



Variational Bayesian Analysis of Skew Spatial Regression Model Based on a Flexible Subclass of Closed Skew-Normal Distribution

Karimi, O. , Hosseini, F.

Department of Statistics, Semnan University, Semnan, Iran.

Corresponding author: Omid Karimi, omid.karimi@semnan.ac.ir

Received: 28/12/2023 Revised: 22/7/2024 Accepted and Published Online: 25/7/2024.

Introduction

Spatial regression models analyze quantitative spatial responses based on linear and non-linear relationships with explanatory variables. Usually, the spatial correlation of responses is modeled with a Gaussian random field. However, in practice, we encounter skewed responses, which are analyzed using skew-normal distributions. Closed skew-normal (CSN) distribution is one of the extended families of skew-normal distribution, which has similar properties to normal distribution. Karimi and Mohammadzadeh (2012) presented a Bayesian analysis for the skew spatial regression model based on realizing a closed skew Gaussian random field. Tan et al. (2013) applied Bayesian methods for generalized linear models in contrast to hierarchical Bayesian methods, which perform well for complex models. This article presents a hierarchical Bayesian analysis based on a flexible subclass of CSN distributions. Due to the time-consuming calculations of Monte Carlo methods in hierarchical Bayes analysis, the variational Bayes approach approximates the posterior distribution. The proposed model is then implemented and analyzed on Iran's real earthquake data.

Material and Methods

Due to the complexity of the likelihood function, Hierarchical Bayes approaches are used to fit the parameters of the skew spatial regression model, and the variational Bayes method is also used to approximate the posterior distribution. The goal is to find a variational distribution as $q(\eta)$, approximating the posterior distribution of $\pi(\eta|y)$ by minimizing the probability distance between variational and posterior distributions.

Results and Discussion

The skew spatial regression model was used to provide a prediction map of the magnitude of earthquakes in all parts of Iran to observe the most dangerous parts of the country. The earthquake's magnitude was considered as the response variable for each location, and the earthquakes' depth, longitude, and latitude were considered explanatory variables. Then, the variational Bayes method and the proposed estimation methods were implemented on the data. Parameters of Gaussian and skew models were estimated and compared using the variational Bayes method. According to the Bayesian credible intervals at the %95 level, the earthquakes' longitude and latitude were insignificant. However, the depth of the earthquakes significantly affected the magnitude of the earthquakes in the last fifteen years because the %95 interval estimate does not include zero.

Conclusion

This paper presented a skew regression model based on a flexible subset of the CSN distribution for modeling skewed spatial data. The variational Bayes method was a suitable alternative to Monte Carlo because of its high calculation speed. In a simulation study, the proposed spatial regression model was evaluated. The simulation results for 100 data sets in different spatial domains showed that the proposed model with the variational Bayesian method performed well. Furthermore, the variational Bayes method demonstrated a 3 to 4 times faster calculation speed on average compared to MCMC algorithms. Finally, the proposed model was applied to real earthquake data from the last 15 years in Iran, and the Gaussian spatial regression model was also fitted to the data for comparison. The cross-validation results clearly indicated the superior performance of the proposed model. This comparison serves to highlight the effectiveness and potential of the proposed model in the context of the Gaussian spatial regression model.

Keywords: Hierarchical Bayesian approach, Optimal variational distribution, Variogram, Spatial data.

Mathematics Subject Classification (2010): 60G15, 62M30.



©The Author(s). The Publisher is Iranian Statistical Society.

This is an open access article distributed under the terms and conditions of [\(CC BY-NC 4.0\)](https://creativecommons.org/licenses/by-nc/4.0/)

تحلیل بیز مقداری مدل رگرسیون فضایی چوله بر اساس زیر رده منعطفی از توزیع چوله نرمال بسته

امید کریمی، فاطمه حسینی

گروه آمار دانشگاه سمنان

نویسنده مسئول: امید کریمی، omid.karimi@semnan.ac.ir

تاریخ دریافت: ۱۴۰۲/۱۰/۷ تاریخ بازنگری: ۱۴۰۳/۵/۱ تاریخ پذیرش و انتشار: ۱۴۰۳/۵/۴

چکیده: مدل‌های رگرسیون فضایی برای تحلیل پاسخ‌های کمی فضایی براساس روابط خطی و غیرخطی با متغیرهای توضیحی به کار گرفته می‌شوند. معمولاً همبستگی فضایی پاسخ‌ها با یک میدان تصادفی گاوسی بر اساس توزیع نرمال چند متغیره مدل می‌شوند. اما در عمل با پاسخ‌های چوله‌ای مواجه می‌شویم که برای تحلیل آن‌ها از خانواده توزیع‌های چوله نرمال استفاده می‌شود. توزیع چوله نرمال بسته یکی از خانواده‌های گرده توزیع چوله نرمال است که خواص مشابه توزیع نرمال دارد. در این مقاله تحلیل بیز سلسله مراتبی این مدل‌ها براساس یک زیر رده منعطف از توزیع چوله نرمال بسته ارائه می‌گردد. به دلیل زمان‌بر بودن محاسبات روش‌های مونت کارلویی در تحلیل بیز سلسله مراتبی از رهیافت بیز مقداری برای تقریب توزیع پسین استفاده می‌شود. سپس مدل پیشنهادی روی داده‌های واقعی زمین لرزه‌ای کشور ایران پیاده‌سازی و مورد تحلیل قرار می‌گیرد.

واژه‌های کلیدی: رهیافت بیزی سلسله مراتبی، توزیع مقداری بهینه، هم‌تغییرنگار، داده‌های فضایی.

کد موضوع بندی ریاضی (۲۰۱۰): 62M30 ، 60G15



۱ مقدمه

معمولاً در زمینه‌های مختلفی همچون زمین‌شناسی، اپیدمیولوژی، جغرافیا و پزشکی با پاسخ‌های کمی مواجه می‌شویم که دارای همبستگی فضایی هستند و این همبستگی به فضای مورد مطالعه آن‌ها بستگی دارد. برای مدل‌بندی این نوع پاسخ‌ها از مدل‌های رگرسیون فضایی استفاده می‌شود. [انسلین \(۱۹۹۰\)](#) برای لحاظ کردن همبستگی فضایی در مدل‌های رگرسیون روش‌های مختلفی پیشنهاد کرد. [باسو و رینسل \(۱۹۹۴\)](#) تحلیل درست‌نمایی را برای مدل رگرسیونی با خطاهای خودهمبسته فضایی ارائه کردند. [اوه و همکاران \(۲۰۰۲\)](#) رهیافت بیزی سلسله مراتبی را برای این نوع مدل‌ها با داده‌های گمشده در متغیر پاسخ مورد مطالعه قرار دادند. [لی \(۲۰۰۵\)](#) تاثیر کوید ۱۹ بر بازارهای مسکن ایالات متحده را به‌وسیله مدل‌های رگرسیون فضایی بررسی کردند. در بیشتر مباحث نظری مدل‌های رگرسیون فضایی فرض بر این است که داده‌ها تحقیقی از یک میدان تصادفی گاوسی هستند، که در آن خطاها دارای توزیع نرمال است. این فرض ممکن است برای بعضی از داده‌ها برقرار نباشد. به‌عنوان مثال: داده‌های بارندگی در مقاله [کیم و مالیک \(۲۰۰۴\)](#)، [داده‌های لرزه‌ای در مقاله کریمی و همکاران \(۲۰۱۰\)](#) و [داده‌های آلودگی هوا که توسط کریمی و محمدزاده \(۲۰۱۲\)](#) مورد استفاده قرار گرفته است. معمولاً در عمل با پاسخ‌هایی مواجه می‌شویم که دارای چولگی هستند و برای مدل‌کردن آن‌ها از توزیع‌های چوله نرمال استفاده می‌شود. [کریمی و محمدزاده \(۲۰۱۲\)](#) تحلیل بیزی برای مدل رگرسیون فضایی چوله بر اساس تحقیقی از یک میدان تصادفی چوله گاوسی بسته ارائه کردند. اما این میدان‌های تصادفی دارای شرط سازگاری حاشیه‌ای نبودند ([کریمی، ۲۰۲۳؛ کریمی و حسینی، ۱۴۰۰](#)). [تان و همکاران \(۲۰۱۳\)](#) روش‌های بیز مقداری^۱ را برای مدل‌های خطی تعمیم‌یافته در مقابل روش‌های بیزی سلسله مراتبی به‌کار گرفتند که برای مدل‌های پیچیده عملکرد مناسبی دارند. [ژانگ و همکاران \(۲۰۲۳\)](#) نیز روش بیز مقداری را برای برازش مدل‌های فضا حالت^۲ ارائه کردند. [کابرال و همکاران \(۲۰۲۴\)](#) رهیافت بیز مقداری را برای برازش مدل‌های پنهان فضایی با تقریب لاپلاس به‌کار گرفتند و نشان دادند که زمان اجرا این رهیافت سریع‌تر از رهیافت‌های دیگر است.

در این مقاله تحلیل بیز مقداری روی مدل‌های رگرسیون فضایی بر اساس یک زیر رده منعطف از توزیع چوله نرمال بسته^۳ (CSN) بیان می‌شود که دارای شرط سازگاری حاشیه‌ای و خوش تعریف است. ساختار مقاله به این صورت است که در بخش ۲ یک زیر رده از توزیع CSN بیان می‌گردد. سپس مدل رگرسیون فضایی چوله به همراه تحلیل بیز مقداری و سلسله مراتبی آن در بخش ۳ ارائه می‌گردد. در بخش ۴ یک مطالعه شبیه‌سازی برای ارزیابی مدل پیشنهادی با رهیافت بیز مقداری بیان می‌شود و روش‌های پیشنهادی برای تحلیل یک مثال واقعی مربوط به داده‌های زمین لرزه‌ای کشور طی ۱۵ سال اخیر به کار می‌رود. در نهایت بحث و نتیجه‌گیری ارائه می‌شود.

¹Variational Bayes²State Space Models³Closed Skew Normal

۲ توزیع چوله نرمال بسته

متغیر تصادفی Y دارای توزیع CSN است (گزنالس و همکاران، ۲۰۰۴)، اگر دارای تابع چگالی

$$f_Y(\mathbf{y}) = \phi_p(\mathbf{y}; \boldsymbol{\mu}, \Sigma) \frac{\Phi_q(\Gamma(\mathbf{y} - \boldsymbol{\mu}); \boldsymbol{\nu}, \Delta)}{\Phi_q(\mathbf{0}; \boldsymbol{\nu}, \Delta + \Gamma \Sigma \Gamma')}, \quad (۱)$$

باشد و با نماد $rmCSN_{p,q}(\boldsymbol{\mu}, \Sigma, \Gamma, \boldsymbol{\nu}, \Delta)$ نمایش داده می‌شود، که در آن $\boldsymbol{\mu} \in R^p$ ، Σ ماتریس معین مثبت $p \times p$ است که شامل پارامتر مقیاس است، Γ ماتریس $q \times p$ است که شامل پارامتر چولگی است، $\boldsymbol{\nu} \in R^q$ و Δ یک ماتریس $q \times q$ معین مثبت است. Φ_q و ϕ_p به ترتیب تابع چگالی و تابع توزیع تجمعی توزیع نرمال p و q متغیره هستند. توزیع CSN تحت تبدیلات خطی و مجموع بردارهای تصادفی مستقل بسته است. هر چند این توزیع دارای خواص بسته بودن و انعطاف‌پذیر است اما تابع چگالی آن دارای پیچیدگی و پارامترهای زیادی است. به عنوان مثال برای q های بزرگ محاسبه Φ_q دشوار و چالش برانگیز است. به همین خاطر یک زیر رده از این توزیع بر اساس ایده مارکر و گزنالس (۲۰۲۲) برای تحلیل مدل رگرسیون فضایی چوله ارائه می‌شود. برای اطلاعات بیشتر در مورد این زیر رده منعطف و نحوه تعریف یک میدان تصادفی فضایی چوله خوش تعریف (دارا بودن خاصیت سازگاری حاشیه‌ای) به مقاله کریمی و حسینی (۱۴۰۰) مراجعه شود. فرض کنید متغیر تصادفی X_i دارای توزیع CSN به صورت $X_i \sim CSN_{1,1}(\circ, 1, \lambda_i, \circ, 1)$ و X_i به ازای $i = 1, \dots, n$ مستقل باشند، آنگاه تابع چگالی X به صورت $f_X(\mathbf{x}) = \mathfrak{r}^n \phi_n(\mathbf{x}; \mathbf{0}, \mathbf{I}_n) \Phi_n(\Gamma \mathbf{x})$ به دست می‌آید که در آن $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)'$ ، $\Gamma = \text{Diag}(\boldsymbol{\lambda})$ ، بنابراین $\mathbf{X} \sim CSN_{n,n}(\circ, \mathbf{I}_n, \Gamma, \mathbf{0}, \mathbf{I}_n)$ و $\Phi_n(\Gamma \mathbf{x}) = \prod_{i=1}^n \Phi(\lambda_i x_i)$ در آن صورت $\lambda_i \in R$ و $\Phi_n(\Gamma \mathbf{x}) = \prod_{i=1}^n \Phi(\lambda_i x_i)$ ، $b = (\mathfrak{r}/\pi)^{\frac{1}{2}}$ ، $\boldsymbol{\tau} = (\tau_1, \dots, \tau_n)'$ ، $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)'$ که در آن $V(\mathbf{X}) = \Omega$ و $E(\mathbf{X}) = b\boldsymbol{\delta}$ ، $\tau_i = (1 - b^2 \delta_i^2)^{-\frac{1}{2}}$ ، $\delta_i = \lambda_i (1 + \lambda_i^2)^{-\frac{1}{2}}$ ، $\Omega = \text{Diag}(\boldsymbol{\tau})$ و $\boldsymbol{\tau}_i$ را می‌توان به صورت $\mathbf{X}$ بردار تصادفی را می‌توان به صورت $\mathbf{Z} = \Omega^{-1}(\mathbf{X} - b\boldsymbol{\delta})$ استاندارد کرد، به این ترتیب $E(\mathbf{Z}) = \mathbf{0}$ و $V(\mathbf{Z}) = \mathbf{I}_n$ است. در ادامه بنابر بسته بودن توزیع‌های CSN نسبت به تبدیلات خطی، مدل رگرسیون فضایی چوله تعریف و برآورد بیزی آن ارائه می‌شود.

۳ مدل رگرسیون فضایی چوله

با در نظر گرفتن بردار تصادفی استاندارد شده $\mathbf{Z}$ ، مدل رگرسیون فضایی چوله برای بردار پاسخ $\mathbf{y}_n$ در n موقعیت فضایی $(s_1, \dots, s_n)$ به صورت

$$\mathbf{y}_n = H\boldsymbol{\beta} + \sigma C\mathbf{Z}. \quad (۲)$$

تعریف می‌شود، که در آن H یک ماتریس $n \times p$ از متغیرهای توضیحی، $\boldsymbol{\beta}$ یک بردار p بعدی از ضرایب رگرسیونی، $\sigma > 0$ پارامتر مقیاس و C یک ماتریس $n \times n$ که از تجزیه چولسکی ماتریس همبستگی فضایی به صورت

$\Sigma_n = CC'$ به دست می‌آید. در واقع Σ_n ماتریس کوواریانس فضایی بردار تصادفی $\mathbf{y}_n$ با مولفه‌های σ_{ij} است. $\sigma^2 \rho(y(s_i), y(s_j); \varphi)$ یک تابع همبستگی فضایی با پارامتر دامنه فضایی φ است. بنابراین بردار پاسخ‌های فضایی $\mathbf{y}_n$ با توجه به خاصیت تبدیلات خطی توزیع CSN دارای توزیع CSN به صورت $\mathbf{y}_n \sim \text{CSN}_{n,n}(\boldsymbol{\mu}_y, \sigma_y, \frac{1}{\sigma} \Gamma \Omega C^{-1}, \mathbf{0}, \mathbf{I}_n)$ است، که در آن $\boldsymbol{\mu}_y = H\boldsymbol{\beta} - b\sigma C\Omega^{-1}\boldsymbol{\delta}$ و $\Sigma_y = \sigma^2 C\Omega^{-1}C'$ از این رو تابع چگالی به صورت

$$f_{\eta}(\mathbf{y}) = \tau^n \phi_n(\mathbf{y}; \boldsymbol{\mu}_y, \sigma_y) \Phi_n\left(\frac{1}{\sigma} \Gamma \Omega C^{-1}(\mathbf{y} - \boldsymbol{\mu}_y)\right)$$

است، که در آن $\boldsymbol{\beta} = (\boldsymbol{\beta}, \sigma, \varphi, \boldsymbol{\lambda})'$ پارامترهای مدل است. لگاریتم تابع درست‌نمایی به صورت

$$\begin{aligned} \ell(\boldsymbol{\beta}|\mathbf{y}) &= \log(f_{\eta}(\mathbf{y})) \\ &= n \log(\tau) - \frac{n}{\tau} \log(\tau \pi) - \frac{V'V}{\tau} - n \log(\sigma) - \log(|C|) \\ &\quad + \sum_{i=1}^n \{ \log(\Phi(\lambda_i v_i)) + \log(\tau_i) \}, \end{aligned} \quad (3)$$

ساده می‌شود، که در آن $V = \frac{1}{\sigma} \Omega C^{-1}(\mathbf{y} - \boldsymbol{\mu}_y)$ و v_i مولفه i ام بردار V است. با توجه به پیچیدگی تابع درست‌نمایی برای برازش پارامترهای مدل رگرسیون فضایی چوله از رهیافت‌های بیز سلسله مراتبی استفاده می‌شود و رهیافت بیز مقداری نیز برای تقریب توزیع پسین به کار گرفته می‌شود که در ادامه به آن می‌پردازیم.

۳.۱ برازش مدل

برای برآورد پارامترهای مدل رگرسیون فضایی (۲) از رهیافت بیزی سلسله مراتبی با توجه به پیچیدگی تابع درست‌نمایی (۳) استفاده می‌شود. با در نظر گرفتن پیشین‌های معمول برای پارامترها به صورت

$$\boldsymbol{\beta} \sim N_p(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \sigma_{\boldsymbol{\beta}}^2 \mathbf{I}_p), \quad \sigma \sim IG(a, b), \quad \varphi \sim G(\alpha, \gamma), \quad \lambda_i \sim N(\mu_{\lambda_i}, \sigma_{\lambda_i}^2),$$

توزیع پسین به صورت

$$\begin{aligned} \pi(\boldsymbol{\beta}|\mathbf{y}) &\propto \phi_n(\mathbf{y}; \boldsymbol{\mu}_y, \sigma_y) \Phi_n\left(\frac{1}{\sigma} \Gamma \Omega C^{-1}(\mathbf{y} - \boldsymbol{\mu}_y)\right) \phi_p(\boldsymbol{\beta}; \boldsymbol{\mu}_{\boldsymbol{\beta}}, \sigma_{\boldsymbol{\beta}}^2 \mathbf{I}_p) \pi(\sigma; a, b) \\ &\times \pi(\varphi; \alpha, \gamma) \prod_{i=1}^n \phi(\lambda_i; \mu_{\lambda_i}, \sigma_{\lambda_i}^2) \pi(\boldsymbol{\mu}_{\boldsymbol{\beta}}) \pi(\sigma_{\boldsymbol{\beta}}^2) \\ &\times \pi(a) \pi(b) \pi(\alpha) \pi(\gamma) \prod_{i=1}^n \pi(\mu_{\lambda_i}) \pi(\sigma_{\lambda_i}^2), \end{aligned}$$

خلاصه می‌شود، که در آن $\pi(\cdot)$ توزیع پیشین ابر پارامترهای $\mu_\beta, \sigma_\beta^2, a, b, \alpha, \gamma, \mu_{\lambda_i}$ و $\sigma_{\lambda_i}^2$ است. توزیع پیشین ابر پارامترها به صورت

$$\begin{aligned}\mu_\beta &\sim N_p(\mathbf{0}, \delta \mathbf{I}_p), \quad \sigma_\beta^2 \sim \text{IG}(\nu, \tau), \quad a, b \sim U(1, \delta), \\ \alpha, \gamma &\sim U(1, 10), \quad \mu_{\lambda_i} \sim N(0, \delta), \quad \sigma_{\lambda_i}^2 \sim \Gamma(0.5, 0.5),\end{aligned}$$

لحاظ شده‌اند. لازم به ذکر است که مقادیر معلوم برای هر یک از پارامترهای ابر پیشین‌ها طوری در نظر گرفته شده‌اند که حساسیت پیشین‌ها با بالا بردن پراکندگی آن‌ها کم شود. به عنوان مثال در توزیع پیشین ابر پارامتر μ_{λ_i} واریانس ۵ لحاظ شده است که پارامتر میانگین توزیع پیشین پارامتر چولگی از اطلاع کمی برخوردار باشد و حساسیت آن در استنباط کمتر شود. با توجه به پیچیده بودن توزیع پسین لازم است از روش‌های مونت کارلویی استفاده شود. در این مقاله از رهیافت بیز مقداری (تان و همکاران، ۲۰۱۳) برای تقریب توزیع پسین با سرعت محاسبات نسبتاً بالا به عنوان یک جایگزین مناسب برای روش‌های MCMC استفاده می‌شود.

۳.۲ رهیافت بیز مقداری

هدف بیز مقداری یافتن توزیع مقداری $q(\beta)$ است، که توزیع پسین $\pi(\beta|\mathbf{y})$ را از طریق مینیم کردن فاصله احتمال بین توزیع مقداری و توزیع پسین تقریب می‌زند (تان و همکاران، ۲۰۱۳). فاصله احتمال توسط رابطه کولبک - لایبلر^۱ (KL) به صورت

$$\begin{aligned}\text{KL}(q(\beta) \parallel \pi(\boldsymbol{\eta} | \mathbf{y})) &= \int \ln \left(\frac{q(\boldsymbol{\eta})}{\pi(\boldsymbol{\eta} | \mathbf{y})} \right) q(\boldsymbol{\eta}) d\boldsymbol{\eta} \\ &= \mathbb{E}_q[\ln q(\boldsymbol{\eta})] - \mathbb{E}_q[\ln \pi(\boldsymbol{\eta} | \mathbf{y})] \\ &= \mathbb{E}_q[\ln q(\boldsymbol{\eta})] - \mathbb{E}_q[\ln \pi(\boldsymbol{\eta}, \mathbf{y})] + \ln f(\mathbf{y}).\end{aligned}\tag{۴}$$

اندازه‌گیری می‌شود. در این روش با مینیم کردن رابطه KL به توزیع هدف

$$q^*(\boldsymbol{\eta}) = \arg \min_q \text{KL} (q(\boldsymbol{\eta}) \parallel \pi(\boldsymbol{\eta} | \mathbf{y})).$$

به عنوان تقریبی از توزیع پسین می‌رسد. چون $\ln f(\mathbf{y})$ شکل بسته‌ای ندارد، رابطه KL از نظر تحلیلی قابل حل نیست. معادله (۴) را به صورت

$$\text{RKL}(q(\boldsymbol{\eta})) = \ln f(\mathbf{y}) - \text{KL}(q(\boldsymbol{\eta}) \parallel \pi(\boldsymbol{\eta} | \mathbf{y})).\tag{۵}$$

^۱Kullback-Leibler

می‌نویسیم. از آنجا که رابطه KL همیشه مثبت است، معادله (۴) نشان می‌دهد که توزیع مقداری بهینه را می‌توان با ماکسیم کردن رابطه (۵) به‌دست آورد. توزیع مقداری باید توسط تحلیلگر انتخاب شود. که معمولاً از تقریب‌های گاوسی استفاده می‌شود. یک راه حل برای تعیین توزیع مقداری استفاده از بیز مقداری میدان میانگین^۱ است. در این روش چگالی توزیع مقداری به‌صورت حاصلضرب مولفه‌های مجزایه صورت

$$q(\eta) = \prod_{j=1}^J q(\eta_j) = q(\beta)q(\sigma)q(\phi)q(\lambda),$$

مشخص می‌شود، که در آن $j \in \{1, \dots, J\}$ مجموعه اندیس‌گذار پارامترهای مدل است. این مشخصه‌سازی باعث استقلال پسینی بین پارامترها می‌شود. اورمرود و وند (۲۰۱۰) چگالی مقداری بهینه را به‌صورت $q^*(\eta_j) \propto \exp[\mathbb{E}_{-\eta_j}(\log \pi(\mathbf{y}, \eta))]$ به‌دست آوردند، که در آن نماد $-\eta_j$ یعنی همه مولفه‌های بردار η بجز j امین مولفه آن است و

$$\mathbb{E}_{-\eta_j}(\log \pi(\mathbf{y}, \eta)) = \int \prod_{\substack{k=1 \\ k \neq j}}^J q(\eta_k) \log \pi(\mathbf{y}, \eta) d\eta_1 \dots d\eta_{j-1} d\eta_{j+1} \dots d\eta_J.$$

در صورتی که $q^*(\eta_j)$ دارای شکل توزیعی مشخصی باشد، توزیع مقداری آن به‌راحتی محاسبه می‌شود و به آن مزدوج شرطی گفته می‌شود.

۴ مطالعه شبیه‌سازی

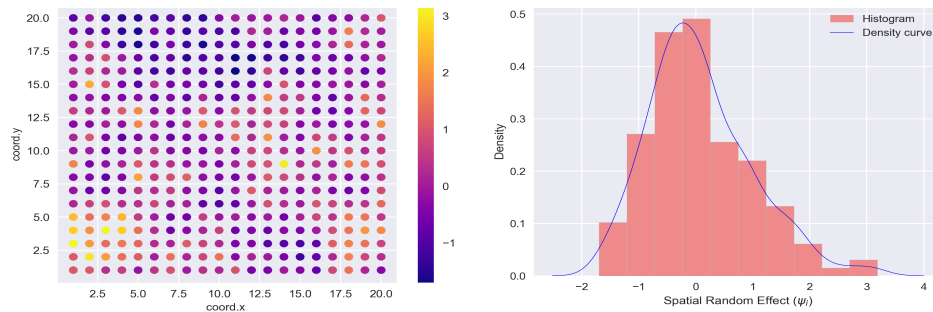
در این بخش یک مطالعه شبیه‌سازی بر اساس مدل رگرسیون فضایی چوله (۲) در شبکه‌های منظم 5×5 ، 10×10 و 20×20 با رهیافت بیز مقداری برای ارزیابی مدل پیشنهادی صورت می‌گیرد. برای این منظور متغیر پاسخ را از مدل رگرسیون فضایی چوله به‌صورت زیر تولید می‌کنیم:

$$y_i = \beta_0 + \beta_1 x_i + \psi_i, \quad i = 1, 2, \dots, n \in \{25, 100, 400\},$$

که در آن متغیر توضیحی x_i از توزیع نرمال با میانگین صفر و انحراف معیار یک تولید می‌شود و ψ_i مولفه i ام عبارت σCZ در مدل (۲) است. ساختار همبستگی فضایی همسانگرد نمایی برای محاسبه ماتریس C و پارامترهای مدل را با مقادیر معلوم $\beta_0 = 5$ ، $\beta_1 = 2$ ، $\sigma^2 = 1$ ، $\varphi = 6$ ، $\lambda = 3$ ، $\lambda_i = \lambda$ در نظر می‌گیریم. همچنین برای ارزیابی عملکرد مدل پیشنهادی، مدل رگرسیون فضایی گاوسی به عنوان یک مدل رقیب پایه جهت مقایسه به کار گرفته می‌شود.

¹Mean Field Variational Bayes

اگر در مدل چوله پارامتر چولگی λ را صفر قرار دهیم، مدل گاوسی به دست می آید که نشان می دهد مدل رگرسیون فضایی گاوسی حالت خاص مدل رگرسیون فضایی چوله است. از این رو فرایند شبیه سازی را ۱۰۰ بار تکرار کردیم و یک نمونه از اثرات فضایی شبیه سازی شده مشبکه ی منظم 20×20 در شکل ۱ رسم شده است. برای ارزیابی برآورد بیز مقداری



(ب)

(الف)

شکل ۱. نمونه ای از داده های شبیه سازی شده مدل رگرسیون فضایی چوله: (الف) هیستوگرام اثرات فضایی چوله، (ب) نمودار پراکنش اثرات فضایی در مشبکه منظم 20×20 .

جدول ۱. نتایج برآورد بیز مقداری برای دو مدل رگرسیون فضایی چوله و مدل گاوسی براساس ۱۰۰ مجموعه داده شبیه سازی شده در مشبکه منظم 20×20

مدل گاوسی			مدل چوله			واقعی	پارامتر
RMSE	انحراف معیار	برآورد	RMSE	انحراف معیار	برآورد		
۰/۴۱۳	۰/۷۲۴	۵/۸۸۲	۰/۳۷۶	۰/۶۳۸	۵/۳۷۲	۵	β_0
۰/۴۱۹	۰/۵۱۳	۲/۶۱۱	۰/۱۲۲	۰/۲۵۲	۱/۸۹۳	۲	β_1
۰/۹۱۱	۰/۱۴۷	۱/۸۲۲	۰/۸۲۸	۰/۰۵۱	۱/۳۰۶	۱	σ^2
۱/۸۷۱	۱/۵۴۴	۴/۷۵۸	۰/۱۳۳	۰/۹۵۷	۶/۲۷۷	۶	φ
—	—	—	۰/۰۲۷	۰/۳۱۳	۲/۵۲۹	۳	λ

پارامترهای دو مدل از معیار ریشه دوم میانگین توان دوم خطاها $RMSE(\hat{\eta}) = \sqrt{\frac{1}{m} \sum_{i=1}^m (\hat{\eta}_i - \eta_i)^2}$ با ۱۰۰ مجموعه داده شبیه سازی استفاده شده است، که در آن $m = 100$ تعداد مجموعه داده های شبیه سازی برای هر حالت و $\hat{\eta}_i$ برآورد بیز مقداری پارامترهای مدل برای مجموعه داده شبیه سازی شده i ام است. نتایج برآورد پارامترها برای ۱۰۰ مجموعه داده شبیه سازی شده با مشبکه منظم 20×20 در جدول ۱ ارایه شدند. نتایج نشان می دهد که مدل رگرسیون فضایی چوله بر اساس رهیافت بیز مقداری نسبت به مدل گاوسی به خوبی عمل نموده است. البته نتایج RMSE بیان می کنند که پارامترهای همبستگی فضایی در دو مدل از دقت کمتری برخوردار هستند. با توجه به نتایج می توان گفت که دقت برآورد پارامترهای مدل با رهیافت بیز مقداری با توجه به سرعت محاسبات بالاتر

۳۷۰ تحلیل بیز مقداری مدل رگرسیون فضایی چوله

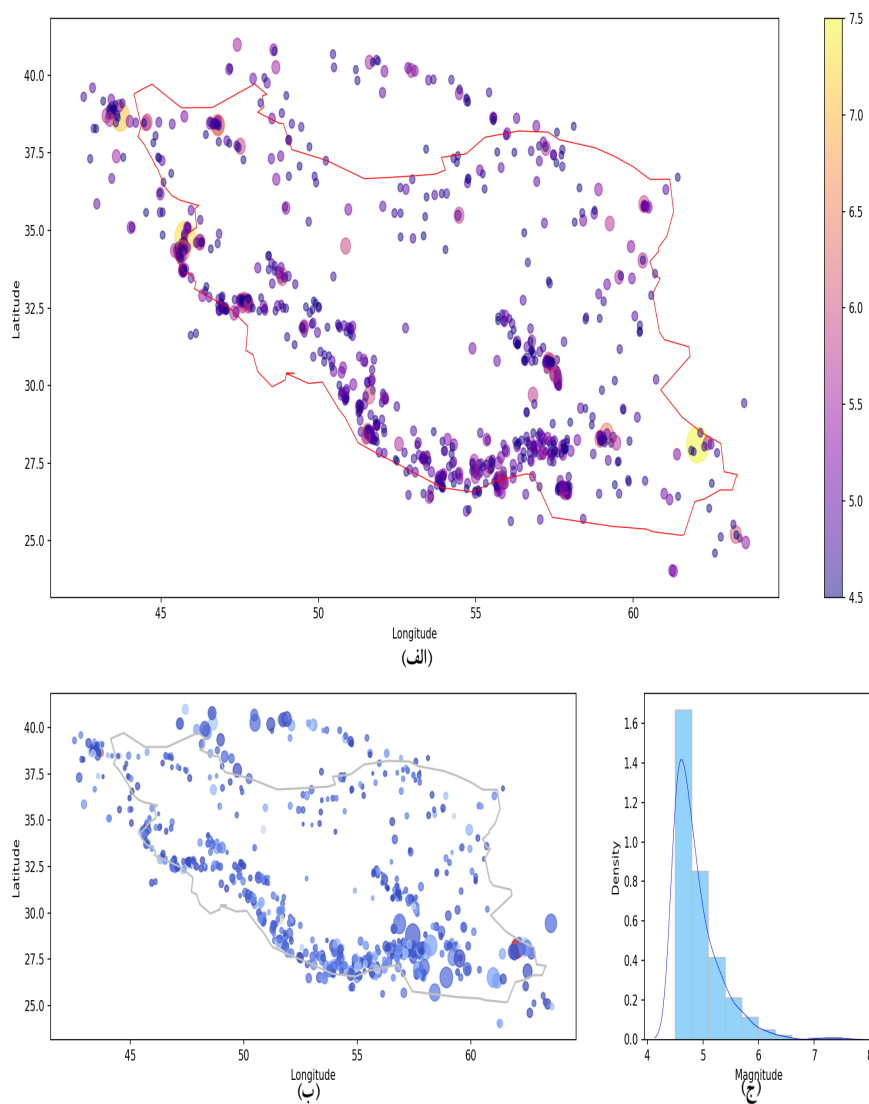
نسبت به روش‌های MCMC نسبتاً خوب است. لازم به ذکر است که برای یک مجموعه داده شبیه‌سازی شده در این مطالعه شبیه‌سازی به‌طور متوسط رهیافت بیز مقداری ۳ الی ۴ برابر سریعتر است. همچنین نتایج روش بیز مقداری دو مدل رگرسیون فضایی چوله و گاوسی برای شبکه‌های منظم 5×10 و 10×10 در جدول ۲ خلاصه شده‌اند. نتایج RMSE نشان می‌دهد که برآورد بیز مقداری برای شبکه‌های منظم با حجم کم نیز عملکرد مناسبی دارد و می‌توان گفت که با افزایش حجم نمونه برآوردهای ضرایب رگرسیونی و دامنه فضایی نسبتاً بهتر می‌شود.

جدول ۲. نتایج برآورد بیز مقداری برای دو مدل رگرسیون فضایی چوله و مدل گاوسی براساس ۱۰۰ مجموعه داده‌ی شبیه‌سازی شده در شبکه‌های منظم 5×10 و 10×10

مدل گاوسی		مدل چوله							
10×10		5×5		10×10		5×5			
RMSE	برآورد	RMSE	برآورد	RMSE	برآورد	RMSE	برآورد	واقعی	پارامتر
۰.۵۳۳	۵.۹۸۲	۰.۹۰۳	۶.۵۲۲	۰.۳۶۵	۵.۶۱۵	۰.۵۱۱	۵.۹۶۱	۵	β_0
۰.۶۰۹	۲.۸۶۸	۰.۷۱۲	۲.۹۵۵	۰.۱۲۱	۱.۷۱۲	۰.۲۸۸	۱.۶۶۲	۲	β_1
۱.۰۵۵	۱.۸۹۹	۱.۰۴۹	۱.۹۸۵	۰.۹۲۲	۱.۴۳۷	۱.۰۸۷	۱.۴۹۷	۱	σ^2
۱.۸۹۷	۴.۶۶۹	۱.۸۸۵	۴.۵۸۷	۰.۴۹۱	۶.۸۶۵	۱.۹۵۶	۷.۵۲۲	۶	φ
—	—	—	—	۰.۰۳۹	۲.۴۴۳	۰.۰۴۵	۲.۴۷۸	۳	λ

۵ تحلیل داده‌های زمین لرزه‌ای ایران

در این بخش به تحلیل داده‌های زمین لرزه‌ای کشور طی ۱۵ سال اخیر بر اساس مدل رگرسیون فضایی چوله و روش‌های معرفی شده می‌پردازیم. داده‌های زمین لرزه‌ای که توسط مرکز لرزه‌نگاری کشور (مؤسسه‌ی ژئوفیزیک تهران) طی سال‌های ۲۰۰۶ الی ۲۰۲۰ ثبت شده است برای تحلیل در این مقاله در نظر گرفته شده است. داده‌های زمین لرزه‌ای شامل زمان وقوع زلزله (Date)، عرض جغرافیایی (Lat)، طول جغرافیایی (Long)، عمق (Depth)، شهر (City)، استان (Province) و بزرگی زمین لرزه (Mg) است. نقشه‌ی نقاط بزرگی زمین لرزه‌های بالای ۴/۵ در شکل ۲(الف) به همراه هیستوگرام آن‌ها (شکل ۲(ج)) و عمق زمین لرزه‌ها (شکل ۲(ب)) رسم شده‌اند. با توجه به شکل ۲ می‌توان همبستگی مکانی (فضایی) داده‌ها را مشاهده کرد. با توجه به نقشه عمق زمین لرزه‌ها بیشتر زمین لرزه‌ها عمیق در جنوب کشور رخ داده است. در مطالعات زمین لرزه‌ای، بررسی عمق زمین لرزه‌ها از اهمیت زیادی برخوردار است چرا که هرچه عمق زمین لرزه‌ها کم‌تر باشد شدت خسارت آن بیش‌تر می‌شود. همچنین چولگی در نمودار هیستوگرام بزرگی زمین لرزه‌ها دیده می‌شود. در ادامه مدل رگرسیون فضایی چوله (۲) به‌منظور آرایه نقشه پیشگویی بزرگی زمین لرزه‌ها در کل نقاط کشور برای مشاهده نقاط پرخطر کشور به‌کار گرفته می‌شود. با در نظر گرفتن بزرگی زمین لرزه به عنوان متغیر پاسخ (Mg_i) برای هر موقعیت s_i و عمق (D_i)، طول جغرافیایی ($Long_i$) و

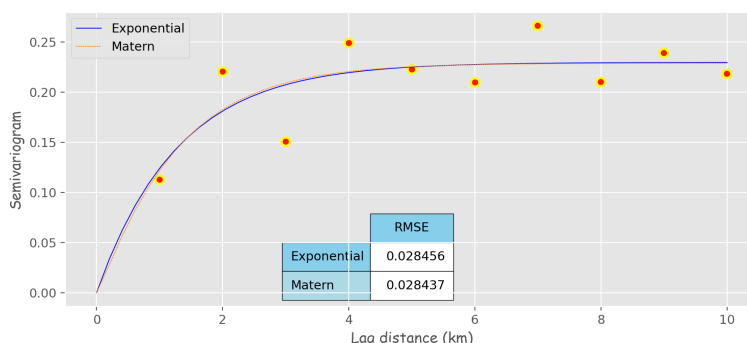


شکل ۲. (الف) نقشه موقعیت زمین لرزه‌های کشور بیشتر از ۴٫۵ طی پانزده سال اخیر، (ب) نقشه عمق زمین لرزه‌ها و (ج) نمودار هیستوگرام آن‌ها

عرض جغرافیایی (Lat_i) زمین لرزه‌ها به عنوان متغیرهای توضیحی مدل رگرسیون فضایی به صورت

$$Mg_i = \beta_0 + \beta_1 D_i + \beta_2 Long_i + \beta_3 Lat_i + \psi_i, \quad i = 1, 2, \dots, \quad (6)$$

تعریف می‌شود، که در آن ψ_i مولفه i ام عبارت σCZ در مدل (۲) است. در این مدل به دلیل شناسایی پذیر بودن پارامتر چولگی به صورت $\lambda_i = \lambda$ در نظر گرفتیم. برای تعیین مدل همبستگی فضایی داده‌ها از تغییرنگار



شکل ۳. نمودار تغییرنگار تجربی داده‌های زمین لرزه‌های کشور

تجربی داده‌ها استفاده کرده‌ایم. شکل ۳ نمودار تغییرنگار تجربی به همراه برازش دو مدل تغییرنگار نمایی و ماترن را نشان می‌دهد. همانطور که مشاهده می‌شود مدل نمایی به خوبی برازش شده است بنابراین این مدل برای تعیین ساختار همبستگی فضایی ماتریس Σ_n در نظر گرفته می‌شود. سپس رهیافت بیز مقداری به همراه روش‌های برآورد پیشنهادی در بخش قبل روی داده‌ها پیاده‌سازی شد. بدلیل پیچیدگی توزیع پسین در روش بیز سلسله مراتبی لازم است از روش‌های MCMC استفاده کرد که این روش‌ها نسبت به بیز مقداری زمان‌بر بود به طوری که در این مطالعه زمان اجرا با یک سیستم لب‌تاپ شخصی معمولی (Intel Core i5 CPU, 4 GB RAM) به‌طور متوسط حدود ۵ برابر بیشتر طول می‌کشد و نتایج برازش مدل هم مشابه است.

مدل رگرسیون فضایی با اثرات فضایی گاوسی را به عنوان مدل رقیب پایه برای مقایسه با مدل رگرسیون فضایی چوله (۶) در نظر گرفته‌ایم. مدل رگرسیون فضایی گاوسی حالت خاصی از مدل چوله است زمانی که پارامتر چولگی λ صفر در نظر گرفته شود. نتایج برآورد پارامترهای دو مدل به روش بیز مقداری در جدول ۳ ارائه شده است که با توجه به فاصله‌های قابل باور بیزی در سطح ۹۵ درصد، پارامترهای β_1 و β_2 معنی‌دار نیستند یعنی طول و عرض جغرافیایی تاثیر معنی‌داری روی بزرگی زمین لرزه‌ها ندارد. در واقع از شمال به جنوب و از شرق به غرب کشور روند معنی‌داری از نظر بزرگی زمین لرزه‌ها مشاهده نمی‌شود. اما عمق لرزه‌ها تاثیر معنی‌داری روی بزرگی لرزه‌ها طی پانزده سال اخیر داشته است چون برآورد فاصله‌ای ۹۵ درصد شامل صفر نیست. با استفاده از روش اعتبار سنجی متقابل دو مدل رگرسیون فضایی چوله و گاوسی مورد مقایسه قرار گرفتند. برای این منظور از معیار ریشه دوم میانگین توان

جدول ۳. نتایج برآورد بیز مقداری پارامترهای مدل رگرسیون فضایی برای داده‌های زمین لرزه‌ای کشور

مدل گاوسی				مدل چوله				پارامتر
چندک ۰/۹۷۵	چندک ۰/۲۵	انحراف استاندارد	برآورد	چندک ۰/۹۷۵	چندک ۰/۲۵	انحراف استاندارد	برآورد	
۵۷۵۲	۳۰۶۳	۰/۴۷۲	۴۵۲۳	۴۳۴۸	۲۸۹۱	۰/۳۷۲	۳۶۱۹	β_0
۰/۱۱۴	۰/۰۶۲	۰/۰۳۱	۰/۰۸۹	۰/۰۹۵	۰/۰۵۵	۰/۰۱۰	۰/۰۷۵	β_1
۰/۱۵۲	-۰/۲۱۱	۰/۰۷۲	-۰/۰۱۶	۰/۰۷۲	-۰/۱۱۹	۰/۰۴۹	-۰/۰۲۳	β_2
۰/۱۹۸	-۰/۱۴۵	۰/۰۹۳	۰/۰۳۹	۰/۱۷۴	-۰/۰۷۶	۰/۰۶۴	۰/۰۴۸	β_3
۲/۰۹۳	۰/۰۱۱	۰/۹۸۴	۰/۵۶۲	۱/۱۴۱	۰/۰۸۱	۰/۵۶۱	۰/۲۳۱	σ
۹/۴۵۳	۰/۰۷۸	۲/۲۱۵	۶/۱۵۱	۶/۸۶۴	۰/۲۴۳	۱/۰۱۷	۳/۸۶۱	φ
-	-	-	-	۴/۴۷۲	۰/۶۸۱	۱/۵۲۳	۲/۱۳۱	λ
۰/۰۸۷۹۴				۰/۰۳۲۴۸				RMSPE

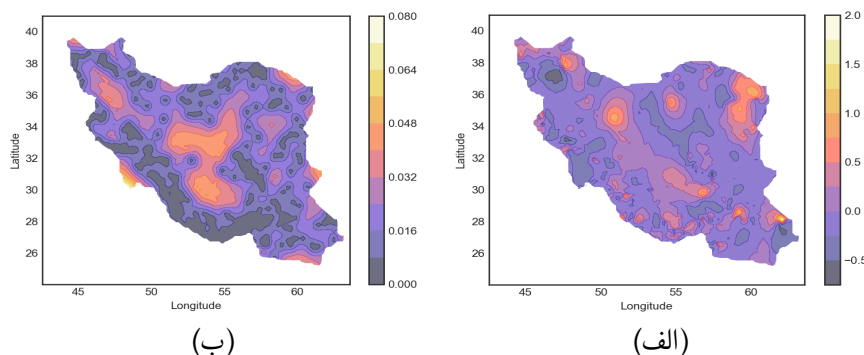
دوم خطای پیشگویی (RMSPE) به صورت

$$RMSPE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_{(i)})^2}$$

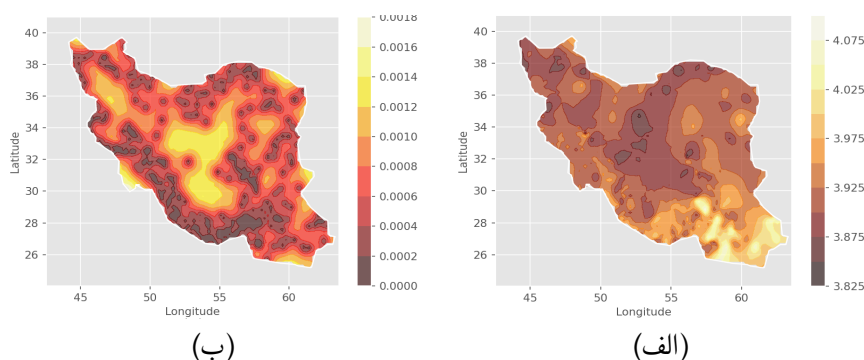
استفاده شد، که در آن مقدار پیشگویی $\hat{y}_{(i)}$ آمین پاسخ بر اساس تمام مشاهدات بدون حضور i آمین مشاهده است. نتایج RMSPE برای دو مدل نیز در جدول ۳ ارائه شده است که نشان می‌دهد مدل رگرسیون فضایی چوله نسبت به مدل گاوسی برای این داده‌ها بهتر است. نقشه پیشگویی فضایی و دقت پیشگویی برای اثرات فضایی مدل رگرسیون فضایی چوله در شکل ۴ رسم شده است که نقشه پیشگویی نواحی همبسته فضایی مشابه و نواحی با اثرات فضایی بیشتر را نشان می‌دهد. با توجه به نقشه پیشگویی شکل ۴ (ب)، نواحی مرکزی کشور دقت پیشگویی فضایی کمتری دارند. همچنین نقشه پیشگویی فضایی برای متغیر پاسخ (بزرگی زمین لرزه‌ها) به همراه دقت آن در شکل ۵ ارائه شده است که نقاط پر خطر از نظر بزرگی زمین لرزه‌ها را روی نقشه ایران نشان می‌دهد. در واقع نقاط هم‌رنگ از نظر میزان بزرگی لرزه مشابه هستند و به نوعی همبستگی مکانی داده‌ها را نشان می‌دهد. همچنین نقاط جنوب و جنوب شرق کشور بیشتر در خطر وقوع زمین لرزه‌های بزرگ هستند. نقشه واریانس پیشگویی شکل ۵ (ب) هم نشان می‌دهد که دقت پیشگویی در نقاط مرکزی و برخی نقاط شمال غربی نسبت به بقیه نقاط کمتر است.

بحث و نتیجه‌گیری

در این مقاله یک مدل رگرسیون چوله براساس زیر رده منعطفی از توزیع چوله نرمال بسته برای مدل‌بندی داده‌های فضایی چوله ارائه شد. با توجه به پیچیدگی تابع درست‌نمایی مدل، رهیافت بیز سلسله مراتبی به کار گرفته شد و برای



شکل ۴. (الف) نقشه پیشگویی اثرات فضایی و (ب) دقت پیشگویی آن‌ها برای مدل رگرسیون چوله



شکل ۵. (الف) نقشه پیشگویی فضایی و (ب) دقت پیشگویی داده‌های زمین لرزه‌های کشور طی پانزده سال اخیر

تقریب توزیع پسین از رهیافت بیز مقداری به‌خاطر سرعت محاسبات بالا به عنوان جایگزینی مناسب برای روش‌های مونت کارلویی استفاده شد. در یک مطالعه شبیه‌سازی مدل رگرسیون فضایی پیشنهادی مورد ارزیابی قرار گرفت و نتایج شبیه‌سازی برای ۱۰۰ مجموعه داده در دامنه‌های فضایی مختلف نشان دادند که مدل پیشنهادی با رهیافت بیز مقداری عملکرد مناسبی داشت و در مقایسه با مدل گاوسی به عنوان یک مدل پایه به‌طور متوسط $RMSE$ برآوردهای آن کمتر بودند. همچنین سرعت محاسبات رهیافت بیز مقداری در مقایسه با الگوریتم‌های MCMC به‌طور متوسط ۳ الی ۴ برابر بیشتر بود (از نرم‌افزار پایتون ۳ روی یک سیستم لب‌تاپ شخصی i5 CPU با 4GB RAM استفاده شده است). در نهایت مدل پیشنهادی روی داده‌های واقعی زمین لرزه‌ای کشور طی ۱۵ سال اخیر اجرا گردید و مدل رگرسیون فضایی گاوسی نیز به‌عنوان یک مدل رقیب به داده‌ها برازش شد. نتایج اعتبارسنجی متقابل نشان داد که مدل پیشنهادی عملکرد بهتری دارد. فاصله‌های قابل باور بیزی ۹۵ درصد برای بررسی تاثیر متغیرهای مدل محاسبه شدند که بیان کردند عمق زمین لرزه‌ها تاثیر معنی‌داری در بزرگی زمین لرزه‌های ثبت شده طی ۱۵ سال اخیر

داشته است. در نهایت نقشه پیشگویی فضایی روی کل ناحیه فضایی رسم شد که نشان داد بخش‌هایی از جنوب و شرق کشور بیشتر از بقیه نواحی در معرض زمین لرزه‌های به بزرگی ۴/۵ و بیشتر هستند. همچنین نقشه‌های دقت پیشگویی نشان دادند که پیشگویی فضایی بخشی‌هایی از نقاط مرکزی و شمال غربی کشور از دقت کمتری برخوردار هستند.

تقدیر و تشکر

نویسندگان مقاله از پیشنهادات و نظرات ارزنده داوران محترم، ویرایش ادبی سردبیر، هیئت تحریریه و ویراستار مجله که در بهبود کیفیت مقاله بسیار مؤثر واقع شد، تقدیر و تشکر می‌نمایند.

مراجع

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

Anselin, L. (1990), Spatial Dependence and Spatial Structural Instability in Applied Regression Analysis, *Journal of Regional Science*, **30**, 185–207.

Basu, S. and Reinsel, G.C. (1994), Regression Models with Spatially Correlated Errors, *Journal of the American Statistical Association*, **89**, 88–99.

Cabral, R., Bolin, D. and Rue, H., (2024), Fitting Latent Non-Gaussian Models Using Variational Bayes and Laplace Approximations, *Journal of the American Statistical Association*, DOI: 10.1080/01621459.2023.2296704.

Cressie, N. (1993), *Statistics for Spatial Data*, Revised Edition, John Wiley, New York.

Gonzalez-Farias, G., Dominguez-Molina, A. and Gupta, A. K. (2004), The Closed Skew Normal Distribution. In: *Genton M. G., ed. Skew-elliptical distributions and their applications: A journey beyond normality. Boca Raton, FL: Chapman and Hall CRC*, 2542.

- Karimi, O. (2023), A Hamiltonian Monte Carlo EM algorithm for Generalized Linear Mixed Models with Spatial Skew Latent Variables, *Statistical Paper*, <https://doi.org/10.1007/s00362-023-01419-y>.
- Karimi, O. and Mohammadzadeh, M. (2012), Bayesian Spatial Regression Models with Closed Skew Normal Correlated Errors and Missing, *Statistical Papers*, **53**(1), 205-218.
- Karimi, O., Omre, H. and Mohammadzadeh, M. (2010), Bayesian Closed-skew Gaussian Inversion of Seismic AVO Data for Elastic Material Properties, *Geophysics*, **75**, R1-R11.
- Márquez-Urbina, O.U. and González-Farías, G. (2022), A Flexible Special Case of the CSN for Spatial Modeling and Prediction, *Spatial Statistics*, **47**, 100556.
- Kim, H.M. and Mallick, B.K. (2004), A Bayesian Prediction using the Skew Gaussian Distribution, *Journal of Statistical Planning and Inference*, **120**, 85–101.
- Lee, J. and Huang, Y. (2022), Covid-19 Impact on US Housing Markets: Evidence from Spatial Regression Models, *Spatial Economic Analysis*, **17**:3, 395-415.
- Oh, M., Shina, D.W. and Kim, H.J. (2002), Bayesian Analysis of Regression Models with Spatially Correlated Errors and Missing Observations, *Computational Statistics and Data Analysis*, **39**, 387–400.
- Ormerod, J. T. and Wand, M. P. (2010), Explaining Variational Approximations, *The American Statistician*, **64**(2), 140–153.
- Tan, L. S. and Nott, D. J. (2013), Variational Inference For Generalized Linear Mixed Models Using Partially Noncentered Parametrizations, *Statistical Science*, **28**, 168–188.
- Zhang, Q., Shan, L., Xie, L., Shaowu G. and Hongye S. (2023), Variational Bayesian State Space Model for Dynamic Process Fault Detection, *Journal of Process Control*, **124**, 129-141.