

تحلیل حساسیت طرح‌های نمونه‌گیری فضایی به‌منظور پیش‌گویی بهینه

نیر فرزانه خراجو و مجید جعفری خالدی*

دانشگاه تربیت مدرس

چکیده. در تحلیل داده‌های فضایی، معمولاً پیش‌گویی مقدار نامعلوم کمیت مورد مطالعه در موقعیت‌های دلخواه از ناحیه بر اساس مشاهدات بدست آمده در نقاط نمونه‌گیری مورد نظر است. برای این منظور با فرض معلوم بودن اندازه‌ی نمونه، لازم است یک طرح نمونه‌گیری به‌گونه‌ای انتخاب شود که بر اساس مشاهدات حاصل از آن بهترین پیش‌گویی در نقاط مورد نظر بدست آید، که در این صورت آن را، طرح نمونه‌گیری فضایی به‌منظور پیش‌گویی می‌نامند. در این مقاله، ضمن ارائه‌ی شیوه‌ی تعیین طرح بهینه، با استفاده از شبیه‌سازی، نحوه‌ی قرارگیری نقاط در آن‌ها مورد بررسی قرار گرفته و سپس، اثر تغییر پارامترهای مدل بر این‌گونه طرح‌ها ارزیابی شده است. در انتها برای یک مجموعه داده‌ی واقعی مربوط به میزان عیار فلز روی، در ناحیه‌ای واقع در شهرستان خوی، طرح بهینه تعیین و با طرح اولیه مقایسه شده است.

واژگان کلیدی. داده‌های فضایی؛ پیش‌گویی؛ میدان تصادفی گاوسی؛ طرح نمونه‌گیری؛ تحلیل حساسیت؛ شبیه‌سازی؛ داده‌های میزان عیار روی.

۱ مقدمه

در آمار فضایی، تحلیل داده‌هایی مورد نظر می‌باشد که همبسته بوده و این وابستگی ناشی از موقعیت و مکان قرارگیری آن‌ها در یک فضای مورد مطالعه است. چنین وابستگی که با فاصله‌ی بین مشاهدات

*نویسنده‌ی عهده‌دار مکاتبات

مرتبط است، همبستگی فضایی نامیده می‌شود. فرض کنید $Z(\cdot)$ کمیت مورد مطالعه‌ای باشد که روی یک ناحیه‌ی پیوسته‌ی دلخواه $D \subseteq R^d$ ، $d \geq 1$ ، تغییر می‌کند و داده‌های متناظر با این کمیت همبسته فضایی باشند. معمولاً داده‌های فضایی، در n موقعیت $S = \{s_1, s_2, \dots, s_n\} \subset D$ به صورت $z = (z(s_1), z(s_2), \dots, z(s_n))$ مشاهده می‌شوند. برای مدل‌بندی چنین داده‌هایی به منظور تحلیل آن‌ها، معمولاً از یک میدان تصادفی به صورت $Z(\cdot) = \{Z(s), s \in D\}$ استفاده می‌شود. اغلب فرض می‌شود میدان تصادفی $Z(\cdot)$ گاوسی است، یعنی توزیع‌های با ابعاد متناهی آن نرمال هستند. یکی از موضوعات مهم در تحلیل داده‌های فضایی، پیش‌گویی است که در آن با استفاده از مشاهدات بدست آمده در نقاط نمونه‌گیری، مقدار نامعلوم کمیت مورد مطالعه در یک موقعیت دلخواه از ناحیه برآورد می‌شود. به دلیل تأثیر موقعیت‌های نمونه‌ای در عملکرد پیشگو، با فرض آن‌که اندازه‌ی نمونه مشخص باشد، لازم است یک طرح نمونه‌گیری به‌گونه‌ای انتخاب گردد که بر اساس مشاهدات حاصل از آن بهترین پیش‌گویی در موقعیت‌های مورد نظر بدست آید، که در این صورت آن را طرح نمونه‌گیری فضایی بهینه به منظور پیش‌گویی می‌نامند. از این طرح، به عنوان مثال می‌توان برای انتخاب ایستگاه‌های هواشناسی و همچنین تعیین مکان‌هایی مناسب برای حفر چاه در یک ناحیه استفاده کرد.

به طور کلی تعیین طرح بهینه به منظور پیش‌گویی از موضوعات اساسی و پرکاربرد در مبحث آمار فضایی بشمار می‌رود. در این خصوص، بسیاری از افراد مانند بن‌هنی و کامبانیز (۱۹۹۲)، سو و کامبانیز (۱۹۹۳) و ریتر (۱۹۹۶) طرح‌های بهینه به منظور پیش‌گویی را تحت فرض معلوم بودن ساختار همبستگی مورد بررسی قرار دادند. در این حالت، خالدی و ریواز (۱۳۸۴) نیز الگوریتم‌های مختلفی که به منظور جستجوی طرح بهینه به کار می‌روند را مورد مطالعه قرار دادند. همچنین ژئو (۲۰۰۲) و زیمرمن (۲۰۰۶) طرح‌های بهینه را در حالتی بررسی نمودند که ساختار همبستگی نامعلوم بوده و بر اساس یک نمونه‌گیری مقدماتی یا مشاهداتی که از قبل موجود است، برآورد می‌شوند.

تاکنون در مطالعات صورت گرفته، عموماً تأکید بر روش تعیین یک طرح بهینه بوده است. در این مقاله ضمن ارائه‌ی شیوه‌ی بدست آوردن چنین طرحی، با استفاده از شبیه‌سازی، نحوه‌ی قرارگیری نقاط در یک طرح بهینه مورد بررسی قرار گرفته و سپس اثر تغییر پارامترهای مدل بر این‌گونه طرح‌ها ارزیابی شده است. برای این منظور، در بخش ۲، تحت یک مدل گاوسی نوفه‌دار، پیش‌گویی فضایی ارائه شده است. سپس در بخش ۳، طرح بهینه با استفاده از یک الگوریتم بهینه‌سازی با عنوان الگوریتم نوردیدن شبیه‌سازی شده بدست می‌آید. در بخش ۴، با استفاده از شبیه‌سازی، نحوه‌ی قرارگیری نقاط در طرح‌های بهینه و عملکرد آن‌ها نسبت به تغییر پارامترهای مدل مورد مطالعه قرار گرفته است. در بخش ۵ نیز برای یک مجموعه داده‌ی واقعی مربوط به میزان عیار فلز روی، در ناحیه‌ای واقع در شهرستان خوی، طرح بهینه

تعیین و با طرح اولیه مقایسه شده است. در انتها، بحث و نتیجه‌گیری در بخش ۶ ارائه می‌شود.

۲ پیش‌گویی در میدان تصادفی گاوسی

فرض کنید کمیت مورد نظر در نقاط $S = \{s_1, s_2, \dots, s_n\}$ ، $S \subset D$ ، به صورت $z = (z(s_1), z(s_2), \dots, z(s_n))$ مشاهده شده و این مشاهدات تحقق‌ی از یک میدان تصادفی گاوسی $\{Z(s); s \in D\}$ با

$$Z(s) = f'(s)\beta + \sigma\delta(s) + \tau\epsilon(s)$$

باشند، که در آن $f(s) = (f_1(s), \dots, f_p(s))'$ یک بردار $1 \times p$ از مؤلفه‌های غیر تصادفی معلوم و β یک بردار $1 \times p$ از ضرایب رگرسیونی است. همچنین $\sigma^2 = \text{var}(Z(s))$ واریانس و $\delta(\cdot)$ یک میدان تصادفی گاوسی ایستای مرتبه‌ی دوم با میانگین صفر، واریانس ۱ و تابع همبستگی معین مثبت $\rho(\cdot; \xi)$ وابسته به پارامتر نامعلوم ξ در نظر گرفته می‌شود. به علاوه $\epsilon(\cdot)$ یک میدان تصادفی گاوسی ناهمبسته با میانگین صفر و واریانس یک است که برای مدل‌بندی خطاهای اندازه‌گیری یا نوفه به کار می‌رود. پارامتر τ^2 نیز واریانس اثر قطعه‌ای یا نوفه را نمایش می‌دهد. بر این اساس میدان تصادفی $Z(\cdot)$ گاوسی نوفه‌دار نامیده می‌شود. بنا بر این،

$$E(Z(s)) = f'(s)\beta$$

$$\text{cov}(Z(s_i), Z(s_j)) = \sigma^2\rho(s_i - s_j; \xi) + \tau^2U\{s_i = s_j\}$$

که در آن U تابع نشانگر است که اگر $s_i = s_j$ باشد، یک و در غیر این صورت صفر است. با فرض آن‌که $\theta = (\sigma^2, \xi, \tau^2)$ بردار پارامترهای تابع کوواریانس را نمایش دهد داریم

$$Z \sim N(X\beta, \Sigma_\theta)$$

که در آن X یک ماتریس $n \times p$ از مقادیر معلوم به صورت

$$X = \begin{pmatrix} f'(s_1) \\ \vdots \\ f'(s_n) \end{pmatrix}$$

بوده و ماتریس کوواریانس نیز به صورت

$$\Sigma_{\theta} = \text{cov}(Z(s_i), Z(s_j)) = \sigma^2 R_{\xi} + \tau^2 I$$

است، که $R_{\xi} = (\rho(s_i - s_j; \xi))$ ماتریس همبستگی مشاهدات و I ماتریس همانی می‌باشد.

بنا بر این، لگاریتم تابع درست‌نمایی به صورت

$$(1) \quad l(\beta, \theta; z) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln |\Sigma_{\theta}| - \frac{1}{2} (z - X\beta)^T \Sigma_{\theta}^{-1} (z - X\beta)$$

خواهد بود. اکنون پیش‌گویی مقدار نامعلوم میدان تصادفی در نقطه‌ی دلخواه $s \in D$ مورد نظر است. با توجه به گاوسی بودن میدان تصادفی $Z(\cdot)$ ، توزیع توأم $(Z(s), Z)$ نرمال $n + 1$ متغیره با میانگین و ماتریس واریانس کوواریانس

$$\mu^* = \begin{pmatrix} x' \beta \\ X \beta \end{pmatrix}$$

$$\Sigma^* = \begin{pmatrix} \sigma^2 + \tau^2 & k'_{\theta} \\ k_{\theta} & \Sigma_{\theta} \end{pmatrix}$$

است، که در آن $x' = (f_1(s), \dots, f_p(s))$ و k_{θ} یک بردار $n \times 1$ از کوواریانس بین $Z(s)$ و Z می‌باشد. بنا بر این، با فرض معلوم بودن پارامترهای مدل یعنی θ و β پیش‌گویی بهینه به صورت

$$\hat{Z}(s; \beta, \theta) = E(Z(s)|z, \beta, \theta)$$

$$= x' \beta + k'_{\theta} \Sigma_{\theta}^{-1} (z - X \beta)$$

بدست می‌آید. هنگامی که فقط بردار ضرایب رگرسیونی β نامعلوم است، برآورد ماکسیمم درست‌نمایی یا برآورد حد اقل مربعات تعمیم‌یافته‌ی β ، یعنی

$$\hat{\beta}_{\theta} = (X' \Sigma_{\theta}^{-1} X)^{-1} X' \Sigma_{\theta}^{-1} z$$

بدست آمده و در پیش‌گویی بهینه جایگذاری می‌شود. در این حالت با مینیمم کردن واریانس پیش‌گویی

بهترین پیشگویی خطی ناریب^۱ در نقطه‌ی دلخواه $s \in D$ که پیشگوی کریگینگ عمومی (Universal Kriging) نامیده می‌شود، به صورت

$$\begin{aligned} \hat{Z}(s; \theta) &= \lambda^T(s, \theta)z \\ (2) \quad &= (\Sigma_\theta^{-1}k_\theta + \Sigma_\theta^{-1}X(X^T\Sigma_\theta^{-1}X)^{-1}(x - X^T\Sigma_\theta^{-1}k_\theta))^T z \end{aligned}$$

بدست می‌آید (کرسی، ۱۹۹۳) که در آن λ ضریب کریگینگ را نمایش می‌دهد. همچنین واریانس پیشگوی کریگینگ، به‌عنوان معیاری از اندازه‌ی نزدیکی پیشگو به مقدار واقعی آن، عبارت است از

$$\begin{aligned} M(s; S, \theta) &= \sigma^2 + \tau^2 - k_\theta^T \Sigma_\theta^{-1} k_\theta \\ (3) \quad &+ (x - X^T \Sigma_\theta^{-1} k_\theta)^T (X^T \Sigma_\theta^{-1} X)^{-1} (x - X^T \Sigma_\theta^{-1} k_\theta). \end{aligned}$$

هنگامی که پارامترهای تابع کوواریانس نامعلوم هستند، در روش بسامدی^۲ پیشگویی فضایی، با ماکسیم کردن تابع درست‌نمایی نیم‌رخ $l(\hat{\beta}_\theta, \theta; z)$ ، برآورد ماکسیم درست‌نمایی پارامتر نامعلوم θ تعیین می‌شود. سپس با جایگذاری این پارامتر برآورد شده در پیشگوی کریگینگ (۲)، پیشگوی کریگینگ با جایگذاری^۳ به صورت

$$\hat{Z}(s; \hat{\theta}) = \lambda^T(s, \hat{\theta})z$$

بدست می‌آید. از آن‌جا که برآورد ماکسیم درست‌نمایی θ با استفاده از روش‌های عددی تعیین می‌شود (ماریا و مارشال، ۱۹۸۴) محاسبه‌ی تحلیلی واریانس پیشگوی با جایگذاری، یعنی $E(\hat{Z}(s; \hat{\theta}) - Z(s; \theta))^2$ بسیار دشوار است. در این حالت معمولاً از $M(s; S, \hat{\theta})$ به‌عنوان برآوردی از واریانس پیشگویی استفاده می‌شود. اگرچه تاکنون با استفاده از بسط سری تیلور، تقریب‌های دیگری نیز برای واریانس پیشگویی ارائه شده است (هارویله و جسک، ۱۹۹۲ و زیمرمن و کرسی، ۱۹۹۲) اما به‌دلیل مشکلات محاسباتی، کم‌تر مورد استفاده قرار می‌گیرند.

۳ طرح نمونه‌گیری فضایی

حال می‌خواهیم طرح نمونه‌گیری را به‌گونه‌ای تعیین کنیم که پیشگویی روی ناحیه‌ی D بهینه بوده و یا به‌عبارت دیگر، واریانس پیشگوی کریگینگ مینیمم باشد. در این صورت چنین طرحی، طرح نمونه‌گیری

فضایی بهینه با هدف پیش‌گویی نامیده می‌شود. هنگامی که پارامترهای تابع کوواریانس معلوم هستند، یک مزیت مهم استفاده از واریانس پیشگوی کریگینگ (۳) در تعیین طرح نمونه‌گیری، عدم وابستگی آن به بردار مشاهدات است. در واقع فقط به طرح نمونه‌گیری و موقعیت مورد نظر برای پیش‌گویی وابسته است. از آن‌جا که مینیمم کردن واریانس پیش‌گویی روی کل ناحیه‌ی D امکان‌پذیر نیست، از معیار طرح متداول متوسط واریانس کریگینگ، که برابر است با

$$V(S; \theta) = \frac{1}{|D|} \int_D M(s; S, \theta) w(s) ds$$

استفاده می‌شود، که در آن $w(s)$ وزن تعیین‌شده در موقعیت s و $|D|$ مساحت ناحیه‌ی D را نمایش می‌دهد. به‌طور کلی $w(s)$ اهمیت مکان s برای پیش‌گویی را نشان داده و نحوه‌ی انتخاب آن به مسئله‌ی مورد بررسی بستگی دارد. به‌عنوان مثال در مسئله‌ی تعیین موقعیت ایستگاه‌های سنجش آلاینده‌های هوا، پیش‌گویی مکان‌هایی که در آن‌ها تراکم جمعیت بالاتر است، از اهمیت بیشتری برخوردار است. در این مقاله برای سادگی $w(s) \equiv 1$ در نظر گرفته می‌شود. همان‌طور که مشاهده می‌شود معیار طرح $V(S; \theta)$ ، به پارامترهای تابع کوواریانس، یعنی θ وابسته است. لذا به‌منظور تعیین طرح نمونه‌گیری بهینه، هنگامی که θ نامعلوم است، از $M(s; S, \hat{\theta})$ در تعریف معیار متوسط واریانس کریگینگ استفاده می‌شود. به‌عبارت دیگر، برآورد θ به جای مقدار واقعی آن به کار می‌رود. بنا بر این، در ادامه، بحث مبتنی بر معلوم بودن پارامترهای کوواریانس ارائه می‌شود.

طرحی که معیار در نظر گرفته‌شده را مینیمم کند، طرح بهینه به‌منظور پیش‌گویی نامیده می‌شود. در نتیجه طرح بهینه به‌صورت

$$S^* = \arg \min_{S \subset D} V(S; \theta)$$

بدست می‌آید. چون ناحیه‌ی مورد مطالعه‌ی D پیوسته است، تعداد ناشمارا طرح وجود داشته و پیدا کردن طرح نمونه‌گیری بهینه ممکن نیست. بنابراین ناحیه‌ی D را بوسیله‌ی یک شبکه‌ی مناسب با N گره تقریب زده و فرض می‌کنیم طرح S ، زیرمجموعه‌ای n عضوی از D_N باشد. لازم به‌ذکر است شبکه‌ای که تعداد گره‌های آن کم بوده و یا نقاط آن در ناحیه‌ی مورد مطالعه پراکنده نشده باشند، نمی‌تواند تقریب مناسبی از ناحیه ارائه نموده و برای تعیین طرح بهینه مفید باشد. با تبدیل ناحیه‌ی پیوسته‌ی D به ناحیه‌ی گسسته‌ی D_N ، معیار طرح متوسط واریانس کریگینگ به‌صورت

$$V'(S; \theta) = \frac{1}{N} \sum_{s \in D_N} M(s; S, \theta)$$

خواهد بود. بنا بر این، طرح نمونه‌گیری فضایی بهینه به صورت

$$S^* = \arg \min_{S \subset D_N} V'(S; \theta)$$

بدست می‌آید. حال اگر N و n هر دو کوچک باشند تعداد طرح‌های ممکن یعنی $\binom{N}{n}$ نیز نسبتاً کوچک بوده و می‌توان تمامی آن‌ها را بررسی و آن طرحی که معیار طرح را مینیمم می‌کند به عنوان طرح نمونه‌گیری بهینه انتخاب کرد. ولی در حالت کلی برای تقریب مناسب ناحیه‌ی پیوسته، باید تعداد گره‌ها بزرگ اختیار شود، که در این صورت تعداد طرح‌های ممکن نیز زیاد خواهد بود. به عنوان مثال، اگر $N = 100$ و $n = 10$ اختیار شود، $\binom{N}{n} = 17310309456440$ بوده و در این حالت حتی با استفاده از کامپیوترهای بسیار پیشرفته نیز تعیین طرح بهینه بسیار دشوار است. در این حالت برای پیدا کردن طرح نمونه‌گیری بهینه، از یک الگوریتم بهینه‌سازی با عنوان الگوریتم نوردیدن شبیه‌سازی شده^۴ استفاده می‌شود. ایده‌ی این الگوریتم بر اساس علم فیزیک آماری می‌باشد. فرض کنید قرار گرفتن ماده در مکانی با کم‌ترین انرژی مورد نظر باشد به طوری که آن ماده منجمد شود. برای این منظور باید ماده از موقعیت i با انرژی $E(i)$ به موقعیت j با انرژی $E(j)$ در دمای T منتقل شود، به طوری که اگر $E(j) \leq E(i)$ باشد انتقال پذیرفته می‌شود و اگر $E(j) > E(i)$ باشد انتقال با احتمال بولتزمن^۵ که مساوی $\exp\{-\frac{E(i)-E(j)}{T}\}$ است، پذیرفته می‌شود. اکنون اگر به تعداد لازم از انتقال‌ها روی هر دمای داده‌شده انجام شود و دما نیز به تدریج کاهش یابد، در این صورت با احتمال زیاد ماده در مکانی با کم‌ترین انرژی جای گرفته و منجمد می‌شود.

در مسائل ریاضی به منظور تعیین نقاط ماکسیم یا مینیمم یک تابع، در صورتی که استفاده از روش‌های عددی مورد نیاز باشد، یک راه حل، به کارگیری الگوریتم نوردیدن شبیه‌سازی شده است (کیرک پاتریک و دیگران، ۱۹۸۳). برای تعیین طرح نمونه‌گیری فضایی بهینه، ون‌گرونینگ و اشتین (۱۹۹۸) و ژئو و اشتین (۲۰۰۵) استفاده از الگوریتم نوردیدن شبیه‌سازی شده را پیشنهاد کردند با این تشابه که موقعیت ذکر شده، طرح نمونه‌گیری $S(i)$ بوده و انرژی، همان تابع معیار طرح است.

برای ارائه‌ی الگوریتم مورد نظر، ابتدا مفروضات زیر در نظر گرفته می‌شود:

دمای اولیه T ، دمایی که الگوریتم با آن شروع می‌شود و هر عدد مثبتی می‌تواند باشد.

دمای پایانی T_e ، که با رسیدن دما به این مقدار، الگوریتم پایان می‌یابد و باید کوچک و نزدیک به صفر اختیار شود.

شعاع δ ، که با استفاده از آن نقطه‌ی جدید و بالاخره طرح جدید بدست می‌آید و مقدار آن با توجه به ناحیه‌ی مورد مطالعه تعیین می‌شود.

فاکتور کاهش‌ی دما α_1 و فاکتور کاهش‌ی شعاع α_2 ، که با استفاده از آن‌ها دما و شعاع به ترتیب به صورت $T_{k+1} = \alpha_1 T_k$ و $\delta_{k+1} = \alpha_2 \delta_k$ ، $k = 0, 1, \dots$ کاهش می‌یابند و مقادیری که این دو فاکتور می‌گیرند بین صفر و یک است.

تعداد تکرارها در هر سطح دما Ir ، که حلقه‌ی الگوریتم در هر سطح دما به تعداد آن تکرار می‌شود و هر عدد مثبتی می‌تواند باشد.

لازم به ذکر است که انتخاب مقادیر پارامترهای مورد نیاز الگوریتم، به پیچیدگی مسئله وابسته است. مثلاً اگر فاکتور کاهش‌ی دما α_1 نزدیک به یک اختیار شود، در آن صورت تعداد سطح دمای مورد بررسی برای الگوریتم بیشتر می‌شود و یا اگر تعداد تکرارهای Ir بزرگ در نظر گرفته شود، در آن صورت تعداد انتقال‌های بیشتر برای هر سطح دما بررسی می‌شود. در هر دو حالت دقت الگوریتم بیشتر می‌شود ولی زمان اجرای الگوریتم نیز افزایش می‌یابد.

نحوه‌ی انجام این الگوریتم به این صورت است که ابتدا برای طرح اولیه‌ی داده‌شده‌ی S_0 ، معیار طرح یعنی $V(S_0)$ محاسبه می‌شود. سپس در سطح دما و شعاع اولیه‌ی داده‌شده، حلقه‌ی الگوریتم شروع می‌شود، به طوری که ابتدا برای بدست آوردن طرح جدید S_1 ، یک نقطه از طرح اولیه‌ی S_0 ، مانند (x_0, y_0) به‌طور تصادفی انتخاب و حذف می‌شود. سپس با استفاده از زاویه‌ی تصادفی $\theta \in (0, 2\pi)$ شعاع داده‌شده‌ی δ نقطه‌ی جدیدی به‌صورت $(x_1, y_1) = (x_0 + \delta \cdot \cos(\theta), y_0 + \delta \cdot \sin(\theta))$ بدست می‌آید. حال اگر این نقطه‌ی جدید داخل ناحیه‌ی D باشد، الگوریتم ادامه می‌یابد. در غیر این صورت، اگر نقطه خارج از ناحیه باشد، روند انتخاب ادامه می‌یابد تا نقطه‌ی جدید داخل ناحیه قرار گیرد. با فرض قرارگیری نقطه‌ی جدید داخل ناحیه، این نقطه با $(n-1)$ نقطه‌ی قبلی طرح مقایسه می‌شود تا نقطه‌ی جدید بر هیچ یک از نقاط قبلی طرح منطبق نباشد. زیرا در این صورت طرح از n موقعیت به $(n-1)$ مکان کاهش خواهد یافت. در صورتی که نقطه‌ی جدید منطبق بر یکی از $(n-1)$ نقطه‌ی قبلی طرح باشد، این نقطه‌ی جدید مورد قبول نبوده و باید دوباره نقطه‌ی جدید دیگری بدست آید. ولی در صورت تمایز نقطه‌ی جدید با نقاط قبلی طرح، این نقطه‌ی جدید مورد قبول بوده و با جایگزینی آن به‌جای نقطه‌ی حذف‌شده از طرح اولیه، طرح جدید S_1 بدست می‌آید. اکنون معیار طرح برای این طرح جدید به‌صورت

$V(S_1)$ محاسبه شده و با $V(S_0)$ مقایسه می‌شود. اگر $V(S_1) \leq V(S_0)$ باشد، طرح جدید جایگزین طرح اولیه می‌شود. در غیر این صورت، مقدار $p = \exp\left\{\frac{V(S_0) - V(S_1)}{T}\right\}$ محاسبه و با عدد تصادفی a که از فاصله‌ی $(0, 1)$ انتخاب شده است، مقایسه می‌شود. اگر $a \leq p$ باشد مجدداً انتقال پذیرفته شده و طرح S_1 جایگزین طرح S_0 می‌شود. در غیر این صورت طرح S_0 انتخاب می‌شود. این حلقه به اندازه‌ی تعداد تکرارهای Tr که در ابتدای الگوریتم در نظر گرفته شده است، تکرار می‌شود. سپس در انتهای حلقه، با استفاده از دو فاکتور کاهش α_1 و α_2 دما و شعاع کاهش یافته و برای سطح جدید دما و شعاع بدست آمده، حلقه مانند قبل اجرا می‌شود. این روند تا زمانی که الگوریتم به دمای پایانی برسد، ادامه می‌یابد. سرانجام با اتمام الگوریتم، آخرین طرح بدست آمده، طرح نمونه گیری بهینه می‌باشد.

یک مسئله در استفاده از الگوریتم‌های بهینه‌سازی از جمله الگوریتم نوردیدن شبیه‌سازی شده، بعضاً حساس بودن نتایج به نقطه‌ی شروع (طرح اولیه) الگوریتم می‌باشد. بنا بر این، برای بررسی این مسئله، چند طرح مختلف را به‌عنوان طرح اولیه‌ی الگوریتم در نظر گرفته و برای هر کدام از طرح‌ها الگوریتم را اجرا می‌کنیم. اگر طرح‌های حاصل از این الگوریتم‌ها یکسان و یا مشابه هم باشند، نشان‌دهنده‌ی عدم حساسیت به طرح اولیه است. برای مشاهده‌ی برنامه‌های کامپیوتری مربوط به این الگوریتم، به فرزانه (۱۳۸۶) مراجعه شود.

۴ تحلیل حساسیت طرح نمونه‌گیری فضایی بهینه

در این بخش با استفاده از مثال‌های شبیه‌سازی، نحوه‌ی قرارگیری نقاط در طرح‌های بهینه با هدف پیش‌گویی و عملکرد آن‌ها نسبت به تغییر مقادیر پارامترها بررسی می‌شود. برای این منظور، میانگینی میدان تصادفی، ثابت در نظر گرفته شده (بنا بر این $p = 1$) و برای مدل‌بندی ساختار همبستگی از تابع نمایی $\rho(h; \xi) = \exp\left\{-\frac{\|h\|}{\xi}\right\}$ ، $h \in R^2$ ، استفاده می‌کنیم.

ابتدا رابطه‌ی بین ساختار یک طرح و اندازه‌ی معیار طرح مربوط به آن مورد مطالعه قرار می‌گیرد.

مثال ۱. فضای مورد مطالعه‌ی D را مربع $[0, 1] \times [0, 1]$ و شبکه‌ی نمونه‌گیری را به‌صورت $D_N = \{0, 0.5, 1\} \times \{0, 0.5, 1\}$ در نظر می‌گیریم. همچنین اندازه‌ی نمونه را $n = 15$ و مقادیر پارامتری را به‌صورت $\xi = 0.2$ و $\sigma^2 = 1$ فرض می‌کنیم. سه طرح نمونه‌گیری فضایی پراکنده، خوشه‌ای و ترکیبی از این دو حالت را در نظر می‌گیریم. لازم به ذکر است که معمولاً برای انتخاب مکان‌های نمونه‌گیری، این سه طرح، مورد استفاده قرار می‌گیرند (زیمرمن، ۲۰۰۶). در شکل ۱ این سه طرح نمونه‌گیری و مقادیر معیار طرح مربوط به هر یک از آن‌ها نشان داده شده است. با توجه به این شکل، مشاهده می‌کنیم که وقتی طرح شامل یک سری نقاط نزدیک به هم و به‌عبارت دیگر خوشه‌ای است، اندازه‌ی معیار طرح

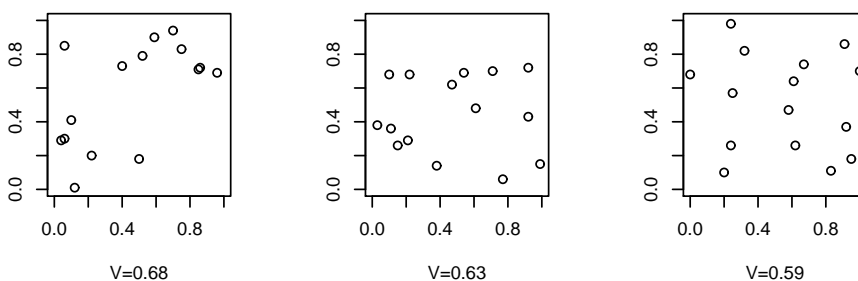
جدول ۱. معیار طرح بهینه‌ی متناظر با مقادیر مختلف پارامترها

ξ	۰/۰۱	۰/۰۵	۰/۱	۰/۱۵	۰/۲	۱	۱۰
$V'(S^*; \theta)$	۰/۳۸۷۳	۰/۳۸۲۶	۰/۳۷۹۴	۰/۳۷۷۴	۰/۳۷۱۵	۰/۳۶۷۲	۰/۳۵۱۱
τ^2	۰/۰۱	۰/۰۵	۰/۱	۰/۱۵	۰/۲	۱	۱۰
$V'(S^*; \theta)$	۰/۲۱۱۵	۰/۲۱۴۶	۰/۲۲۰۶	۰/۲۲۶۹	۰/۲۳۰۱	۰/۲۴۸۸	۰/۲۷۰۳

بزرگ می‌شود ولی هنگامی که نقاط طرح روی کل ناحیه مورد مطالعه‌ی D پراکنده می‌شوند، معیار طرح کاهش می‌یابد. از آن‌جا که طرح نمونه‌گیری بهینه باید دارای کم‌ترین معیار طرح باشد، بنا بر این، طرح نمونه‌گیری بهینه به‌منظور پیش‌گویی، طرحی است که نقاط آن در ناحیه‌ی D پراکنده شده‌اند. این موضوع برای کاهش واریانس پیش‌گویی روی ناحیه مورد مطالعه است.

در مثال بعد، حساسیت طرح نمونه‌گیری بهینه به مقادیر پارامترها بررسی می‌شود. از آن‌جا که پارامترهای همبستگی و اثر قطعه‌ای در تحلیل‌های زمین آمار نقش مهمی دارند، رفتار طرح با تغییر مقادیر آن‌ها مطالعه می‌شود.

مثال ۲. فرض کنید ناحیه‌ی مورد مطالعه‌ی D و شبکه‌ی نمونه‌گیری به‌صورت مثال ۱ باشند. مقادیر $\sigma^2 = ۱$ و اندازه‌ی نمونه را $n = ۱۵$ اختیار می‌کنیم. معیارهای طرح بهینه متناظر با $\tau^2 \in \{۰/۰۱, ۰/۰۵, ۰/۱, ۰/۱۵, ۰/۲, ۱, ۱۰\}$ و $\xi \in \{۰/۰۱, ۰/۰۵, ۰/۱, ۰/۱۵, ۰/۲, ۱, ۱۰\}$ محاسبه و در جدول ۱ نمایش داده شده است. همان‌طور که از این جدول مشاهده می‌شود با افزایش مقادیر ξ و τ^2 ، معیار طرح بهینه تغییرات تدریجی دارد. پس می‌توان نتیجه گرفت شدت همبستگی و اندازه‌ی اثر قطعه‌ای، تأثیر کمی روی طرح‌های نمونه‌گیری به‌منظور پیش‌گویی داشته و طرح‌ها نسبت به تغییر این پارامترها تا حدودی نیرومند هستند. لازم به ذکر است با افزایش شعاع همبستگی ξ ،



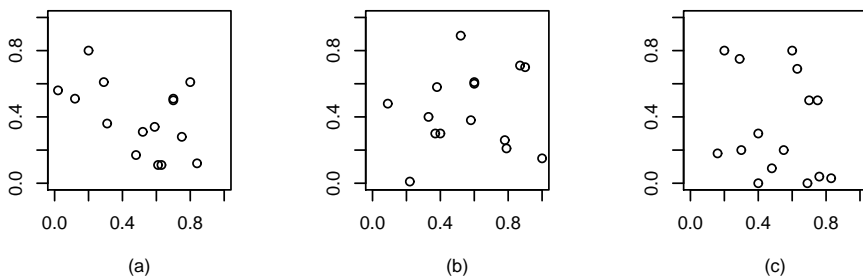
شکل ۱. سه طرح نمونه‌گیری مختلف با معیارهای متناظرشان

چون برای پیش‌گویی یک موقعیت از اطلاعات موقعیت‌های مجاور بیش‌تری استفاده می‌شود، معیار طرح کاهش می‌یابد. در حالی‌که با افزایش τ^2 ، به دلیل افزایش خطای اندازه‌گیری، معیار طرح افزایش یافته است.

در ادامه می‌خواهیم عملکرد طرح‌های نمونه‌گیری بهینه را با تغییر مقادیر پارامترها بررسی کنیم.

مثال ۳. فرض کنید ناحیه‌ی مورد مطالعه‌ی D و شبکه‌ی نمونه‌گیری به صورت مثال ۱ باشند. مقادیر پارامترهای مورد بررسی را به صورت $\tau^2 \in \{0.01, 1, 100\}$ ، $\xi \in \{0.01, 1, 100\}$ و $\sigma^2 = 1$ و اندازه‌ی نمونه را $n = 15$ اختیار می‌کنیم. در شکل ۲ سه طرح نمونه‌گیری بهینه با ξ های متفاوت و مقادیر پارامتری $\tau^2 = 0.01$ و $\sigma^2 = 1$ نشان داده شده است، به طوری که مقادیر ξ برای طرح‌های (a)، (b) و (c) به ترتیب 0.01 ، 1 و 100 می‌باشد. با توجه به این شکل مشاهده می‌کنیم که برای مقادیر کوچک ξ ، ضمن آن‌که طرح نمونه‌گیری بهینه پراکنده است، تا حدودی حالت خوشه‌ای نیز داشته و شامل بعضی نقاط نزدیک به هم می‌باشد. ولی با افزایش شدت همبستگی فضایی، دیگر حالت خوشه‌ای در طرح وجود نداشته و طرح بهینه کاملاً پراکنده است. زیرا وقتی شدت همبستگی فضایی میدان کوچک است، موقعیت‌های نمونه‌ای به هم نزدیک می‌شوند تا از اطلاعات یکدیگر برای پیش‌گویی مناسب‌تر استفاده کنند. در غیر این صورت مشاهدات از یکدیگر مستقل شده و این موضوع به معنای عدم لحاظ همبستگی مکانی در پیش‌گویی است که به دنبال آن نتایج چندان مناسب نخواهد شد. ولی با افزایش مقدار همبستگی، دیگر نیازی نیست تا موقعیت‌های نمونه‌ای متمرکز باشند و در نتیجه طرح کاملاً پراکنده می‌شود.

در شکل ۳، سه طرح نمونه‌گیری بهینه با τ^2 های متفاوت و مقادیر پارامتری $\xi = 0.01$ و $\sigma^2 = 1$ نشان داده شده است، به طوری که مقادیر τ^2 برای طرح‌های (a)، (b) و (c) به ترتیب 0.01 ، 1 و 100 می‌باشد. همان‌طور که مشاهده می‌شود وقتی τ^2 افزایش می‌یابد، ضمن این‌که طرح حالت پراکنده دارد، بعضی نقاط نزدیک به هم نیز در طرح وجود دارد. زیرا وقتی τ^2 ، یعنی خطای اندازه‌گیری کوچک است،

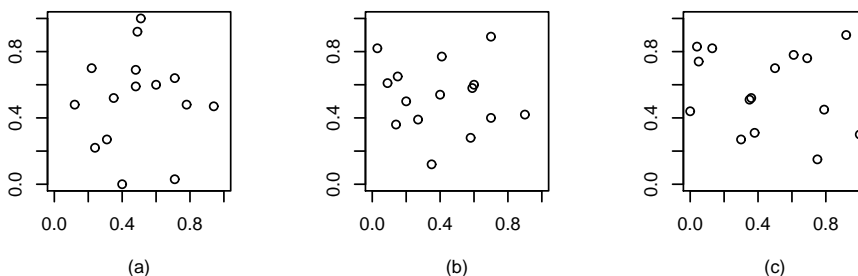


شکل ۲. طرح‌های نمونه‌گیری بهینه با مقادیر مختلف پارامتر ξ

نیازی نیست که از همسایگی خیلی نزدیک به یک موقعیت نمونه‌ای، موقعیت نمونه‌ای دیگری انتخاب شود. ولی با افزایش خطای اندازه‌گیری، به منظور افزایش دقت پیش‌گویی لازم است در مجاورت یک موقعیت نمونه‌ای، موقعیت‌های نمونه‌ای دیگری نیز در نظر گرفته شود تا اطلاعات مناسب‌تری در رابطه با پیش‌گویی بدست آید. .

۵ مثال کاربردی

در سال ۱۳۸۳ سازمان زمین شناسی کشور، در ناحیه‌ای به ابعاد ۹۰ در ۶۰ کیلومتر واقع در شهرستان خوی، میزان عیار فلز روی را به منظور تعیین مکان‌هایی مناسب برای استخراج آن، در ۵۴ موقعیت، اندازه‌گیری کرد. مختصات جغرافیایی موقعیت‌های نمونه‌ای به همراه مقادیر مشاهده‌شده در جدول ۲ ارائه شده و موقعیت مشاهدات در شکل ۴ رسم گردیده است. واحد اندازه‌گیری طول و عرض جغرافیایی، درجه و واحد اندازه‌گیری میزان عیار فلز روی، کیلوگرم بر تن می‌باشد. برای تحلیل فضایی داده‌های میزان عیار فلز روی، ماهیت اولیه‌ی مشاهدات از نظر وجود ایستایی (ایستایی در میانگین و ایستایی تابع کوواریانس) و نرمال بودن آن‌ها توسط خالدی (۱۳۸۴) مورد بررسی قرار گرفته و برقراری این فروض نشان داده شده است. اکنون به تعیین طرح‌های نمونه‌گیری بهینه با هدف پیش‌گویی می‌پردازیم. با استفاده از داده‌های موجود، و بکار بردن تابع $likfit$ از کتابخانه $geoR$ از نرم‌افزار R ، برآورد ماکسیمم درست‌نمایی پارامترها به صورت $\hat{\xi} = 4/65$ و $\hat{\sigma}^2 = 1/02$ و $\hat{\tau}^2 = 0/12$ بدست آمده‌اند. در ادامه با در نظر گرفتن ناحیه‌ی $D = [1, 3/7] \times [2/5, 5]$ ، شبکه‌ی نمونه‌گیری $D_N = \{2/5, 2/51, \dots, 5\} \times \{1, 1/01, \dots, 3/7\}$ و استفاده از الگوریتم نوردیدن شبیه‌سازی شده، طرح نمونه‌گیری بهینه با هدف پیش‌گویی به صورت شکل ۵ با معیار طرح $V'(S^*; \hat{\theta}) = 0/03$ است که نسبت به معیار طرح اولیه که $V'(S; \hat{\theta}) = 0/05$ می‌باشد، کمتر است. بنا بر این، نتیجه‌گیری می‌شود که طرح اولیه برای پیش‌گویی تقریباً به صورت مناسب طراحی



شکل ۳. طرح‌های نمونه‌گیری بهینه با مقادیر مختلف پارامتر τ^2

جدول ۲. داده‌های میزان عیار فلز روی (Z) همراه با مختصات نقاط

x	y	Z(x, y)	x	y	Z(x, y)	x	y	Z(x, y)
۳٫۷	۲٫۵	۱٫۱۵	۳٫۷	۲٫۷	۰٫۸۹	۳٫۶	۲٫۹	۰٫۹۵
۳٫۶	۳٫۲	۰٫۳۹	۲٫۸	۱٫۲	۰٫۷۲	۲٫۹	۱٫۴	۰٫۵۲
۲٫۹	۱٫۶	۰٫۵۷	۲٫۹	۱٫۸	۰٫۷۲	۲٫۸	۲٫۱	۱٫۱۶
۲٫۷	۲٫۳	۱٫۱۲	۲٫۸	۱٫۵	۱٫۰۷	۲٫۷	۲٫۷	۰٫۷
۲٫۹	۲٫۹	۰٫۸۵	۲٫۹	۳٫۱	۱٫۰۲	۳٫۵	۲٫۳	۰٫۹۹
۳٫۵	۲	۰٫۰۳	۳٫۷	۱٫۸	۰٫۶۱	۳٫۷	۱٫۷	۰٫۶۷
۳٫۶	۱٫۴	۰٫۶۲	۳٫۶	۱٫۲	۰	۴٫۳	۳٫۱	۰٫۵۵
۴٫۵	۳	۰٫۴	۴٫۵	۲٫۷	۱٫۸۸۱	۴٫۳	۲٫۵	۰٫۵۳
۴٫۴	۲٫۳	۱٫۱۶	۴٫۴	۲	۰٫۹	۴٫۴	۱٫۸	۱٫۲۵
۴٫۴	۱٫۶	۱٫۲۱	۴٫۳	۱٫۴	۰٫۰۹	۴٫۴	۱٫۲	۰
۲٫۵	۱٫۷	۰٫۳۲	۲٫۵	۱٫۳	۰٫۱۲	۳٫۶	۳٫۶	۱٫۰۱
۳٫۳	۲٫۲	۰٫۶۵	۳٫۱	۲٫۲	۱٫۱۶	۲٫۸	۲٫۲	۰٫۹۹
۲٫۵	۲٫۲	۱٫۴۹	۳٫۵	۲٫۱	۱٫۰۸	۳٫۶	۲	۰٫۶۸
۴٫۴	۲٫۴	۰٫۳۴	۴٫۴	۱٫۹	۱٫۵	۴٫۹	۱٫۵	۰٫۶
۳٫۹	۳٫۱۲	۱٫۰۴	۴٫۹	۲٫۵	۰٫۴۴	۴٫۹	۲٫۸	۰٫۷۱
۴٫۱	۱٫۵	۱	۴٫۱	۱٫۷	۱٫۰۴	۲٫۵	۲٫۶	۰٫۳۸
۴٫۱	۲٫۹	۰٫۵۸	۳٫۳	۱٫۶	۰٫۸۷	۳٫۳	۱٫۹	۱٫۱۶
۳٫۳	۲٫۷	۰٫۷۶	۳٫۳	۳	۰٫۱۲	۲٫۵	۳	۰٫۱۷

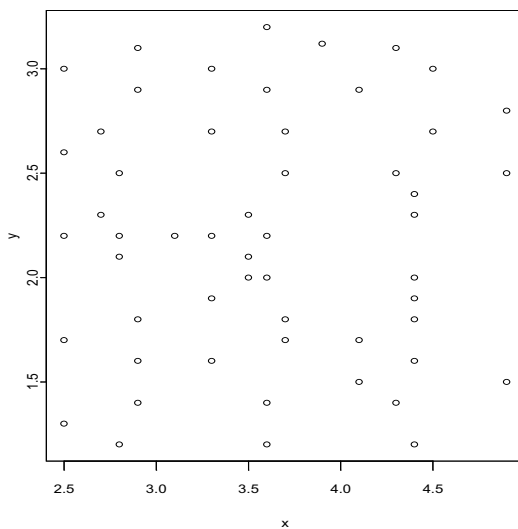
شده است.

۶ بحث و نتیجه‌گیری

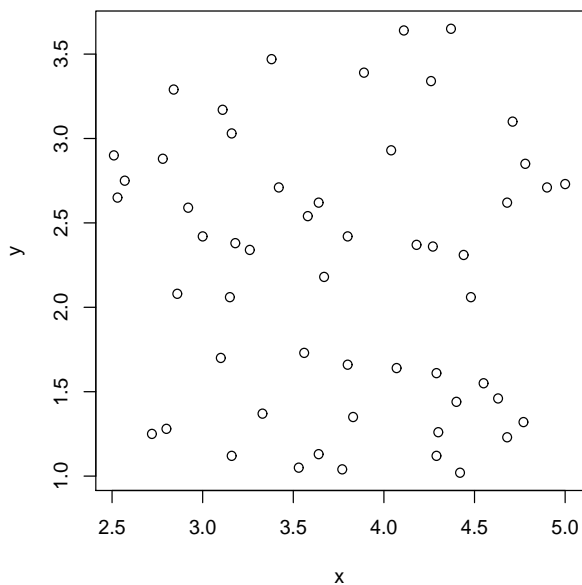
در این مقاله، با هدف پیش‌گویی فضایی، طرح‌های نمونه‌گیری بهینه تعیین گردیدند. برای این منظور، الگوریتم نوردیدن شبیه‌سازی‌شده مورد استفاده قرار گرفته و نتایج زیر بدست آمدند.

۱. طرح‌های نمونه‌گیری بهینه به‌منظور پیش‌گویی، شامل نقاط پراکنده روی کل ناحیه‌ی مورد مطالعه می‌باشند.

۲. در طرح‌های نمونه‌گیری به‌منظور پیش‌گویی، برای مقادیر کوچک شدت همبستگی، ضمن آن‌که طرح نمونه‌گیری بهینه پراکنده است، تا حدودی حالت خوشه‌ای نیز داشته و شامل بعضی نقاط



شکل ۴. محل قرار گرفتن مشاهدات



شکل ۵. طرح نمونه‌گیری بهینه با هدف پیش‌گویی

نزدیک به هم می‌باشد. ولی با افزایش شدت همبستگی فضایی، دیگر حالت خوشه‌ای در طرح وجود نداشته و طرح بهینه کاملاً پراکنده است. همچنین وقتی خطای اندازه‌گیری افزایش می‌یابد، ضمن این که طرح حالت پراکنده دارد، بعضی نقاط نزدیک به هم نیز در طرح وجود دارد.

۳. شدت همبستگی و اندازه‌ی اثر قطعه‌ای، تأثیر کمی روی طرح‌های نمونه‌گیری به‌منظور پیش‌گویی داشته و طرح‌ها نسبت به تغییر این پارامترها نیرومند هستند.

با فرض آن‌که پارامترهای تابع همبستگی معلوم بوده و یا نمونه‌گیری مقدماتی امکان‌پذیر باشد، طرح‌های نمونه‌گیری بهینه مورد بررسی قرار گرفتند. اما هنگامی که این پارامترها نامعلوم بوده و نمونه‌گیری مقدماتی نیز امکان‌پذیر نباشد، از دو روش مینیماکس و بیزی می‌توان برای تعیین طرح بهینه استفاده کرد.

در روش مینیماکس به‌ازای مقادیر مختلف پارامتر، ماکسیمم معیار طرح را تعیین کرده و سپس از بین طرح‌های موجود، طرحی که این مقدار را مینیمم کند، به‌عنوان طرح نمونه‌گیری مینیماکس انتخاب می‌شود. در روش بیزی میانگین موزون معیار طرح با وزن‌های توزیع پیشینی بدست آمده و سپس از بین طرح‌های موجود، طرحی که این مقدار را مینیمم کند به‌عنوان طرح نمونه‌گیری فضایی بیزی اختیار می‌شود. بررسی و مقایسه‌ی این دو طرح به ویژه تحلیل حساسیت آن‌ها به تغییر مقادیر پارامترها توسط نویسندگان این مقاله تحت بررسی قرار دارد.

توضیحات

۱. Best Linear Unibased Predictor

۲. Frequentist Method

۳. Plug-in Predictor

۴. Simulated Annealing

۵. Boltzmann Probability

۶. Robust

مرجع‌ها

- جعفری خالدی، م. (۱۳۸۴). تجزیه و تحلیل برای پیش‌گویی فضایی. رساله دکتری، دانشگاه تربیت مدرس.
- جعفری خالدی، م. و ریواز، ف. (۱۳۸۴). تعیین طرح نمونه‌گیری بهینه برای تحلیل داده‌های فضایی، مجله پژوهش‌های آماری ایران، دوره ۲، شماره ۱. صص ۵۳-۶۰.
- فرزانه، ن. (۱۳۸۶). طرح نمونه‌گیری بهینه برای برآورد ساختار همبستگی و پیش‌گویی فضایی، پایان‌نامه کارشناسی ارشد، دانشگاه تربیت مدرس.
- Benhenni, K. and Cambanis, S. (1992). Sampling design for estimating integrals of stochastic processes. *Ann. Statist.* **20**, 161-194.
- Harville, D.A. and Jeske, D.R. (1992). Mean squared error of estimation or prediction under a general linear model. *J. Amer. Statist. Assoc.* **87**, 724-731.
- Jackson, R.B. and Caldwell, M.M. (1993). Geostatistical patterns of soil heterogeneity around individual perennial plants. *J. Ecol* **81**, 683-692.
- Kirkpatrick, S., Gelatt, C.D. and Vecchi, M. P. (1983). Optimization by simulated annealing. *Science* **220**, 671-68
- Mardia, K.V. and Marshal, R.J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71**, 135-146.
- Ritter, K. (1996). Asymptotic optimality of regular sequence designs. *Ann. Statist.* **24**, 2081-2096.
- Su, Y. and Cambanis, S. (1993). Sampling designs for estimation of a random process. *Stochastic process. Appl.* **46**, 47-89.
- Van Groenigen, J.W. and Stein, A. (1998). Constrained optimization of spatial sampling using continuous Simulated Annealing. *J. Environ. Quality.* **43**, 684-691.
- Zhu, Z. (2002). *Optimal Sampling Design and Parameter Estimation of Gaussian Random Fields*. PH.D. Thesis, Department of Statistics, University of Chicago.
- Zhu, Z. and Stein, M. (2005). Spatial sampling design for parameter estimation of the covariance function. *J. Statist. Plann. Inference* **134**, 583-603.
- Zimmerman, D.L. and Cressie, N. (1992). Mean squared prediction error in the spatial linear model with estimated covariance parameters. *Ann. Inst. Statist. Math.* **44**, 27-43.
- Zimmerman, D.L. (2006). Optimal network design for spatial prediction, covariance parameter estimation and empirical prediction. *Environmetrics* **17**, 635-652.

مجیدي جعفری خالدي
گروه آمار، دانشكده‌ی علوم ریاضی،
دانشگاه تربیت مدرس،
تهران، ایران.
رایانشانی: jafari-m@modares.ac.ir

نیر فرزانه خراجو
گروه آمار، دانشكده‌ی علوم ریاضی،
دانشگاه تربیت مدرس،
تهران، ایران.
رایانشانی: