https://doi.org/10.22034/jr_iss.2024.728280
اندیشه آماری، سال ۱۴۰۳، دوره ۲۹، شماره ۱، ص ۱۰۶-۹۷

# ارزیابی بیزی مدل رگرسیون فضایی چوله با میدان تصادفی چوله گاوسی منعطف با استفاده از الگوریتم مونت‌کارلو همیلتونی

**فاطمه حسینی[۱]* و امید کریمی[۲]**

[۱،۲] گروه آمار، دانشکده ریاضی، آمار و علوم کامپیوتر، دانشگاه سمنان

تاریخ دریافت: ۱۴۰۴/۰۴/۰۷
تاریخ پذیرش: ۱۴۰۴/۰۵/۱۹

**چکیده:**

در عمل توزیع داده‌های فضایی دارای چولگی هستند که به‌دلیل پیچیدگی‌های ذاتی آن‌ها، مدل‌سازی آماری را با چالش‌های جدی مواجه می‌سازد. مدل‌های میدان تصادفی چوله گاوسی چارچوبی منعطف برای تحلیل این نوع داده‌ها فراهم می‌کنند، اما بسیاری از آن‌ها با مشکلاتی مانند پیچیدگی محاسباتی و عدم شناسایی‌پذیری پارامترها روبه‌رو هستند که دقت تحلیل را کاهش می‌دهد. در این مقاله، یک مدل رگرسیون فضایی بر پایه‌ی توزیع چوله نرمال بسته‌ی منعطف توسعه داده شده است که دارای مزایایی چون شناسایی‌پذیری کامل، بسته بودن تحت حاشیه‌سازی و شرطی‌سازی، و انعطاف‌پذیری بالا در مدل‌سازی ساختارهای پیچیده‌ی فضایی است. تحلیل بیزی مدل با بهره‌گیری از الگوریتم مونت‌کارلو همیلتونی انجام شده است؛ روشی پیشرفته در چارچوب الگوریتم‌های زنجیر مارکوفی که با استفاده از مشتقات توزیع هدف، نرخ پذیرش و سرعت همگرایی را افزایش می‌دهد. برای ارزیابی عملکرد مدل پیشنهادی، یک مطالعه شبیه‌سازی انجام شده و نتایج حاصل از روش مونت‌کارلو همیلتونی با روش‌های کلاسیک زنجیر مارکوفی مونت‌کارلو مقایسه شده است. نتایج بیانگر بهبود در دقت و کارایی الگوریتم پیشنهادی هستند. همچنین، مدل پیشنهادی توانایی بالایی در تحلیل داده‌های فضایی با ابعاد بالا دارد.

**واژه‌های کلیدی:** توزیع چوله نرمال بسته، شناسایی‌پذیری، تغییرنگار تجربی فضایی، روش مونت‌کارلو همیلتونی.

## ۱ مقدمه

در بسیاری از مسائل عملی، داده‌ها دارای وابستگی‌های مکانی هستند؛ به‌عنوان مثال در تحلیل داده‌های آلودگی هوا این باور وجود دارد که میزان آلودگی به موقعیت جغرافیایی ایستگاه‌های اندازه‌گیری بستگی دارد. معمولاً برای تحلیل چنین داده‌هایی از میدان‌های تصادفی گاوسی استفاده می‌شود. با این حال، گاهی داده‌ها دارای چولگی هستند و استفاده از میدان‌های تصادفی گاوسی برای تحلیل داده‌های فضایی چوله می‌تواند منجر به کاهش دقت پیش‌گویی‌ها و برآورد پارامترها شود. در سال‌های اخیر، برای تحلیل داده‌های فضایی چوله، میدان‌های تصادفی چوله گاوسی به‌عنوان تعمیمی از میدان‌های تصادفی گاوسی معرفی شده‌اند، اما اغلب این میدان‌ها خوش‌تعریف نیستند و مشکلاتی در کاربرد آن‌ها وجود دارد. به‌عنوان مثال با استفاده از میدان تصادفی [۶]، [۱۱] برای اولین بار میدان تصادفی چوله گاوسی تقریباً مانا را تعریف کردند. پس از آن، با اصلاح این مدل، نسخه‌ای مانا از میدان چوله گاوسی پیشنهاد گردید، [۵]. در ادامه، این ایده‌ها برای مدل‌سازی و برآورد مدل‌های آمیخته خطی تعمیم‌یافته فضایی در چارچوب بیزی تقریبی به‌کار گرفته شد، [۲]. همچنین، یک زیررده انعطاف‌پذیر از توزیع چوله نرمال بسته [۱] (CSN) معرفی شد و بر پایه‌ی آن، میدان تصادفی فضایی‌ای تعریف گردید که پارامترهای آن قابل شناسایی و تفسیر هستند [۱۰]. توزیع *CSN* منعطف[۲] (FCSN) ساختاری مشابه توزیع‌های چوله [۸] دارد، هرچند تفاوت‌هایی در ساختار همبستگی فضایی و پارامترها مشاهده می‌شود.

در ادامه، [۱] ویژگی‌های میدان تصادفی فضایی چوله گاوسی منعطف را بر اساس توزیع *FCSN* ارائه و مدل فضایی مناسبی برای داده‌های چوله معرفی کردند که پارامترهای آن با استفاده از روش

*نویسنده مسئول: fatemeh.hoseini@semnan.ac.ir

[1] Closed Skew Normal
[2] Flexible Closed Skew Normal

درست‌نمایی، تخمین زده شد. همچنین [۳] با بهره‌گیری از یک زیرده منعطف از توزیع چوله نرمال بسته، به تحلیل بیزی سلسله‌مراتبی مدل‌های رگرسیون فضایی پرداختند و برای کاهش زمان محاسباتی، از روش بیز مقداری برای تقریب توزیع پسین استفاده کردند.

با توجه به پیچیدگی مدل‌های فضایی چوله به‌دست آوردن برآوردهای ماکسیمم درست‌نمایی به‌راحتی امکان‌پذیر نیست؛ بنابراین، استفاده از رهیافت بیزی توصیه می‌شود. اما تحلیل بیزی این مدل‌ها نیز به دلیل پیچیدگی توزیع پسین، نیازمند استفاده از الگوریتم‌هایی مانند متروپلیس–هستینگس و گیبز است که ممکن است زمان‌بر بوده و همگرایی آن‌ها دشوار باشد.

در این راستا، [۴] پیشنهاد کردند که در تحلیل مدل‌های پیچیده و داده‌های با بعد بالا، به جای الگوریتم‌های مذکور، از الگوریتم مونت‌کارلو همیلتونی[3] ($HMC$) به‌عنوان یکی از روش‌های پیشرفته در چارچوب روش زنجیر مارکوف مونت‌کارلویی[4] ($MCMC$) استفاده می‌شود. بر اساس میدان تصادفی چوله گاوسی معرفی شده در [۱۱]، متغیرهای پنهان همبستگی فضایی در مدل‌های آمیخته خطی تعمیم‌یافته فضایی مدل‌سازی شده و یک الگوریتم جدید ترکیبی از الگوریتم بیشینه‌سازی امیدریاضی و $HMC$ برای برآورد ماکسیمم درست‌نمایی پارامترها ارائه گردید. [۹] نیز از این الگوریتم در مطالعه داده‌های زلزله‌ای که دارای چولگی هستند بهره برد. در این مقاله، بر اساس میدان تصادفی فضایی چوله گاوسی منعطف معرفی‌شده در [۱]، داده‌های فضایی چوله مدل‌سازی شده و به‌منظور کاهش زمان محاسبات و افزایش سرعت همگرایی، یک تحلیل بیزی با بهره‌گیری از الگوریتم مونت‌کارلو همیلتونی پیشنهاد گردیده است. این رویکرد از نظر ساختار مدل و روش برآورد با مقالات [۵]، [۴] و [۳] تفاوت اساسی دارد. ساختار مقاله بدین ترتیب است: در بخش ۲، میدان تصادفی چوله گاوسی منعطف به‌کاررفته در این پژوهش معرفی می‌شود؛ در بخش ۳، مدل رگرسیون فضایی چوله، رهیافت بیزی و الگوریتم پیشنهادی برای برآورد پارامترهای مدل ارائه می‌گردد و در نهایت، در بخش‌های ۴ و ۵، به‌ترتیب یک مطالعه شبیه‌سازی و یک مطالعه واقعی برای ارزیابی مدل و رهیافت پیشنهادی مورد بررسی قرار می‌گیرد.

## ۲ میدان تصادفی چوله گاوسی منعطف

فرض کنید بردار $U \in R^{p+q}$ دارای توزیع نرمال چندمتغیره به‌صورت

$$U = \begin{pmatrix} U_1 \\ U_2 \end{pmatrix} \sim N_{p+q}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right), \qquad (۱)$$

باشد. آن‌گاه $X = [U_1 | U_2 \le \mathbf{0}]$ دارای توزیع CSN با تابع چگالی

$$\begin{aligned} f_X(x) &= k\phi_p(x;\mu_1,\Sigma_{11}) \\ &\quad \times \Phi_q(\mathbf{0};\mu_2+\Sigma_{21}\Sigma_{11}^{-1}(x-\mu_1),\Sigma_{2|1}), \end{aligned} \qquad (۲)$$

است، که در آن $k^{-1} = \Phi_q(\mathbf{0};\mu_2,\Sigma_{22})$ و $\Sigma_{2|1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$ و $\phi$ و $\Phi$ به‌ترتیب تابع چگالی و تابع توزیع تجمعی توزیع نرمال چند متغیره با پارامترهای مورد نظر هستند. معمولا توزیع $CSN$ را به شکل

$$\mathrm{CSN_{p,q}}(\mu_1,\Sigma_{11},\Sigma_{21}\Sigma_{11}^{-1},\mu_2,\Sigma_{2|1})$$

نمایش می‌دهند. یک زیرده منعطف از توزیع $CSN$ توسط [۱۰] معرفی شده که مدل‌های فضایی مانا و شناسایی‌پذیر تولید می‌کند و آن را با نماد $FCSN$ نمایش می‌دهند.

**تعریف ۱.۲.** فرض کنید $\lambda \in \mathbb{R}$، $\sigma > 0$، $\mu \in \mathbb{R}$ و $\Sigma_n$ یک ماتریس معین مثبت $n \times n$ باشد. با تعریف $b = (2/\pi)^{\frac{1}{2}}$، $\delta = \lambda(1+\lambda^2)^{-\frac{1}{2}}$ و $\tau = (1-b^2\delta^2)^{-\frac{1}{2}}$، اگر بردار $n$–بُعدی $Z_n$ یک توزیع $CSN$ به‌صورت

$$CSN_{n,n}(\mu - b\delta\sigma\tau\Sigma_n^{\frac{1}{2}}\mathbf{1_n}, \sigma^2\tau^2\Sigma_n, \frac{\lambda}{\sigma\tau}\Sigma_n^{-\frac{1}{2}}, \mathbf{0}, I_n).$$

باشد، $\mathbf{1_n}$ بردار یکه $n$–بُعدی و $I_n$ ماتریس همانی $n \times n$، آن‌گاه این توزیع یک توزیع $FCSN$ به شکل $Z_n \sim FSCSN_n(\mu,\Sigma_n,\sigma,\lambda)$ است.

امید ریاضی و واریانس بردار تصادفی $Z_n$ به‌صورت

$$E(Z_n) = \mu, \quad \mathrm{Var}(Z_n) = \sigma^2\Sigma_n,$$

به‌دست می‌آید، که ملاحظه می‌شود میانگین و واریانس به پارامتر چولگی وابسته نیست. این مزیتی است که باعث می‌شود ساختار همبستگی فضایی داده‌ها را بتوان براساس تغییرنگار تجربی بدون تاثیرگذاری میزان چولگی داده‌ها تعیین کرد. [۱] نشان دادند که در این حالت میزان چولگی داده‌ها تاثیری بر تعیین مدل تغییرنگار ندارد. تابع چگالی توزیع $FCSN$ مطابق رابطه (۲) به‌صورت

$$f_{Z_n}(z) = 2^n\phi_n(z;\mu - b\delta\sigma\tau\Sigma_n^{\frac{1}{2}}\mathbf{1_n}, \sigma^2\tau^2\Sigma_n)$$

[3] Hamiltonian Monte Carlo
[4] Markov chain Monte Carlo

$$\times\Phi_n(\frac{\lambda}{\sigma\tau}\Sigma_n^{-\frac{1}{2}}(z-\mu)+b\delta\lambda\mathbf{1_n};\mathbf{0},I_n),$$

خلاصه می‌شود، که در آن پارامتر $\lambda$ چولگی توزیع را کنترل می‌کند. نکته حائز اهمیت این است که توزیع $FCSN$ نسبت به حاشیه‌سازی بسته است و این شرط خاصیت سازگاری حاشیه‌ای در تعریف میدان تصادفی فضایی چوله را تضمین می‌کند.

**تعریف ۲.۲.** فرض کنید مجموعه $\{Z(s);s\in D\subseteq\mathbb{R}^d,d\in\mathbb{N}\}$ یک میدان تصادفی فضایی است و برای هر تعداد متناهی از موقعیت‌های فضایی $s_1,\ldots,s_n\in D$ بردار $Z_n=(Z(s_1),\ldots,Z(s_n))'$ را در نظر بگیرید. به‌علاوه، فرض کنید $X$ یک ماتریس $n\times p$ از متغیرهای کمکی در موقعیت‌های فضایی، $\beta$ یک بردار $p$ بُعدی از ضرائب رگرسیونی، $\lambda$ پارامتر چولگی و $\Sigma_n=\sigma^2C_n$ ماتریس کوواریانس است، که $C_n$ ماتریس همبستگی فضایی بین مقادیر $Z_n$ می‌باشد. ماتریس $C_n$ با استفاده از یک تابع همبستگی فضایی معتبر $\rho(h;\varphi)$ تعریف می‌شود، به‌گونه‌ای که:

$$[C_n]_{ij}=\rho(h_{ij};\varphi),\quad h_{ij}=\|s_i-s_j\|_2,$$

که در آن $\varphi>0$ پارامتر دامنه فضایی تابع همبستگی و $h_{ij}$ فاصله اقلیدسی بین مکان‌های $s_i$ و $s_j$ است. اگر توزیع توام $Z_n$ یک FCSN به‌صورت $FSCSN_n(X\beta,\Sigma_n,\sigma,\lambda)$ باشد، آن‌گاه $\{Z(s);s\in D\}$ یک میدان تصادفی فضایی $FCSN$ است. ([۱])

بنابراین پارامترهای میدان تصادفی فضایی $FCSN$ به‌صورت $\eta=(\beta,\lambda,\sigma,\varphi)$ می‌باشد. تابع همبستگی در این میدان به‌آسانی برآورد می‌شود چون در میدان تصادفی $FSN$ پارامتر چولگی روی میانگین و واریانس میدان تصادفی تاثیری ندارد، در حالی که میدان تصادفی فضایی چوله مانا [۲] دارای این ویژگی نیست و برآورد تابع همبستگی براساس تغییرنگار تجربی به پارامتر چولگی وابسته است.

# ۳ مدل رگرسیون فضایی چوله

مدل رگرسیون فضایی چوله برای بردار پاسخ $\mathbf{Y_n}$ در $n$ موقعیت فضایی $(s_1,\ldots,s_n)$ را به‌صورت

$$Y_n=X\beta+Z_n,\qquad(۳)$$

تعریف می شود، که در آن $X$ یک ماتریس $n\times p$ از متغیرهای توضیحی، $\beta$ یک بردار $p$ بعدی از ضرائب رگرسیونی و $Z_n$ یک تحقق از میدان تصادفی $FCSN$ با میانگین صفر به صورت $FCSN_n(0,C_n,\sigma,\lambda)$ است.

بنابراین بردار پاسخ‌های فضایی $\mathbf{Y_n}$ با توجه به خاصیت تبدیلات خطی توزیع $CSN$ به‌صورت

$$Y_n\sim CSN_{n,n}(\mu_y,\Sigma_y,\Gamma_y,\mathbf{0},\mathbf{I_n})$$

به‌دست می‌آید، که در آن $\mu_y$، $\Sigma_y$ و $\Gamma_y$ به‌صورت زیر

$$\mu_y=X\beta-b\delta\sigma\tau C_n^{\frac{1}{2}}\mathbf{1_n},\quad \Sigma_y=\sigma^2\tau^2C_n,\ \Gamma_y=\frac{\lambda}{\sigma\tau}C_n^{-\frac{1}{2}},$$

می‌باشند، که

$$b=\sqrt{\frac{2}{\pi}},\quad \delta=\frac{\lambda}{\sqrt{1+\lambda^2}},\quad \tau=\frac{1}{\sqrt{1-b^2\delta^2}}.$$

بنابراین تابع چگالی توزیع $Y_n$ به‌صورت

$$f_\eta(y)=2^n\,\phi_n(y;\mu_Y,\Sigma_Y)\,\Phi_n\left(\Gamma_Y(y-\mu_Y)\right).\qquad(۴)$$

خلاصه می‌شود. در این حالت، بردار پارامترهای مدل به‌صورت $\eta=(\beta,\sigma,\varphi,\lambda)'$ تعریف می‌شود، که در آن $\beta$ ضرایب رگرسیونی، $\sigma$ پارامتر مقیاس، $\varphi$ پارامتر دامنه تابع همبستگی فضایی، $\lambda$ پارامتر چولگی مدل است. در این مقاله هدف بهره‌گیری از رهیافت بیزی برای برآورد پارامترهای مدل می‌باشد، که با تعیین توزیع‌های پیشین مناسب برای پارامترها، الگوریتم‌های $MCMC$ و رهیافت همیلتونی برای نمونه‌برداری از توزیع پسین به کار گرفته می‌شوند. در بخش‌های بعدی به جزئیات این روش‌ها و نحوه پیاده‌سازی آن‌ها پرداخته خواهد شد.

## ۱.۳ برآورد بیزی مدل

در چارچوب تحلیل بیزی مدل رگرسیون فضایی چوله، هدف اصلی برآورد توزیع پسین پارامترها با استفاده از اطلاعات مشاهدات و توزیع‌های پیشین مناسب است. برای این منظور، مجموعه پارامترهای مدل شامل ضرایب رگرسیونی $\beta$، پارامتر مقیاس $\sigma$، پارامتر دامنه همبستگی فضایی $\varphi$ و پارامتر چولگی $\lambda$ در نظر گرفته و بردار کلی پارامترها به‌صورت $\eta=(\beta,\sigma,\varphi,\lambda)$ تعریف می‌شود. برای مدل‌سازی بیزی، توزیع‌های پیشین مستقل زیر برای پارامترها به‌صورت

$$\beta\sim\mathcal{N}_p(\mu_\beta,\sigma_\beta^2I_p),\quad \sigma\sim\mathsf{IG}(a,b),$$
$$\varphi\sim\mathsf{G}(\alpha,\gamma),\quad \lambda\sim\mathcal{N}(\mu_\lambda,\sigma_\lambda^2).$$

فرض شد، که در این توزیع‌ها برای $\beta$ یک توزیع نرمال چندمتغیره با میانگین $\mu_\beta$ و کوواریانس $\sigma_\beta^2I_p$ به‌عنوان پیشین ناآگاهی‌بخش؛ توزیع معکوس گاما $\mathsf{IG}(a,b)$ برای $\sigma$ به‌دلیل تضمین مثبت بودن و انعطاف‌پذیری در کنترل پراکندگی؛ توزیع گاما برای $\varphi$ به‌دلیل مثبت بودن و شکل تحلیلی مناسب و توزیع نرمال برای $\lambda$ به‌منظور پوشش

چولگی‌های مثبت و منفی در نظر گرفته شده است. در نتیجه، توزیع پسین توأم پارامترها به‌صورت

$$\begin{aligned}\pi(\eta|y) &\propto f_\eta(y)\,\pi(\beta)\,\pi(\sigma)\,\pi(\varphi)\,\pi(\lambda)\\ &\propto \phi_n(y;\mu_y,\Sigma_y)\,\Phi_n(\Gamma_y(y-\mu_y))\\ &\times \phi_p(\beta;\mu_\beta,\sigma_\beta^2 I_p)\,\mathsf{IG}(\sigma;a,b)\,\mathsf{G}(\varphi;\alpha,\gamma)\,\mathcal{N}(\lambda;\mu_\lambda,\sigma_\lambda^2),\end{aligned}$$

حاصل می‌شود. با توجه به پیچیدگی توزیع پسین و حضور تابع توزیع تجمعی چندمتغیره نرمال در آن، استفاده از روش‌های تحلیلی مستقیم ممکن نیست. بنابراین، نمونه‌گیری عددی از توزیع پسین با استفاده از الگوریتم‌های زنجیر مارکوف مونت‌کارلویی (*MCMC*) ضروری است. با این حال، روش‌های کلاسیک *MCMC* مانند متروپلیس–هستینگس یا گیبز، در مدل‌هایی با ابعاد بالا یا ساختارهای همبستگی پیچیده مانند مدل حاضر، دچار همگرایی کند، نرخ پذیرش پایین و خودهمبستگی شدید نمونه‌ها می‌شوند. در این شرایط، الگوریتم مونت‌کارلو همیلتونی که یک الگوریتم مشتق‌مبنا در چارچوب *MCMC* است، به‌عنوان گزینه‌ای کارآمد استفاده می‌شود.

الگوریتم *HMC* با افزودن یک بردار کمکی $r$ به فضای پارامترها، سیستم دینامیکی همیلتونی به‌صورت

$$\begin{aligned}r &\sim \mathcal{N}(\mathbf{0},M),\\ p(r,\eta) &= \mathcal{N}(r;\mathbf{0},M)\pi(\eta|y),\end{aligned}$$

تعریف می‌کند، که در آن $M$ ماتریس کوواریانس تقریب‌زده‌شده از نمونه‌ها در مرحله داغیدن[5] است و معمولاً $M = I_{p+3}$ انتخاب می‌شود.

## ۴ مطالعه شبیه‌سازی

در این بخش، مطالعه شبیه‌سازی از میدان تصادفی فضایی چوله منعطف در یک مشبکه منظم $20\times 20$ با تابع همبستگی فضایی نمایی همسانگرد بر روی مدل پیاده سازی می‌شود. لازم به ذکر است که طرح شبیه‌سازی به‌صورت ۱۰۰ بار تکرار مستقل اجرا شده است. برای مقایسه عملکرد روش پیشنهادی *HMC* از الگوریتم‌های متروپلیس–هستینگس و گیبز به‌عنوان روش‌های سنتی *MCMC* استفاده شده است. این انتخاب به‌دلیل سادگی پیاده‌سازی و رایج بودن آن در مدل‌های فضایی صورت گرفته است. و رهیافت پیشنهادی مورد ارزیابی قرار می‌گیرد. فرض کنید بردار $Z_n$ یک تحقق به حجم $n$ از میدان تصادفی $FSCSN_n(X\beta,\Sigma_n,\sigma,\lambda)$ با پارامترهای معلوم

$$\beta_0 = 1, \beta_1 = 2, \sigma = 1, \varphi = 6,\ \lambda = 2.5$$

باشد، که در آن دامنه فضایی یک مشبکه منظم $20\times 20$ به‌صورت

$$D = \{s_i = (x_i,y_i); (x_i,y_i) \in \{1,\dots,20\}\times\{1,\dots,20\}\}$$

$i = 1,\dots,400$، در نظر گرفته شد. مولفه $i$ام میانگین میدان تصادفی به‌صورت $\beta_0 + \beta_1 x_i$ و تابع همبستگی نمایی همسانگرد $\rho(h) = e^{-\frac{h}{\varphi}}$ لحاظ شده است. این تابع برای ساخت ماتریس وابستگی فضایی $C_n$ استفاده می‌شود که در آن $h$ فاصله اقلیدسی بین دو موقعیت در دامنه فضایی است.

شکل ۱ یک تحقق $Z_n$ از میدان تصادفی چوله منعطف را نمایش می‌دهد. در شکل ۱ (ب) موقعیت داده‌های تولید شده با تفکیک رنگ برای مقدار

توجه داشته باشید که بعد بردار کمکی $r$ برابر با بعد فضای پارامترهای مدل، یعنی $r \in \mathbb{R}^{p+3}$ است. انرژی کل سیستم تعریف‌شده به‌صورت تابع همیلتونی

$$\begin{aligned}H(r,\eta) &= -\log p(r,\eta)\\ &= \underbrace{\tfrac{1}{2} r^\top M^{-1} r}_{\text{انرژی جنبشی}} + \underbrace{U(\eta)}_{=-\log\pi(\eta|y)},\end{aligned}$$

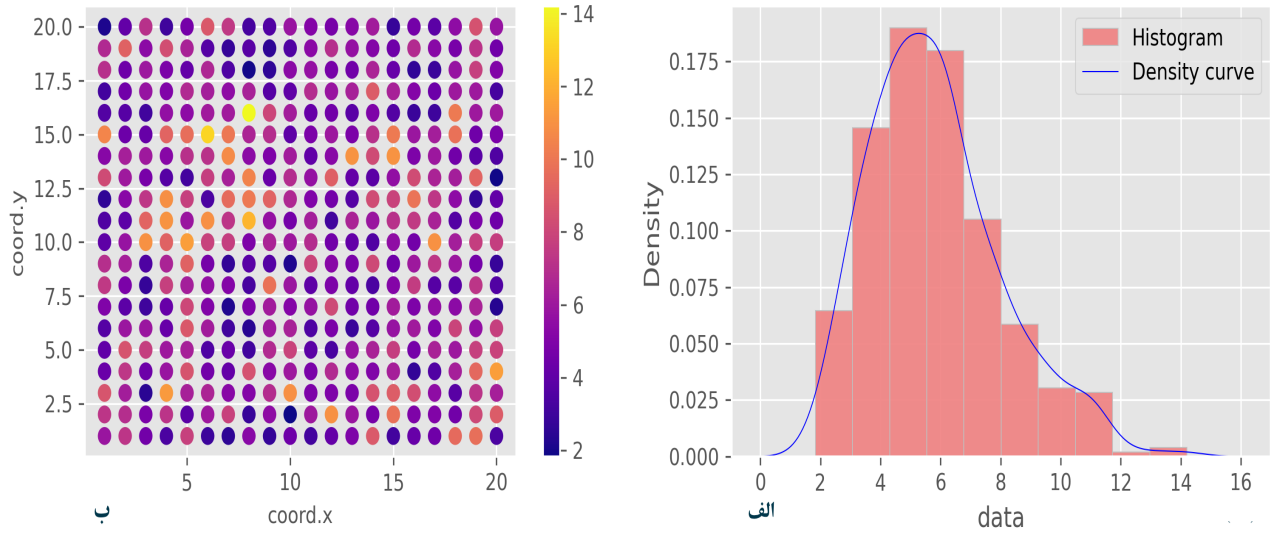
در نظر گرفته می‌شود. معادلات حرکت در فضای همیلتونی به‌صورت

$$\frac{d\eta}{dt} = \nabla_r H = M^{-1} r, \qquad (۵)$$

$$\frac{dr}{dt} = -\nabla_\eta H = -\nabla_\eta U(\eta). \qquad (۶)$$

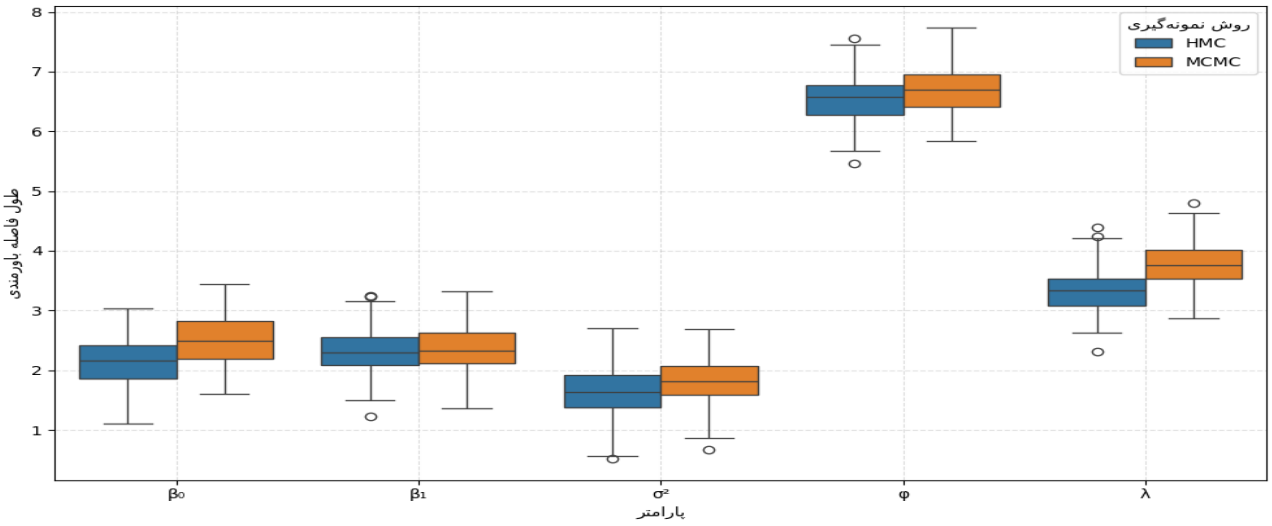
تعریف می‌شوند. با شبیه‌سازی مسیر دینامیکی این سیستم از طریق انتگرال‌گیری عددی، الگوریتم قادر به تولید نمونه‌هایی با همبستگی پایین، نرخ پذیرش بالا و پوشش دقیق‌تر فضای پارامترها است.از آنجا که حل تحلیلی معادلات همیلتونی امکان‌پذیر نیست، روش لیپ فراگ برای انتگرال‌گیری استفاده می‌شود. الگوریتم ۱ مراحل انجام روش *HMC* را مرحله به مرحله بیان می‌کند. استفاده از روش *HMC* مزایایی مانند افزایش نرخ پذیرش، کاهش خودهمبستگی نمونه‌ها و افزایش کارایی نمونه‌گیری در مدل‌های با ابعاد بالا را به همراه دارد. در مدل رگرسیون فضایی چوله، این روش باعث کاهش زمان همگرایی و بهبود دقت برآورد پارامترهای مدل می‌شود. در ادامه برای ارزیابی عملکرد روش پیشنهادی، یک مطالعه شبیه‌سازی ارائه خواهد شد.

[5] Burn-in




**شکل ۱:** یک تحقق از میدان تصادفی چوله نرمال منعطف، الف) هیستوگرام داده‌های فضایی تولید شده، ب) موقعیت داده‌های فضایی تولید شده در شبکه $۲۰ \times ۲۰$.

داده‌ها رسم شده است و هیستوگرام داده‌ها در شکل ۱ (الف) چولگی به راست داده‌ها را نمایش می‌دهد. یک راه تشخیص مدل تغییرنگار داده‌های فضایی استفاده از تغییرنگار تجربی و رسم آن است که براساس آن مدل مناسب برازش شود. تغییرنگار تجربی داده‌های فضایی تولید شده با توجه به برآورد ارائه شده توسط [۷] محاسبه و در شکل ۳ (پ) رسم شده است. با روش کمترین توان‌های دوم خطاها یک مدل نمایی همسانگرد به تغییرنگار تجربی داده‌ها برازش داده شد. سپس با در نظر گرفتن مدل همبستگی نمایی، برآورد بیزی پارامترهای میدان تصادفی بر اساس روش‌های مونت کارلو و همیلتونی محاسبه و در جدول ۱ خلاصه شده‌اند. نتایج جدول ۱ میانگین‌گیری از برآوردهای حاصل از ۱۰۰ تکرار شبیه‌سازی می‌باشد. نتایج جدول ۱ نشان می‌دهد که برآورد بیزی پارامترهای $\beta_0$، $\beta_1$ و $\sigma$ با دقت نسبتا خوبی محاسبه شده‌اند، اما پارامترهای $\varphi$ و $\lambda$ با توجه به مقادیر میانگین توان دوم خطاها ($RMSE$) و انحراف معیار از دقت کمتری برخوردار هستند. روش $HMC$ علاوه بر دقت محاسباتی تقریبا قابل قبول در مقایسه با روش $MCMC$ از سرعت محاسباتی بالایی بهره‌مند است که زمان اجرای آن حدود ۳۵ درصد سریعتر بود. نتایج شبیه‌سازی نشان می‌دهد که مدل چوله منعطف قادر به مدل‌سازی ساختار فضایی همراه با چولگی داده‌ها است و الگوریتم همیلتونی در مقایسه با روش $MCMC$ هم از لحاظ دقت و هم از نظر سرعت اجرا عملکرد بهتری دارد. علاوه بر مقایسه میانگین برآوردها و محاسبه $RMSE$ نحوه عملکرد فواصل باورمندی نیز براساس نرخ پوشش مورد ارزیابی قرار گرفت. نتایج نشان داد که نرخ پوشش برای اکثر پارامترها در حدود ۹۰ درصد تا ۹۶ درصد است که نشان‌دهنده معتبر بودن فواصل در چارچوب برآورد پسین است. همچنین توزیع طول فواصل باورمندی در ۱۰۰ تکرار مستقل نشان داد که الگوریتم $HMC$ نسبت به $MCMC$ طول فواصل کوتاه‌تر و پراکندگی کمتری دارد و این امر بیانگر پایداری و دقت بالاتر در برآورد پارامترها است. برای پارامترهای ساختاری مانند چولگی و پارامتر همبستگی فضایی، هر دو روش با کاهش نرخ پوشش مواجه‌اند که ناشی از حساسیت بالا در برآورد آن‌هاست. در شکل ۲، نمودار جعبه‌ای طول فواصل باورمندی ۹۵ درصد برای پارامترهای مدل براساس ۱۰۰ مجموعه داده مستقل شبیه‌سازی برای دو روش نمونه‌گیری $HMC$ و $MCMC$ نمایش داده شده است. همان‌طور که مشاهده می‌شود، روش $HMC$ در تمامی پارامترها طول فاصله کوتاه‌تری نسبت به $MCMC$ ایجاد کرده است. این موضوع به‌ویژه در پارامترهای حساس مانند $\lambda$ و $\varphi$ نمایان‌تر است که نشان می‌دهد الگوریتم $HMC$ در مدل‌سازی ساختار همبستگی فضایی و چولگی عملکرد کارآمدتری دارد. در مقابل، روش $MCMC$ در برخی پارامترها مانند $\sigma^2$ و $\lambda$، فواصل طولانی‌تر را نشان می‌دهد. در مجموع، این نمودار تأیید می‌کند که الگوریتم $HMC$ با حفظ نرخ پوشش مناسب، فواصل باورمندی کوتاه‌تر و با پراکندگی کمتر تولید می‌کند. در شکل ۳، دو نقشه برای مقایسه بازیابی اثر فضایی ارائه شده‌اند. شکل ۳ (الف) اثر فضایی واقعی تولیدشده در فرایند شبیه‌سازی را نشان می‌دهد و در نهایت، پیشگویی فضایی در این مشبکه منظم بر اساس روش $HMC$ محاسبه و نقشه پیشگویی فضایی روی کل ناحیه




**شکل ۲:** نمودار جعبه‌ای طول فواصل باورمندی ۹۵ درصد بیزی برای ۱۰۰ مجموعه داده شبیه‌سازی.

فضایی با نمودار تراز داده‌های فضایی شبیه‌سازی شده در شکل ۳ (ب) رسم شد. شباهت ساختاری این دو نقشه نشان می‌دهد که مدل قادر به بازسازی الگوی فضایی به‌صورت مناسب است. این شکل نحوه همبستگی مکانی داده‌ها را به‌صورت طیف‌های رنگی مختلفی نشان می‌دهد که چه مناطقی همبستگی مشابه‌ای دارند.

**جدول ۱:** نتایج برآورد بیزی *HMC* و *MCMC* براساس ۱۰۰ تکرار مستقل از مدل چوله گاوسی منعطف. نرخ پوشش ۹۵٪ برای هر پارامتر در ستون‌های آخر گزارش شده است.

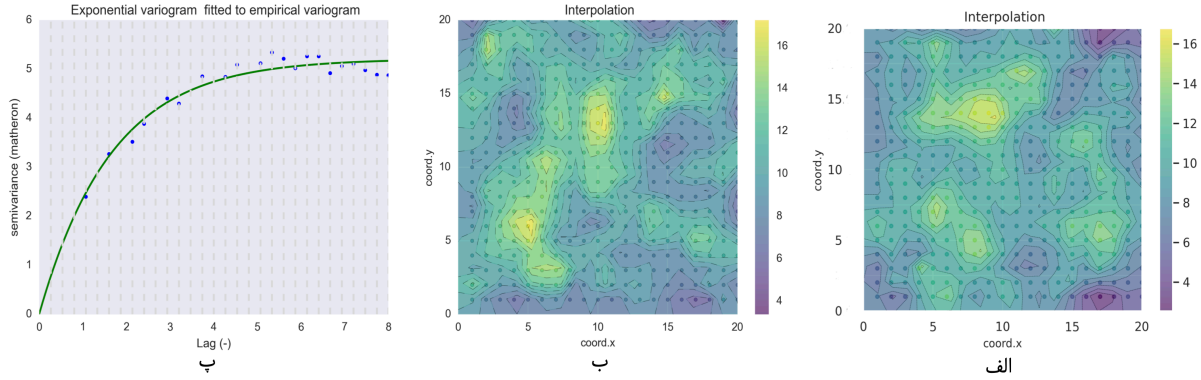
| | | روش *MCMC* | | | روش *HMC* | | |
|---|---|---|---|---|---|---|---|
| پارامتر | مقدار واقعی | برآورد | *RMSE* | نرخ پوشش (%) | برآورد | *RMSE* | نرخ پوشش (%) |
| $\beta_0$ | ۱ | ۰٫۸۹۴ | ۰٫۴۸۶ | ۹۳ | ۰٫۸۸۲ | ۰٫۴۲۹ | ۹۵ |
| $\beta_1$ | ۲ | ۲٫۱۱۷ | ۰٫۵۱۷ | ۹۱ | ۲٫۰۸۹ | ۰٫۵۰۴ | ۹۴ |
| $\sigma^2$ | ۱ | ۱٫۳۱۱ | ۰٫۴۰۱ | ۹۰ | ۰٫۹۷۷ | ۰٫۳۷۰ | ۹۶ |
| $\varphi$ | ۶ | ۵٫۳۱۲ | ۱٫۷۵۱ | ۸۶ | ۵٫۴۱۸ | ۱٫۷۱۲ | ۹۱ |
| $\lambda$ | ۲٫۵ | ۱٫۵۶۳ | ۱٫۰۱۵ | ۸۲ | ۱٫۹۴۱ | ۰٫۹۷۰ | ۸۸ |

## ۵ مطالعه موردی: بررسی غلظت فلزات سنگین در جنوب هلند

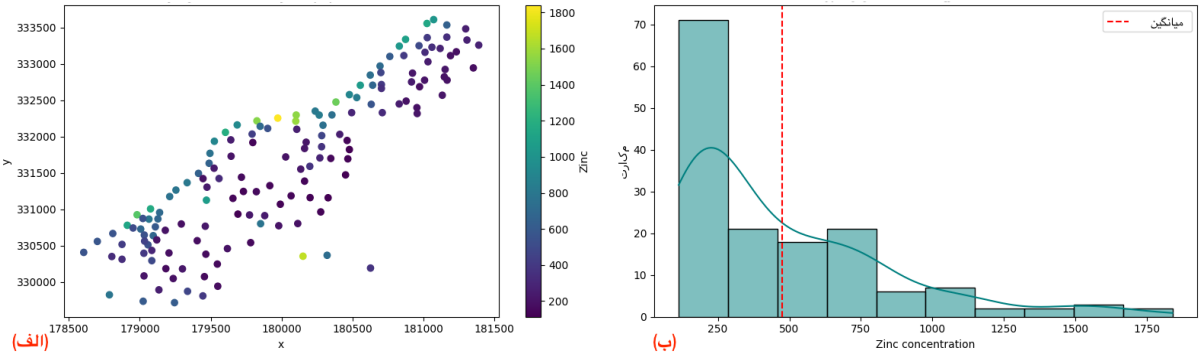
در این مطالعه از مجموعه‌داده‌ی *meuse* موجود در محیط نرم‌افزاری *R* و متعلق به بسته‌ی *sp*، که دارای ویژگی‌های مکانی و چولگی بوده و به‌طور گسترده در تحلیل‌های آماری فضایی استفاده می‌شود، بهره‌برداری شده است. این داده‌ها مربوط به غلظت فلزات سنگین (از جمله روی، سرب، کادمیوم و مس) در نمونه‌های خاک اطراف رودخانه میوس [۶] در جنوب هلند می‌باشند. داده‌ها شامل مختصات مکانی نقاط نمونه‌برداری به همراه مقادیر اندازه‌گیری‌شده متغیرهای محیطی و شیمیایی خاک است. متغیر وابسته در این تحقیق، غلظت عنصر روی[۷] است.

در شکل ۴ (الف) موقعیت مکانی نقاط نمونه‌برداری را در فضای دوبعدی نشان می‌دهد. رنگ هر نقطه بیانگر غلظت عنصر روی در آن موقعیت است. توزیع غلظت‌ها به صورت ناهمگن در منطقه پراکنده شده‌اند و مناطقی با غلظت بالا (به رنگ زرد روشن) در میان نقاط با

[6] Meuse
[7] Zinc




**شکل ۳:** الف) نمودار تراز و اثر فضایی واقعی، ب) نمودار تراز و پیشگویی اثر فضایی، پ) نقاط تغییرنگار تجربی و مدل نمایی برازش شده.




**شکل ۴:** الف) موقعیت مکانی نمونه‌ها، ب) توزیع غلظت زینک.

غلظت پایین دیده می‌شوند که نشانه‌ای از ساختار فضایی و نوسانات منطقه‌ای در داده‌ها است. شکل ۴ (ب) نمودار هیستوگرام داده‌های غلظت روی را نشان می‌دهد. توزیع داده‌ها به وضوح دارای چولگی مثبت (راست‌چوله) است و داده‌ها از توزیع نرمال انحراف قابل توجهی دارند. بنابراین، استفاده از مدل‌های انعطاف‌پذیرتر نظیر میدان تصادفی چوله گاوسی در این شرایط ضروری است. به‌منظور تحلیل دقیق‌تر ساختار مکانی غلظت عنصر روی در خاک، از دو متغیر کمکی که در مجموعه‌داده موجود بودند، یعنی متغیر کادمیوم[8] و سرب[9] استفاده شد. سایر متغیرها به دلیل هم‌خطی، نبود معناداری آماری یا افزایش پیچیدگی مدل حذف شدند. کادمیوم یکی از عناصر فلزی سنگین با منشأ صنعتی که به‌دلیل رفتار مشابه با روی از نظر جذب در خاک، به‌عنوان یک متغیر توضیحی مؤثر وارد مدل شد. عنصر سرب نیز به‌عنوان یک آلاینده‌ی رایج در مناطق صنعتی و کشاورزی در نظر گرفته شد و همبستگی بالایی با روی دارد. برای تحلیل این داده‌ها، از مدل رگرسیون فضایی (۳) استفاده شد. در مدل $\beta_0$، $\beta_1$، $\beta_2$ به ترتیب ضرایب رگرسیون مربوط به عرض از مبدا، متغیر کادمیوم و سرب ، $\lambda$ پارامتر چولگی، $\varphi$ پارامتر همبستگی فضایی و $\sigma$ پارامتر مقیاس هستند. به منظور ارزیابی عملکرد مدل و همچنین بررسی رفتار الگوریتم‌های نمونه‌گیری، دو الگوریتم مونت‌کارلو همیلتونی و الگوریتم متروپلیس هاستینگس (MH) اجرا شدند. برای مقایسه‌ی عملکرد الگوریتم‌ها از دو شاخص استفاده شد: زمان کل صرف‌شده برای نمونه‌گیری پارامترها توسط هر الگوریتم و ریشه میانگین توان دوم خطای پیشگویی بین مشاهدات واقعی و مقادیر برازش شده با استفاده از معیار $(CV-RMSE)$[10] محاسبه شد.

[8] Cadmium
[9] Lead
[10] Cross-validation Root Mean Squared Error

جدول ۲: خلاصه آماری برآورد بیزی پارامترهای مدل با دو الگوریتم *HMC* و متروپلیس–هاستینگس

| نام الگوریتم | پارامتر | میانگین | انحراف معیار | فاصله ۳٪ | فاصله ۹۷٪ | $\hat{R}$ |
|---|---|---|---|---|---|---|
| *HMC* | $\beta_0$ | ۶٫۸۵ | ۹٫۶۸ | −۱۰٫۶۸ | ۲۴٫۸۷ | ۱٫۰۰ |
| | $\beta_1$ | ۱۰٫۷۴ | ۴٫۹۴ | ۰٫۸۷ | ۱۹٫۲۶ | ۱٫۰۰ |
| | $\beta_2$ | ۱۲٫۹۷ | ۴٫۹۴ | ۳٫۹۹ | ۲۲٫۳۲ | ۱٫۰۰ |
| | $\lambda$ | ۲٫۷۲ | ۱٫۱۵ | ۰٫۰۳ | ۵٫۴۱ | ۱٫۰۰ |
| | $\varphi$ | ۰٫۱۷ | ۰٫۰۱ | ۰٫۱۵ | ۰٫۲۰ | ۱٫۰۰ |
| | $\sigma$ | ۶۷٫۶۷ | ۶٫۷۰ | ۵۵٫۶۷ | ۸۰٫۸۸ | ۱٫۰۰ |
| *MH* | $\beta_0$ | ۰٫۵۷ | ۱۰٫۰۸ | −۱۷٫۶۲ | ۲۱٫۰۷ | ۱٫۰۱ |
| | $\beta_1$ | ۱۹٫۲۸ | ۴٫۸۴ | ۹٫۶۳ | ۲۷٫۹۲ | ۱٫۰۱ |
| | $\beta_2$ | ۲۲٫۵۶ | ۵٫۳۱ | ۱۲٫۵۰ | ۳۲٫۲۳ | ۱٫۰۱ |
| | $\lambda$ | ۰٫۹۸ | ۳٫۱۴ | −۵٫۱۳ | ۸٫۲۳ | ۱٫۰۰ |
| | $\varphi$ | ۰٫۵۹ | ۰٫۰۳ | ۰٫۵۵ | ۰٫۶۵ | ۲٫۰۹ |
| | $\sigma$ | ۱۱۱٫۸۵ | ۴٫۶۷ | ۱۰۴٫۲۶ | ۱۲۱٫۳۹ | ۱٫۰۲ |

در جدول ۲ خلاصه‌ای از آماره‌های مربوط به پارامترهای پسین مدل ارائه شده است که با استفاده از دو الگوریتم نمونه‌گیری مونت‌کارلویی برآورد شده‌اند. برای هر پارامتر، میانگین، انحراف معیار، فاصله باورمندی بیزی در بازه ۳٪ تا ۹۷٪ و معیار همگرایی $\hat{R}$ گزارش شده است. مقایسه بین الگوریتم‌ها نشان می‌دهد که *HMC* عملکرد باثبات‌تری از نظر همگرایی داشته است؛ به‌گونه‌ای که مقدار $\hat{R}$ برای همه پارامترها نزدیک به ۱ گزارش شده است. در حالی که در الگوریتم *MH* برای برخی پارامترها به‌ویژه $\varphi$ مقدار $\hat{R}$ بالاتر از ۲ گزارش شده است، که نشانه‌ی همگرایی ضعیف یا نوسان زیاد در زنجیره نمونه‌گیری می‌باشد. این تفاوت تأکید می‌کند که در مدل‌های پیچیده با ساختار فضایی و چولگی، استفاده از الگوریتم‌های پیشرفته‌تر مانند *HMC* توصیه می‌شود. مقایسه ضرایب رگرسیونی $\beta_1$ و $\beta_2$ نشان می‌دهد که الگوریتم MH تمایل به برآورد ضرایب بزرگ‌تری دارد، که می‌تواند به دلیل عدم همگرایی کامل زنجیرها باشد. مقدار $\varphi$ که نقش مهمی در همبستگی فضایی دارد، در الگوریتم HMC برابر با ۰٫۱۷ تخمین زده شده و در MH برابر با ۰٫۵۹ است. همچنین شاخص همگرایی $\hat{R}$ برای این پارامتر در MH بالا است، که نشان‌دهنده عدم همگرایی زنجیرهای متروپلیس–هاستینگس برای این پارامتر است. در حالی که تمام پارامترها در الگوریتم HMC مقدار $\hat{R}$ نزدیک به ۱٫۰۰ دارند، که نشان‌دهنده‌ی همگرایی بسیار خوب این الگوریتم است. در مورد پارامتر مقیاس $\sigma$، تفاوت معناداری میان دو الگوریتم دیده می‌شود، الگوریتم HMC مقدار میانگین $\sigma = ۶۷٫۶۷$ را گزارش کرده، در حالی که MH مقدار بالاتر ۱۱۱٫۸۵ را تخمین زده است، که هم با افزایش انحراف معیار و هم با افزایش بازه فاصله اطمینان همراه است. نتایج جدول ۲ نشان می‌دهد که الگوریتم HMC نه تنها دارای برآوردهای متمرکزتر برای پارامترهای مدل است، بلکه از نظر شاخص‌های همگرایی نیز عملکرد کاملاً قابل قبولی دارد. در مقابل، الگوریتم متروپلیس–هاستینگس برای برخی پارامترها (به‌ویژه $\varphi$) دارای مشکل همگرایی بوده و تخمین‌های آن دارای واریانس بالاتری هستند. در مقایسه عملکرد دو الگوریتم نمونه‌گیری بر اساس شاخص CV-RMSE، مشاهده می‌شود که روش HMC با مقدار خطای ریشه میانگین مربعات برابر با ۴۵٫۹۸ عملکرد مناسب‌تری نسبت به الگوریتم MH با RMSE برابر با ۱۴۷٫۶۶ داشته است. این تفاوت نشان می‌دهد که پیش‌گویی‌های مدل با استفاده از HMC، حدود سه برابر به واقعیت نزدیک‌تر بوده‌اند، در حالی که زمان اجرای الگوریتم MH طولانی‌تر از زمان اجرای HMC بوده است. در اجرای هر دو الگوریتم نمونه‌گیری مونت‌کارلو همیلتونی و متروپلیس هاستینگس، از دو زنجیر نمونه‌گیری استفاده شد و برای هر زنجیر، تعداد ۲۰۰۰ نمونه به عنوان مرحله داغیدن و سپس ۴۰۰۰ نمونه اصلی تولید گردید. بنابراین، در مجموع برای هر الگوریتم، تعداد ۸۰۰۰ نمونه نهایی حاصل شد. بررسی شاخص‌های همگرایی مانند $\hat{R}$ و اندازه مؤثر نمونه نشان داد که الگوریتم HMC با همین تعداد نمونه به همگرایی مناسبی رسیده، اما در مقابل، الگوریتم MH با همین پیکربندی نیازمند تعداد نمونه‌های بیشتری برای همگرایی دقیق‌تر و کاهش خطاهای پیش‌بینی بوده است. این موضوع با مقایسه مقادیر CV-RMSE در دو روش نیز به‌وضوح تأیید می‌شود.

# بحث و نتیجه‌گیری

در تحلیل‌های آماری مبتنی بر رهیافت بیزی، یکی از چالش‌های اصلی در مواجهه با مه‌داده‌ها و مدل‌های پیچیده، ناکارآمدی روش‌های نمونه‌گیری کلاسیکی مانند الگوریتم متروپلیس–هستینگس در همگرایی سریع و نمونه‌برداری مؤثر از توزیع پسین پارامترها است. در چنین شرایطی، استفاده از الگوریتم‌های پیشرفته‌تری مانند الگوریتم مونت کارلو همیلتونی به‌عنوان جایگزین مناسبی توصیه می‌شود. در این مقاله، یک مدل رگرسیون فضایی مبتنی بر میدان تصادفی چوله منعطف با ساختار همبستگی نمایی همسانگرد توسعه داده شد و برای برآورد پارامترهای آن، روش *HMC* در چارچوب بیزی مورد استفاده قرار گرفت. برای ارزیابی عملکرد این روش، از مطالعه شبیه‌سازی در یک مشبکه فضایی منظم استفاده شد و نتایج آن با روش *MCMC* مقایسه گردید. نتایج حاصل از مطالعه شبیه‌سازی نشان داد روش *HMC* قادر است پارامترهای مدل را با دقت قابل قبول و خطای میانگین مربعی پایین برآورد کند. در مقایسه با روش *MCMC* ، *HMC* دارای سرعت اجرای بالاتری (حدود ۳۵ درصد سریع‌تر) و پایداری بهتر در نمونه‌برداری از توزیع‌های پسین است. مدل چوله منعطف به‌خوبی قادر به بازنمایی چولگی و همبستگی مکانی داده‌ها بوده و پیش‌بینی فضایی دقیقی نیز ارائه می‌دهد که البته چالش‌هایی مثل وابستگی مدل به انتخاب اولیه پارامترهای پیشین و حساسیت نتایج به نوع تابع همبستگی و کاهش دقت برآورد در پارامترهای ساختاری مانند $\lambda$ و $\varphi$ در مقایسه با پارامترهای میانگین و واریانس نیز وجود دارد. در مجموع، ترکیب مدل‌های منعطف چوله با الگوریتم‌های نوین نمونه‌گیری بیزی نظیر *HMC*، راهکار مؤثری برای مدل‌سازی داده‌های فضایی پیچیده و نامتقارن فراهم می‌آورد که می‌تواند در کاربردهای مختلف زمین‌آمار، محیط‌زیست، و اپیدمیولوژی فضایی مفید واقع شود. به عنوان پیشنهاد می‌توان مدل و روش به‌کار رفته در این مطالعه را به میدان‌های تصادفی فضایی زمانی تعمیم داد و همچنین می‌توان از روش‌های کارآمدتر نمونه‌گیری مانند الگوریتم‌های مبتنی بر یادگیری ماشین استفاده کرد.

# تقدیر و تشکر

نویسندگان بر خود لازم می‌دانند از سردبیر محترم، دبیر گرامی و داوران ارجمند که با ارائه‌ی نظرات دقیق، پیشنهادهای سازنده و راهنمایی‌های ارزشمند خود، نقش بسزایی در ارتقای کیفیت علمی این مقاله ایفا کردند، صمیمانه قدردانی نمایند. همچنین از تمامی افرادی که به‌طور مستقیم یا غیرمستقیم در انجام این پژوهش همکاری داشته‌اند، سپاسگزاری می‌شود.

# مراجع

[۱] کریمی، الف. و حسینی، ف. (۱۴۰۲). میدان تصادفی چوله نرمال بسته منعطف برای تحلیل داده‌های فضایی چوله. مجله علوم آماری، ۱۷(۲)، ۳۷۱-۳۸۸.

[۲] حسینی، ف. و کریمی، الف. (۱۴۰۳). تحلیل بیزی متغیرهای پنهان در مدل‌های آمیخته خطی تعمیم‌یافته فضایی با میدان تصادفی مانای چوله گاوسی. مجله علوم آماری، ۱۸(۱)، ۵۷-۷۲.

[۳] کریمی، الف. و حسینی، ف. (۱۴۰۳). تحلیل بیز مقداری مدل رگرسیون فضایی چوله بر اساس زیررده منعطفی از توزیع چوله نرمال بسته. مجله علوم آماری، ۱۸(۲)، ۱۲۷-۱۴۲.

[۴] حسینی، ف. و کریمی، الف. (۱۴۰۰). روش‌های مونت‌کارلو همیلتونی برای تحلیل مدل آمیخته خطی تعمیم‌یافته فضایی چوله. مجله اندیشه آماری، ۲۶(۱)، ۳۷-۴۶.

[۵] کریمی، الف. و حسینی، ف. (۱۴۰۰). معرفی یک میدان تصادفی مانای چوله گاوسی. مجله علوم آماری، ۱۵(۲)، ۵۴۹-۵۶۶.

[6] Allard, D., & Naveau, P. (2007). A new spatial skew-normal random field model. *Communications in Statistics—Theory and Methods*, **36**, 1821-1834.

[7] Cressie, N. (1993). *Statistics for Spatial Data*. Wiley, New York.

[8] Hosseini, F., & Karimi, O. (2021). Approximate pairwise likelihood inference in SGLM models with skew normal latent variables. *Journal of Computational and Applied Mathematics*, **398**, 113692.

[9] Karimi, O. (2024). A Hamiltonian Monte Carlo EM algorithm for generalized linear mixed models with spatial skew latent variables. *Statistical Papers*, **65**, 1065-1084.

[10] Márquez-Urbina, O. U., & González-Farías, G. (2022). A flexible special case of the CSN for spatial modeling and prediction. *Spatial Statistics*, **47**, 100556.

[11] Rimstad, K., & Omre, H. (2014). Skew-Gaussian random fields. *Spatial Statistics*, **10**, 43-62.

# Bayesian Assessment of a Skew Spatial Regression Model Based on a Flexible Closed Skew-Normal Random Field using the Hamiltonian Monte Carlo Algorithm

**F. Hosseini[1]* and O. Karimi[2]**

[1,2] Department of Statistics, Faculty of Mathematics, Statistics and Computer Science, Semnan University, Semnan, Iran

**Abstract:**

In practice, spatial data distributions are skewed, which, due to their inherent complexities, poses serious challenges to statistical modeling. Skew-Gaussian random field models offer a flexible framework for analyzing such data; however, many of these models suffer from computational complexity and parameter identifiability issues, which can reduce the accuracy of statistical inference. In this paper, we develop a spatial regression model based on the flexible closed skew-normal distribution. This model benefits from key properties such as full identifiability, closure under marginalization and conditioning, and high flexibility in capturing complex spatial structures. Bayesian inference is performed using the Hamiltonian Monte Carlo algorithm, a modern Markov Chain Monte Carlo method that leverages gradient information from the target distribution to improve acceptance rate and convergence speed. To assess the performance of the proposed model, a simulation study is conducted, comparing the HMC-based estimates with those obtained from classical MCMC methods. The results indicate improved accuracy and computational efficiency for the proposed approach. Moreover, the model demonstrates strong capability in analyzing high-dimensional spatial data.

**Keywords:** Closed Skew Normal Distribution, Identifiability, Spatial Empirical Variogram, Hamiltonian Monte Carlo method.

*Corresponding author: fatemeh.hoseini@semnan.ac.ir