنی و مک برانتی (2007) می باشد که در مقایسه چند روش نمونه برداری روش فرامکعب لاتین را به عنوان بهترین روش معرفی کردند.

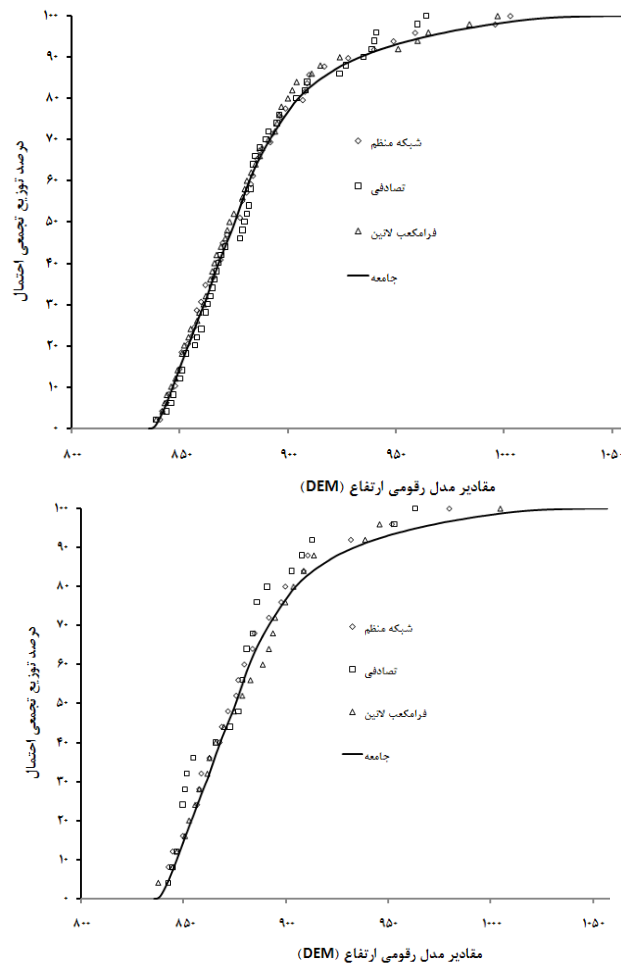
برای مدل رقومی ارتفاع نیز روندی کم و بیش مشابه در تراکم های مختلف نمونه برداری و روش های مختلف مشاهده شد (شکل 7): به گونه ای که در تراکم نمونه برداری زیاد (100 و 50 نقطه) توزیع احتمال نقاط نمونه برداری بسیار نزدیک به یکدیگر و به جامعه اصلی بودند و با کاهش تراکم به 25 نقطه در محدوده مورد مطالعه، روش های شبکه منظم و تصادفی مقداری انحراف

تراکم نمونه برداری 100 نقطه



تراکم نمونه برداری 50 نقطه

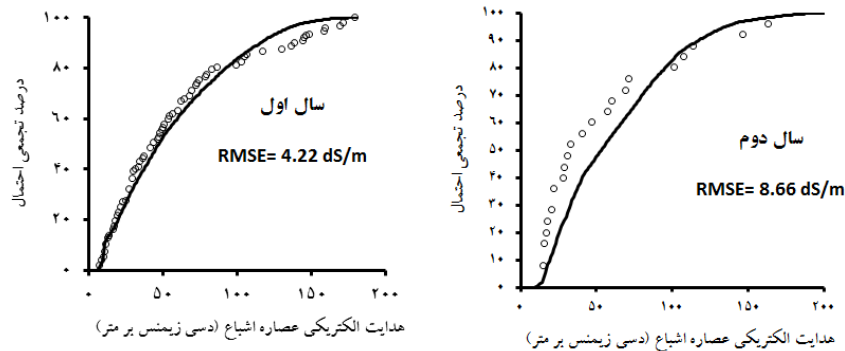
تراکم نمونه برداری 25 نقطه



شکل 7- توزیع تجمعی احتمال جامعه اصلی و شبیه سازی شده به وسیله سه روش فرامکعب لاتین (LHS)، کاملاً تصادفی (Random) و شبکه نمونه برداری منظم (Fishnet) برای مدل رقمی ارتفاع در تراکم های مختلف نمونه برداری

جامعه اصلی شوری خاک در 2 سال و مقادیر نمونه- برداری شده با تراکم 100 و 25 نقطه به روش فرامکعب لاتین نشان می دهد. در این توزیع نیز مانند باندهای تصویر و مدل رقمی ارتفاع با کاهش تعداد نقاط نمونه- برداری انحراف از توزیع جامعه اصلی افزایش می یابد.

مانند پارامترهای ورودی، شوری خاک که هدف اصلی این مطالعه بوده است در جامعه اصلی، دارای توزیع تجمعی کمابیش نرمال و در صورتی نمونه برداری موفقیت آمیز خواهد بود که توزیع نقاط نمونه برداری به توزیع جامعه نزدیک باشد. شکل 8 توزیع تجمعی احتمال



شکل 8- توزیع تجمعی احتمال شوری خاک در جامعه اصلی و نمونه های اخذ شده با تراکم 100 و 25 نقطه در سال اول و دوم

در جدول (2) مقادیر ریشه میانگین مربعات خطا RMSE برای روش‌ها و تراکم‌های مختلف نمونه برداری نشان داده شد. صحت روش‌های کاملاً تصادفی و شبکه منظم به ترتیب در تراکم‌های 100 و 50 نقطه بیشتر از روش دیگر است ولی تفاوت بین روش‌ها چندان چشمگیر نیست. حال آن‌که در تراکم 25 نقطه روش فرامکعب لاتین به طوری چشمگیر دارای صحتی بیشتر در مقایسه با روش کاملاً تصادفی و شبکه منظم می‌باشد.

بحث و نتیجه گیری

در این مقاله ضمن مقایسه روش‌ها و تراکم‌های مختلف نمونه‌برداری از یک جامعه قبل از نمونه‌برداری، روش نمونه‌برداری مناسب نیز معرفی شده است. معیار ارزیابی تناسب نمونه‌برداری که در این مقاله معرفی شده است توانایی روش‌ها و تراکم‌های مختلف نمونه‌برداری برای شبیه‌سازی توزیع احتمال متغیر ورودی بوده است با این فرض که در صورت موفقیت نمونه‌برداری در شبیه‌سازی توزیع احتمال متغیر ورودی، توزیع احتمال متغیرهای خروجی (ویژگی‌های خاک) نیز به‌درستی شبیه‌سازی خواهد شد. با توجه به وجود همبستگی بین شوری خاک و شاخص‌های مستخرج از تصاویر ماهواره‌ای که در چندین مقاله گزارش شده است (تقی‌زاده و همکاران، 1392؛ علوی‌پناه و گوسنر، 2011؛ الخیر، 2003؛ دعایی و همکاران، 2006؛ النجار و نولر، 2010؛ نوروزی و همکاران، 2012؛ اسکودپرو و همکاران، 2014) می‌توان باندهای مختلف این تصاویر را به عنوان جامعه ورودی منظور کرد در نتیجه امکان ارزیابی دقت روش‌ها و تراکم‌های مختلف نمونه‌برداری قبل از اقدام به برداشت نمونه میسر خواهد بود. همچنین نتایج نمونه‌برداری شوری خاک نیز نشان داد که با کاهش تراکم نمونه‌برداری از 100 به 25 نقطه ریشه میانگین مربعات خطا از 4/22 به 8/66 دسی زیمنس بر متر افزایش یافت ولی با صحت قابل قبولی توزیع احتمال جامعه اصلی را شبیه‌سازی نمود. نتایج آزمون‌های (Z) و (χ^2) و ریشه میانگین مربعات خطای متغیرهای کمکی (باندهای یک تا هفت و DEM) نشان داد که می‌توان از متغیرهای کمکی شاخص‌های مستخرج از تصاویر ماهواره‌ای برای الگوی نمونه‌برداری خاک در تراکم کم به روش فرامکعب لاتین برای مطالعات شوری خاک در ناحیه مطالعه شده استفاده کرد. این توانایی روش فرامکعب لاتین مربوط به قابلیت آن برای تعیین نقاط نمونه‌برداری از روی توزیع احتمال متغیرهای کمکی ورودی است، به گونه‌ای که توزیع احتمال همه متغیرهای ورودی را به تعداد نقاط نمونه‌برداری به فواصل منظم تقسیم نموده و سپس با ترکیب آن‌ها سعی می‌کند

شکل تابع توزیع احتمال شوری خاک در جامعه اصلی بسیار مشابه توزیع احتمال مدل رقومی ارتفاع است. به این ترتیب مدل رقومی ارتفاع مهمترین پارامتر ورودی تأثیرگذار در مطالعات شوری خاک است. این موضوع با نقشه توزیع شوری خاک نیز مطابقت دارد به این ترتیب که شورترین نقاط در پست‌ترین نقاط ارتفاعی و نزدیک‌ترین نقاط به خروجی حوزه واقع شده بودند.

جدول (1) آزمون مقایسه میانگین‌های حاصل از نمونه‌برداری به روش‌های مختلف و تراکم‌های مختلف را با جامعه اصلی با استفاده از نمرات استاندارد Z نشان می‌دهد. در این جدول، میانگین نمونه‌هایی که در سطح 95 درصد با میانگین جامعه دارای اختلاف معنی‌دار بوده‌اند، با علامت ستاره مشخص شده‌اند. بر این اساس در تراکم نمونه‌برداری 25 نقطه به روش تصادفی برای همه پارامترها به‌جز باندهای 6، اختلافی معنی‌دار بین میانگین نمونه و جامعه وجود دارد. سایر اختلاف‌ها معنی‌دار نیست. همچنین، به روش مشابه برای آزمون انحراف معیار نمونه‌ها از آزمون کای اسکور¹ استفاده شده است. بر این اساس نیز عمده اختلاف انحراف معیار نمونه‌ها با مقادیر انحراف معیار جامعه اصلی مربوط به تراکم نمونه‌برداری 25 نقطه و روش کاملاً تصادفی بوده است. در تراکم 25 نقطه هر چند اختلاف میانگین و انحراف معیار جامعه با مقادیر حاصل از شبکه منظم نیز زیاد است ولی از لحاظ آماری در سطح 95 درصد معنی‌دار نیست. این موضوع می‌تواند مربوط به پراکنش یکنواخت نقاط نمونه‌برداری شده به روش شبکه منظم در کل منطقه مطالعاتی باشد که در تراکم‌های کمتر نیز پوشش کامل‌تری از منطقه را در مقایسه با روش کاملاً تصادفی ارائه می‌دهد ولی باز هم این پوشش به اندازه روش فرامکعب لاتین به توزیع جامعه نزدیک نیست.

آزمون میانگین و انحراف معیار نمونه‌ها در مقایسه با مقادیر جامعه در جدول (1) به صورت خلاصه نشان داده شده است. در مقایسه مقادیر حاصله با جداول استاندارد مشخص می‌شود که اختلاف معنی‌دار میانگین نمونه‌ها با جامعه اصلی تنها مربوط به تراکم 25 نقطه و روش تصادفی است. بر پایه آزمون کای اسکور نیز هر چند در تراکم‌های 100 و 50 نقطه و روش‌های فرامکعب لاتین و شبکه منظم نیز اختلاف‌های معنی‌دار دیده می‌شود اما در 6 مورد از 10 مورد اختلاف معنی‌دار مربوط به تراکم 25 نقطه و روش تصادفی است.

¹ Chi-square

نقاط را به نحوی انتخاب کند که در هر دسته یک نمونه واقع شده باشد.

جدول 1- آزمون میانگین نمونه‌ها با میانگین جامعه (استاندارد Z) و انحراف معیار نمونه‌ها با انحراف معیار جامعه (کای اسکور). اعدادی که دارای بالانویس ستاره (*) هستند به احتمال 95 درصد دارای اختلاف آماری معنی‌دار با جامعه اصلی هستند.

آزمون	پارامتر	تراکم 100		تراکم 50		تراکم 25		فرامکعب لاتین
		منظم	تصادفی	منظم	تصادفی	منظم	تصادفی	
استاندارد Z	باند 1	0/23	-0/74	0/11	0/11	0/13	-0/08	1/46
	باند 2	-0/10	0/71	-0/02	-0/01	0/28	-0/15	1/43
	باند 3	-0/22	-0/57	-0/25	-0/22	0/33	-0/18	1/72
	باند 4	0/45	-0/03	0/45	-0/66	0/17	-0/28	1/13
	باند 5	0/37	-0/66	-0/82	-0/17	0/61	-0/45	1/66
	باند 6	-0/38	-0/73	-1/42	-0/16	1/21	-1/67	0/86
	باند 7	0/14	-0/96	-1/55	0/33	1/03	-0/59	1/75
کای اسکور	DEM	0/49	-0/19	-0/52	-0/15	-0/01	-0/26	-0/28
	باند 1	111/2	118/8	122/1	58/8	55/0	62/0	34/3
	باند 2	109/9	120/8	116/8	59/3	56/9	55/4	31/6
	باند 3	102/9	110/5	124/8	57/7	53/1	51/1	28/0
	باند 4	96/6	132/1*	124/8	46/1	42/8	42/4	22/0
	باند 5	109/6	107/2	201/1*	54/8	49/5	52/5	25/6
	باند 6	97/5	86/9	136/6*	70/0	40/5	72/1*	22/1
DEM	باند 7	113/3	95/1	123/9	63/7	55/5	51/0	29/6
	DEM	103/0	92/2	96/3	35/1	48/4	48/5	20/8

جدول 2- مقایسه ریشه میانگین مربعات خطا (RMSE) در روش‌ها و تراکم‌های مختلف نمونه برداری

تراکم نمونه	روش نمونه برداری	باند 1	باند 2	باند 3	باند 4	باند 5	باند 6	باند 7	DEM	میانگین
100	شبکه منظم	1/68	1/38	1/31	1/33	1/80	0/77	1/89	0/94	1/39
	کاملاً تصادفی	1/02	0/88	1/22	4/78	1/29	0/73	1/34	2/21	1/68
	فرامکعب لاتین	1/46	1/34	1/83	1/59	3/11	1/08	2/93	2/53	1/98
50	شبکه منظم	3/87	4/45	3/39	2/22	3/24	7/01	4/61	1/58	3/80
	کاملاً تصادفی	4/16	4/43	4/13	3/43	3/79	5/83	3/48	3/16	4/05
	فرامکعب لاتین	4/41	3/73	3/07	2/13	3/81	7/43	3/94	2/20	3/84
25	شبکه منظم	9/85	9/49	10/14	6/82	8/59	5/48	9/18	3/09	7/83
	کاملاً تصادفی	19/90	20/90	19/53	18/24	21/36	8/15	21/37	7/75	17/15
	فرامکعب لاتین	5/72	2/58	2/45	5/15	4/77	3/49	4/83	3/04	4

فهرست منابع:

1. تقی‌زاده، ر.، ف. سرمدیان، م. امید، ن. تومانیان، م.ج. روستا و م.ح. رحیمیان. 1392. مقایسه روش‌های شبکه عصبی مصنوعی و درخت تصمیم در تهیه نقشه رقمی خاک در منطقه اردکان. مجله تحقیقات آب و خاک ایران. 173-182: (2) 44.
2. تقی‌زاده، ر.، ف. سرمدیان، م. تازه، م. امید، ن. تومانیان، م.ج. روستا و م.ح. رحیمیان. 1393. مقایسه روش‌های مختلف نمونه‌برداری جهت پهنه‌بندی رقمی خاک در منطقه اردکان. نشریه علمی پژوهشی مهندسی و مدیریت آبخیز. 363-353: 6(4).
3. Alavi Panah S. K. and R. Goossens. 2001. Relationship Between the Landsat TM, MSS Data and Soil Salinity. *J. Agric. Sci. Technol.*, 3: 21-31.
4. Al-Khaier F. 2003. Soil Salinity Detection using Satellite Remote Sensing. Msc. Thesis in Geo information Science and Earth Observation, Watershed Management, Conservation and River Basin Planning. ITC. Netherland.
5. Cline, M.D. 1944. Principles of Soil Sampling. *Soil Sci.* 58: 275- 288.
6. Douaoui, A.K., H. Nicolas and C. Walter. 2006. Detecting salinity hazards within a semiarid context by means of combining soil and remote-sensing data. *Geoderma*, 134 :217–230
7. Elnaggar, A.A. and J.S. Noller. 2010. Application of Remote-sensing Data and Decision-Tree Analysis to Mapping Salt-Affected Soils over Large Areas. *Remote Sens.*, 2:151-165.
8. Hajrasuliha, S., N. Baniabbasi, J. Metthey and D.R. Nielsen. 1980. Spatial variability of soil sampling for salinity studies in southwest Iran. *Irrigation Science*, 1: 197-208.
9. Hanson, B.R. and S.R. Grattan. 1996. Field sampling of soil, water and plants. In: K.K. Tanji (ed.) *Agricultural salinity assessment and management*. ASCE manuals and reports on engineering practice No. 71.
10. Hantschel, T. and A. Kauerauf. 2009. *Fundamentals of Basin and Petroleum Systems Modeling*. Springer. 476 pages.
11. Helton, J.C. and F. J. Davis. 2003. Latin Hypercube Sampling and the Propagation of Uncertainty in Analyses of Complex Systems. *Reliability Eng. & system safety*. 81: 23-69.
12. Hengl, T., D.G. Rossiter and A. Stein. 2003. Soil sampling strategies for spatial prediction by correlation with auxiliary maps. *Australian Journal of Soil Research*, 41:1403–1422.
13. Iman, R.L. and W.J. Conover. 1980. Small sample sensitivity analysis techniques for computer models, with an application to risk assessment. *Communications in Statistics Theory and Methods A9*, 1749–1874.
14. Kempthorne, O. and R.R. Allmaras. 1986, Errors and Variability of Observations. In: A. Klute (ed.) *Methods of soil analysis*. Part 1, Physical and Mineralogical Methods, Second Edition. ASA-SSSA. Madison, WI.
15. McBratney, A.B., M.L. Mendonc-a-Santos and B. Minasny. 2003. On digital soil mapping. *Geoderma*, 117:3–52.
16. McKay, M.D., R.J. Beckman and W.J. Conover. 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21:239–245.
17. Minasny, B. and A.B. McBratney. 2007. Incorporating taxonomic distance into spatial prediction and digital mapping of soil classes. *Geoderma*, 142: 285–293.
18. Minasny B. and A.B. McBratney. 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computers & Geosciences*, 32: 1378–1388.
19. Minasny, B. and A.B. McBratney. 2002. Uncertainty analysis for pedotransfer functions. *European Journal of Soil Science*, 53:417–430.

20. Noroozi A.A., M. Homaei and A. Farshad. 2012. Integrated Application of Remote Sensing and Spatial Statistical Models to the Identification of Soil Salinity: A Case Study from Garmsar Plain, Iran. ENVIRONMENTAL SCIENCES, 9, (1): 59-74
21. Odeh, I.O.A. and A.B. McBratney. 2000. Using AVHRR images for spatial prediction of clay content in the lower Namoi Valley of eastern Australia. Geoderma, 97:237- 254.
22. Pebesma, E.J. and G.B.M. Heuvelink. 1999. Latin hypercube sampling of Gaussian random fields. Technometrics, 41:303–312.
23. Petersen R.G. and L.D. Calvin. 1986. Sampling. In: A. Klute (ed.) Methods of soil analysis. Part 1, Physical and Mineralogical Methods, Second Edition. ASA-SSSA. Madison, WI.
24. Richards, L.A. 1954. Diagnosis and Improvement of Saline Alkali Soils. US Salinity Laboratory Staff. Agriculture Handbook 60.
25. Scudiero, E., T.H. Skaggs and D.L. Corwin. 2014. Regional scale soil salinity evaluation using Landsat 7, western San Joaquin Valley, California, USA. Geoderma Regional 2–3-: 82–90.
26. Wilson P.J. and C.J. Gallant. 2000. Terrain analysis: principles and applications. John Wiley and Sons Ltd: New York.
27. Yates, F. 1981. Sampling methods for censuses and surveys. Charles Griffin and Co. Ltd., 458p.
28. Zhang, Y. and G.F. Pinder. 2004. Latin-hypercube sample-selection strategies for correlated random hydraulic-conductivity fields. Water Resources Research 39 (Art. No. 1226).