

تحلیل بیزی مدل فضایی گاوسی با خطای اندازه‌گیری در پیش‌گوها

وحید تدین، عبدالرحمن راسخ

گروه آمار، دانشگاه شهید چمران اهواز

چکیده: عدم قطعیت یکی از ویژگی‌های ذاتی در بسیاری از داده‌های زیستی، زمین‌آماري و جغرافیایی، به عنوان داده‌های فضایی است، که در اغلب اوقات ناشی از وجود خطا در اندازه‌گیری کمیت‌های مورد مطالعه است. این در حالی است که در نظر نگرفتن این موضوع می‌تواند اعتبار نتایج حاصل از تحلیل داده‌ها را زیر سؤال برده و برآوردهای حاصل را دچار تورم واریانس و اریبی قابل توجهی نماید. در این مقاله، به تحلیل مدل فضایی گاوسی با خطای اندازه‌گیری در پیش‌گوها در قالب یک چارچوب بیزی پرداخته خواهد شد. با توجه به پیچیدگی شکل توزیع پسین، نمونه‌گیری از این توزیع به کمک الگوریتم‌های مونت کارلوی زنجیر مارکوفی و روش داده‌افزایی انجام خواهد شد. سرانجام عملکرد مدل پیشنهادی را در مقایسه با نتایج تحلیل مدل ناپخته به کمک شبیه‌سازی ارزیابی می‌شود.

واژه‌های کلیدی: مدل فضایی گاوسی، خطای اندازه‌گیری، تحلیل بیزی.

۱ مقدمه

گسترش روزافزون تحقیقات علمی در علوم مختلف از جمله علوم زیستی و جغرافیایی و همچنین ماهیت داده‌های مورد مطالعه در این شاخه‌ها، محققان این رشته‌ها را بر آن داشته است که مطالعات خود را روی تحلیل داده‌های فضایی متمرکز نمایند. از سوی دیگر اگرچه تا کنون روش‌های آماری گسترده‌ای برای تحلیل چنین داده‌هایی ارائه شده‌اند، با این وجود تعداد مطالعاتی که خطای اندازه‌گیری در پیش‌گوها را لحاظ نموده‌اند، به نسبت کمتر از سایر جنبه‌ها است. از جمله این مطالعات، می‌توان به پژوهش لی و همکاران (۲۰۰۹) اشاره نمود که در آن تحلیل فراوانی‌گرای یک مدل رگرسیون خطی ساده فضایی با اثرات تصادفی و خطای اندازه‌گیری در تنها پیش‌گویی مدل مورد مطالعه قرار گرفته است. اگرچه نتایج حاصل، مؤید عملکرد مناسب مدل پیشنهادی توسط ایشان است، اما همان‌طور که بیان شد با وجود تنها یک متغیر پیش‌گو، محاسبات لازم برای ارائه برآوردها به‌طور چشمگیری پیچیده و در به‌کارگیری الگوریتم EM زمان‌بر خواهد بود (تدین و خالدی، ۲۰۱۵ و تدین، ۲۰۱۷). میلیشنو و همکاران (۲۰۱۳) به ارائه تحلیل کیفیت اندازه‌های حاصل از دستگاه موقعیت‌یاب جهانی (GPS) با استفاده از یک مدل فضایی آمیخته پرداختند. برنیدینالی و همکاران (۱۹۹۷) نحوه به‌کارگیری بیز سلسله‌مراتبی فضایی را برای مدل‌بندی نقشه‌برداری‌های بیماری بر مبنای متغیرهای کمکی زیست‌محیطی نادقیق تشریح نمودند. به‌علاوه، دیردان و همکاران (۲۰۱۱) روی تحلیل داده‌های بیماری‌های واگیر عفونی متمرکز شده و مدلی را به‌کار گرفتند که در آن فاصله اقلیدسی بین مشاهدات به عنوان یک عامل مخاطره لحاظ می‌شود. گریپریس و همکاران (۲۰۰۹) تحلیل خطای اندازه‌گیری ناشی از وجود داده‌های بدتر از در داده‌های فضایی را مورد مطالعه قرار دادند. حمزه‌پور و همکاران (۲۰۱۳) عملکرد مناسب‌تر کریگیدن با خطای اندازه‌گیری در مقایسه با کریگیدن معلولی بر اساس داده‌های میزان شوری خاک را نشان دادند. هیوک و همکاران (۲۰۱۶) استفاده از رویکرد رگرسیون خطی ساده نیمه‌پارامتری را برای تحلیل داده‌های فضایی با خطای اندازه‌گیری در پیش‌گو پیشنهاد نمودند، که در این تحلیل نیز حتی با وجود تنها یک متغیر پیش‌گو، پیچیدگی محاسبات چشمگیر است. از جمله سایر مطالعات در این زمینه می‌توان به هیوک و همکاران (۲۰۱۴)، لیگالو و فینگلنتون (۲۰۱۲)، توماس (۲۰۱۳)، وست و همکاران (۱۹۹۵)، شپارد و همکاران (۲۰۱۲) و جیا و کارلین (۱۹۹۸)، حسینی و همایون‌فال (۱۳۹۵)، اشاره نمود.

در این مقاله به تحلیل بیزی یک مدل فضایی گاوسی بر مبنای خطای اندازه‌گیری در پیش‌گوها پرداخته خواهد شد. از آن‌جا که توزیع پسین پارامترهای مورد نظر شکل بسته‌ای ندارند، نمونه‌گیری از آن‌ها به روش‌های

معمول ممکن نیست، بنابراین با به‌کارگیری الگوریتم متروپولیس-هستینگز درون‌گیز^۱ و استفاده از روش داده‌افزایی^۲، از توزیع پسین توأم پارامترها و متغیرهای پنهان مدل نمونه‌گیری می‌شود. در نهایت عملکرد مدل پیشنهادی به کمک شبیه‌سازی در مقایسه با نتایج حاصل از تحلیل مدل ناپخته^۳ (یعنی مدلی که در آن خطای اندازه‌گیری لحاظ نشده است)، ارزیابی خواهد شد.

در بخش ۲ مدل مورد مطالعه معرفی می‌شود. بخش ۳ به تحلیل بیزی و محاسبه توزیع‌های پسین اختصاص می‌یابد. محاسبه توزیع پیش‌گوی بیزی در بخش ۴ ارائه می‌شود. در بخش ۵ عملکرد مدل پیشنهادی ارزیابی خواهد شد. در انتها بحث و نتیجه‌گیری کلی مقاله ارائه می‌شود. جزئیات مربوط به محاسبات توزیع پسین در پیوست ارائه خواهد شد.

۲ مدل آماری

مشابه با ادبیات مرسوم در آمار فضایی، فرض کنید تغییرات فضایی متغیر مورد مطالعه $Y(\cdot)$ براساس یک میدان تصادفی گاوسی به فرم $\{Y(s); s \in R\}$ مدل‌بندی شده است، که در آن $R \subseteq \mathbb{R}^l$ و $l \geq 1$. لذا می‌توان این مدل را براساس پیش‌گوه‌های $X = (X_1, \dots, X_p)$ و بردار پارامترهای رگرسیونی نامعلوم β ، به صورت

$$Y(s) = X(s)\beta + \sigma\varepsilon(s) + \sigma\rho(s) \quad (1)$$

بیان نمود، که در آن $X(s)\beta$ با $\beta' = (\beta_1, \dots, \beta_p)$ روند موجود در داده‌ها را نمایش می‌دهد، $\varepsilon(\cdot)$ یک میدان مانای مرتبه دوم با میانگین صفر، واریانس یک و تابع همبستگی همسانگرد وابسته به بردار پارامتری θ به صورت $Corr\{\varepsilon(s_i), \varepsilon(s_j)\} = C_\theta(\|s_i - s_j\|) = C_\theta(\|h\|)$ است. در این جا از خانواده انعطاف‌پذیر و همسانگرد مترن به صورت

$$C_\theta(\|h\|) = \frac{1}{\sqrt{\theta_2 - 1} \Gamma(\theta_2)} \left(\frac{\|h\|}{\theta_1} \right)^{\theta_2} K_{\theta_2} \left(\frac{\|h\|}{\theta_1} \right)$$

استفاده می‌شود، که در آن $\theta = (\theta_1, \theta_2)$ ، $\theta_1 > 0$ و $\theta_2 > 0$ (کرسی، ۱۹۹۳). به‌علاوه $\rho(\cdot)$ نیز یک میدان تصادفی گاوسی ناهمبسته با میانگین صفر و واریانس یک است که برای مدل‌بندی اثر قطعه‌ای بکار

¹Metropolis-Hastings-within-Gibbs algorithms

²Data Augmentation

³Naive

می‌رود. لازم به ذکر است که در مدل فوق، میدان‌های تصادفی $\rho(\cdot)$ و $\varepsilon(\cdot)$ مستقل از یکدیگر در نظر گرفته می‌شوند. به‌علاوه پارامترهای σ و σ_ρ مثبت بوده و نسبت $\omega^2 = \frac{\sigma_\rho^2}{\sigma^2}$ بیانگر اهمیت نسبی اثر قطعه‌ای σ_ρ^2 در قیاس با واریانس σ^2 است.

حال مواردی در نظر گرفته می‌شود که در آن‌ها اندازه‌گیری پیش‌گوه‌ای فوق، توأم با خطای اندازه‌گیری بوده و می‌توان نوشت $W(s) = X(s) + U(s)$ ، که در آن $E[U(\cdot)] = 0$. در این صورت، مدل (۱) به‌صورت

$$\begin{aligned} Y(s) &= X(s)\beta + \sigma\varepsilon(s) + \sigma_\rho\rho(s) \\ &= (W(s) - U(s))\beta + \sigma\varepsilon(s) + \sigma_\rho\rho(s) \\ &= W(s)\beta + \sigma\varepsilon(s) + \sigma_\rho\rho(s) - U(s)\beta \end{aligned} \quad (۲)$$

بازنویسی می‌شود. مشکل مهمی که در این مدل خودنمایی کرده و تحلیل آن را تحت تأثیر قرار می‌دهد، وابستگی بین $U(s)$ و $W(s)$ است. برای غلبه بر این مشکل، یک جایگزین مناسب به‌صورت

$$X(\cdot) = E[X(\cdot)|W(\cdot)] + V(\cdot)$$

پیشنهاد می‌شود که با اعمال آن، V و $E(X|W)$ ناهمبسته خواهند بود (کرول و همکاران، ۲۰۰۶)، زیرا با توجه به رابطه $X = W - U$ می‌توان نوشت

$$E(X|W) = E(W - U|W) = W - E(U|W)$$

بنابراین $W = E(X|W) + E(U|W)$ و در نتیجه

$$X = E(X|W) + \underbrace{E(U|W) - U}_V.$$

بدیهی است که $E(V(\cdot)) = 0$. در این پژوهش، $V(\cdot)$ یک میدان تصادفی گاوسی ناهمبسته با واریانس واحد در نظر گرفته می‌شود. بنابراین با بازنویسی مجدد مدل (۲) داریم

$$\begin{aligned} Y(s) &= \underbrace{E[X(s)|W(s)]}_{\mu(s)}\beta + \sigma\varepsilon(s) + \sigma_\rho\rho(s) + \sigma_v V(s)\beta \\ &= \mu(s)\beta + \sigma\varepsilon(s) + \sigma_\rho\rho(s) + \sigma_v V(s)\beta. \end{aligned} \quad (۳)$$

بدین ترتیب در مدل (۳) فرض می‌شود که میدان‌های $\rho(\cdot)$ ، $\varepsilon(\cdot)$ و $V(\cdot)$ مستقل از یکدیگر هستند. حال با فرض این‌که n مشاهده از مدل (۱) در اختیار باشد، با قرار دادن $\tau^2 = \frac{\sigma_\rho^2}{\sigma^2}$ و همچنین در نظر گرفتن بردار

پارامترهای $\eta' = (\beta, \sigma^2, \omega^2, \tau^2, \theta)$ ، توزیع بردار پاسخها مشروط بر فرض ثابت بودن پارامترها به صورت $N_n(\mu(s)\beta, \sigma^2[C_\theta + (\omega^2 + \tau^2)I_n])$ حاصل می‌شود، که در آن I_n یک ماتریس همانی با بعد n است. لازم به ذکر است که در محاسبات آتی به منظور اجتناب از پیچیدگی فرمول‌های ریاضی، از نوشتن عبارت s در کنار نمادهای میدان‌های تصادفی و تغییرات مقیاس بزرگ مدل مورد نظر خودداری می‌شود.

۳ تحلیل بیزی

برای تحلیل بیزی مدل، نخست لازم است توزیع‌های پیشین پارامترهای مدل تعیین شوند. از آن‌جا که هر یک از پارامترها ویژگی خاصی از میدان را کنترل می‌کند، فرض استقلال پیشینی پارامترها از یکدیگر، فرضی دور از انتظار نخواهد بود، لذا می‌پذیریم $\pi(\beta, \sigma^2, \omega^2, \tau^2, \theta) = \pi(\beta)\pi(\sigma^2)\pi(\omega^2)\pi(\tau^2)\pi(\theta)$. با توجه به این‌که اتخاذ پیشین‌های ناسره معمول نظیر پیشین جفریز ممکن است منجر به پسین‌های ناسره شود (برگر و همکاران، ۲۰۰۱)، به منظور اطمینان از سره بودن پسین، پیشین‌های تمام پارامترهای مدل سره در نظر گرفته می‌شود. لازم به ذکر است که مقادیر ابرپارامترها را می‌توان به کمک رهیافت بیزی تجربی تعیین نمود، اما به دلیل تعداد زیاد ابرپارامترها در مدل مفروض، برآورد آن‌ها مشکل‌ساز بوده و لذا مقادیر پایه‌ای برای ابرپارامترها اختیار می‌شود. به‌علاوه با هدف ساده‌سازی محاسبات، سعی می‌شود در صورت امکان از توزیع‌های پیشین مزدوج استفاده شود. لازم به ذکر است که در هر یک از موارد زیر، مقادیر ابرپارامترها به‌گونه‌ای انتخاب می‌شوند که توزیع‌های مبهمی را ارائه نمایند.

- برای پارامتر رگرسیونی β که مؤلفه‌های آن حقیقی‌مقدار هستند، یک پیشین مبهم نرمال به صورت $\beta \sim N_p(\underline{c}, c_1 I_p)$ در نظر گرفته می‌شود که در آن c_1 مقداری دلخواه و به اندازه کافی بزرگ است. پیشین‌های جایگزین پیشنهادی عبارتند از پیشین توانی (ابراهیم و چن، ۲۰۰۰) و پیشین نرمال-گاما (گریفین و برون، ۲۰۱۰).

- اغلب برای پارامتر σ^{-2} ، یک پیشین مزدوج مبهم گاما به فرم $\sigma^{-2} \sim \text{Gamma}(c_2, c_3)$ انتخاب می‌شود که در آن c_2 و c_3 مقادیر کوچک دلخواهی هستند (گلمن، ۲۰۰۶).

- برای ω^2 پیشین گاوسی وارون تعمیم‌یافته به شکل $\omega^2 \sim \text{GIG}(\gamma, c_4, c_5)$ مدنظر قرار می‌گیرد.

توزیع GIG نخستین بار توسط براندوف-نیلسن و همکاران (۱۹۸۲) با چگالی

$$f(\omega^2; \gamma, c_4, c_5) = \frac{\left(\frac{c_5}{c_4}\right)^\gamma}{2\kappa_\gamma(c_4 c_5)} (\omega^2)^{\gamma-1} \exp\left\{-\frac{1}{2}\left\{c_4^2(\omega^2)^{-1} + c_5^2 \omega^2\right\}\right\}$$

معرفی شده است، که در آن $\gamma \in \mathbb{R}$ ، $c_4, c_5 \in \mathbb{R}^+$ و k_γ دلالت بر تابع بسط اصلاح‌شده نوع سوم از مرتبه ω دارد. با توجه به آن که ω^2 یک پارامتر نامنفی بوده و معمولاً شانس بالای مقادیر کوچک این پارامتر مورد انتظار است، توزیع‌های چوله به راست مانند گاما و گامای وارون را نیز می‌توان برای آن در نظر گرفت. اما از آن‌جا که خانواده توزیع‌های گاوسی وارون تعمیم‌یافته بسیار کلی بوده و به‌علاوه در حالت خاص $c_4 = 0$ و $\gamma > 0$ به توزیع گاما و در حالت $c_5 = 0$ و $\gamma < 0$ به توزیع گامای وارون تبدیل می‌شود، استفاده از آن توجیه‌پذیر بوده و لذا ارجحیت دارد. این توزیع همچنین در حالت $\gamma = 0$ نیز از انعطاف‌پذیری بالایی برخوردار است و می‌تواند با تغییر پارامترهای c_4 و c_5 شکل‌های متنوعی از چگالی‌ها را ارائه نماید (یبی و سورنسن، ۲۰۰۳).

• برای پارامتر مثبت τ^2 نیز با استدلالی مشابه با آنچه برای ω^2 ارائه شد، از توزیع GIG به صورت $\tau^2 \sim GIG(0, c_6, c_7)$ استفاده می‌شود.

• پیشین برای پارامترهای تابع همبستگی

در این بخش از پیشین‌های ساده‌نمایی برای پارامترهای تابع همبستگی استفاده می‌شود. ویژگی بارز توزیع نمایی، شانس بالا برای رخداد مقادیر کوچک در قیاس با مقادیر بزرگ است. در ادامه پیشین‌های مستقل معرفی شده، با این حال خواننده به مطالعه پیشین وابسته (اشتاین، ۱۹۹۹) ارجاع داده می‌شود. پارامتر دامنه (θ_1) ، ارتباطی وارونه با فاصله اقلیدسی $(\|h\|)$ دارد، به‌گونه‌ای که با افزایش $\|h\|$ شاهد کاهش وابستگی فضایی خواهیم بود. برای لحاظ کردن این موضوع در پیشین فرض می‌شود $\theta_1 \sim \text{Exp}(\frac{c_8}{\text{med}(d)})$ و $\theta_2 \sim \text{Exp}(c_9)$. نشان دهنده میانه کلیه فواصل بین داده‌هاست.

۱.۳ محاسبه توزیع پسین و نمونه‌گیری از آن

ارائه تحلیل بیزی، منوط به محاسبه توزیع‌های پسین پارامترهای مدل بوده، که خود مستلزم تعیین تابع درست‌نمایی است، زیرا که $\pi(\eta|y) \propto L(\eta|y)\pi(\eta)$. با توجه به پیچیده بودن شکل توزیع پسین، بدیهی است که ارائه تحلیل بیزی به‌طور مستقیم (بر اساس آن) در عمل امکان‌پذیر نیست. بنابراین سعی می‌شود تا به کمک روش‌های مونت کارلوی زنجیر مارکوفی از جمله نمونه‌گیری گیبز و روش داده‌افزایی، از توزیع $P(\eta, \varepsilon, V|y)$ نمونه‌گیری شود، که در آن که $\varepsilon' = (\varepsilon(s_1), \dots, \varepsilon(s_n))$. برای اساس با استفاده از نمونه‌های $\{\eta^{(i)}, \varepsilon^{(i)}, V^{(i)}\}_{i=1}^{\ell}$ که در واقع نمونه‌های حاصل شده پس از دوره داغیدن از توزیع پسین $P(\eta, \varepsilon, V|y)$ هستند، می‌توان پیش‌گویی فضایی بیزی را ارائه نمود.

با توجه به توضیحات فوق، هدف این بخش، ارائه نحوه استفاده از الگوریتم نمونه‌گیری گیبز، به منظور نمونه‌گیری از توزیع پسین $P(\eta, \varepsilon, V | \underline{y}) = P(\varepsilon | \eta, v, \underline{y})P(V | \eta, \underline{y})\pi(\eta | \underline{y})$ خواهد بود. برای این منظور بایستی توزیع‌های پسین شرطی کامل هر یک از پارامترهای مدل و همچنین پارامترهای اضافی (متغیرهای پنهان) تعیین شوند. با توجه به حجم زیاد محاسبات، در این بخش از ارائه آن اجتناب می‌شود. توزیع‌های پسین شرطی عبارتند از:

$$\begin{aligned} \varepsilon | v, \eta, \underline{y} &\sim N_n(A_1^{-1} \mathbf{z}, A_1^{-1}), \\ V \boldsymbol{\beta} | \varepsilon, \eta, \underline{y} &\sim N_n(A_V^{-1} \mathbf{r}, A_V^{-1}), \\ \boldsymbol{\beta} | \varepsilon, v, \eta, \underline{y} &\sim N_p(F, A_V^{-1}), \\ \omega^2 | \varepsilon, v, \eta, \underline{y} &\sim GIG(\gamma - \frac{n}{V}, d^*, c_\delta), \\ \pi(\sigma^2 | \varepsilon, v, \eta, \underline{y}) &\propto (\frac{1}{\sigma^2})^{\frac{n}{2} + c_1 + 1} \exp\{-\frac{1}{2\omega^2} \sum_{i=1}^n (q_i^* - \frac{q_i}{\sigma})^2 - \frac{c_2}{\sigma^2}\}, \\ \pi(\tau^2 | \varepsilon, v, \eta, \underline{y}) &\propto \exp\{-\frac{\sigma^2}{2\tau^2\omega^2} \sum_{i=1}^n (r_i^* - \tau^2 v_i' \boldsymbol{\beta})^2\} \\ &\quad \times \exp\{-\frac{1}{2} (V \boldsymbol{\beta} - A_V^{-1} \frac{\mathbf{r}^*}{\tau})' A_V (V \boldsymbol{\beta} - A_V^{-1} \frac{\mathbf{r}^*}{\tau})\} \\ &\quad \times (\frac{\omega^2}{\tau^2 + \omega^2})^{-n} \frac{1}{\tau^2} \exp\{-\frac{1}{2} [\frac{c_3}{\tau^2} + c_4 \tau^2]\}, \\ \pi(\theta_1 | \varepsilon, v, \eta, \underline{y}) &\propto |C_{(\theta_1, \theta_r)}|^{-\frac{1}{2}} \exp\{-\frac{1}{2} \boldsymbol{\varepsilon}' C_{(\theta_1, \theta_r)}^{-1} \boldsymbol{\varepsilon}\} \\ &\quad \times \exp\{-\frac{c_\lambda}{med(d)} \theta_1\}, \\ \pi(\theta_r | \varepsilon, v, \eta, \underline{y}) &\propto |C_{(\theta_1, \theta_r)}|^{-\frac{1}{2}} \exp\{-\frac{1}{2} \boldsymbol{\varepsilon}' C_{(\theta_1, \theta_r)}^{-1} \boldsymbol{\varepsilon}\} \\ &\quad \times \exp\{-c_9 \theta_r\}, \end{aligned}$$

که در آن

$$\begin{aligned} A_1 &= \frac{1}{\omega^2} I_n + C_\theta^{-1}, & A_V &= (\frac{\tau^2}{\omega^2} + 1) I_n, \\ A_V &= \frac{1}{\sigma^2 \omega^2} \mathbf{T}^* \mathbf{T}^* + \frac{1}{c_1} I_p, & F &= \frac{1}{\sigma^2 \omega^2} A_V^{-1} \mathbf{T}^* \mathbf{t}, \\ \mathbf{z}' &= (z_1, \dots, z_n), & z_i &= \frac{y_i - \mu_i' \boldsymbol{\beta} - \sigma \tau v_i' \boldsymbol{\beta}}{\sigma}, \end{aligned}$$

$$\begin{aligned} \mathbf{r}' &= (r_1, \dots, r_n), & r_i &= \frac{y_i - \mu'_i \beta - \sigma \varepsilon_i}{\sigma \tau}, & r_i^* &= \tau r_i, \\ \mathbf{t}' &= (t_1, \dots, t_n), & t_i &= y_i - \sigma \varepsilon_i, \\ \mathbf{T}^{*'} &= (t_1^{*'}, \dots, t_n^{*'}), & t_i^{*'} &= \mu'_i + \sigma \tau v_i', \\ d^* &= c_\tau^\gamma + \frac{1}{\sigma^\gamma} \sum_{i=1}^n d_i^\gamma, & d_i &= y_i - \mu'_i \beta - \sigma \varepsilon_i - \sigma \tau v_i' \beta, \\ q_i^* &= \varepsilon_i + \tau v_i' \beta, & q_i &= y_i - \mu'_i \beta. \end{aligned}$$

بدیهی است که توزیع‌های شرطی کامل پارامترهای $\theta_1, \tau^\gamma, \sigma^\gamma$ و θ_2 شکل بسته‌ای نداشته و نمونه‌گیری از آن‌ها با تکنیک‌هایی چون الگوریتم نمونه‌گیری متروپولیس-هستینگز یا تکنیک SIR انجام خواهد شد.

۴ پیش‌گویی فضایی بیزی

از دیگر موضوعات مهم در تحلیل داده‌های فضایی، پیش‌گویی مقدار (یا مقادیر) نامعلوم میدان تصادفی مورد مطالعه در یک (یا چند) موقعیت دلخواه (اندازه‌گیری نشده) بر اساس سایر موقعیت‌های مشاهده‌شده است. به منظور ارائه پیش‌گویی با رویکرد بیزی، تعیین توزیع پیش‌گوی بیزی امری ضروری است. مقدار متغیر فضایی مورد مطالعه در موقعیت جدید s_0 را با $y_0 = y(s_0)$ نمایش می‌دهیم. در این صورت رابطه

$$\begin{aligned} P(Y_0 | \underline{y}) &= \int_{\mathbb{R}^{n+1}} \int_{\mathbb{R}^{n+1}} \int_{\Omega} P(Y_0, \varepsilon^*, V^*, \eta | \underline{y}) d\eta d\varepsilon^* dV^* \\ &= \int_{\mathbb{R}^{n+1}} \int_{\mathbb{R}^{n+1}} \int_{\Omega} P(Y_0 | \varepsilon^*, V^*, \eta, \underline{y}) P(\varepsilon_0 | \varepsilon, V^*, \eta, \underline{y}) \\ &\quad \times P(V_0 | \varepsilon, V, \eta, \underline{y}) P(\varepsilon, V, \eta | \underline{y}) d\eta d\varepsilon^* dV^*, \end{aligned}$$

برقرار است، که در آن فضای پارامتری η بوده، $\varepsilon^* = (\varepsilon_0, \varepsilon')'$ و $V^* = (V_0, V)$. واضح است که با در اختیار داشتن نمونه‌های حاصل از توزیع $P(\varepsilon, V, \eta | \underline{y})$ ، می‌توان از توزیع $P(V_0 | \varepsilon, V, \eta, \underline{y})$ نمونه‌گیری نمود. به‌طور مشابه با داشتن نمونه‌های جدید حاصل شده، امکان تولید نمونه تصادفی از توزیع $P(\varepsilon_0 | \varepsilon, V^*, \eta, \underline{y})$ فراهم می‌شود. با تکرار این روند، از توزیع $P(Y_0 | \varepsilon^*, V^*, \eta, \underline{y})$ نمونه‌گیری می‌شود و بدین ترتیب نمونه‌هایی از توزیع پیش‌گو تولید می‌شود. بدین ترتیب رابطه

$$(Y_0, \underline{Y}')' | \eta \sim N_n \left(\begin{bmatrix} \mu(s_0) \\ \mu(s) \end{bmatrix}, \sigma^2 [C_\theta^* + (\omega^\gamma + \tau^\gamma) I_{n+1}] \right),$$

حاصل می‌شود، که در آن $C_{\theta}^* = \begin{pmatrix} 1 & r'_{\theta} \\ r_{\theta} & C_{\theta} \end{pmatrix}$ و برای $i = 1, \dots, n$ داریم $r'_{\theta_i} = C_{\theta}(\|s_0 - s_i\|)$ با توجه به ویژگی توزیع نرمال چندمتغیره

نتیجه می‌شود که $Y_0 | \varepsilon^*, V^*, \eta, \underline{y}$ دارای توزیع نرمال، به ترتیب، با میانگین و واریانس

$$E(Y_0 | \varepsilon^*, V^*, \eta, \underline{y}) = \mu'(s_0)\beta + \sigma\varepsilon(s_0) + \sigma\tau v'(s_0)\beta \\ + r'_{\theta} C_{\theta}^{-1}(\underline{y}(s) - \mu(s)\beta - \sigma\varepsilon(s) - \sigma\tau V(s)\beta)$$

$$Cov(Y_0 | \varepsilon^*, V^*, \eta, \underline{y}) = \sigma^2(\omega^2 + \tau^2) - r'_{\theta} C_{\theta}^{-1} r_{\theta}$$

است. به علاوه $\varepsilon_0 | \varepsilon, V, \eta, \underline{y} \sim N(0, 1)$ و $\varepsilon_0 | \varepsilon, V^*, \eta, \underline{y} \sim N(r'_{\theta} C_{\theta}^{-1} \varepsilon, 1 - r'_{\theta} C_{\theta}^{-1} r_{\theta})$. بدین ترتیب با در اختیار داشتن نمونه‌های تولیدشده به شکل $\{y_{0_i}\}_{i=1}^m$ پیش‌گویی فضایی بیزی و واریانس پیش‌گویی متناظر به صورت $\widehat{Y}_0 = \frac{1}{m} \sum_{i=1}^m y_{0_i}$ و $S_{Y_0 | \underline{y}}^2 = \frac{1}{m} \sum_{i=1}^m (y_{0_i} - \widehat{y}_0)^2$ ارائه می‌شوند.

۵ مطالعه شبیه‌سازی

در بخش‌های قبل به معرفی مدل فضایی گاوسی با خطای اندازه‌گیری در پیش‌گوها و تحلیل آن با رویکرد بیزی پرداخته شد. در این بخش تلاش می‌شود که عملکرد مدل مذکور را در مقایسه با نتایج حاصل از تحلیل مدل ناپخته مورد مطالعه قرار گیرد. در حالت کلی، اهداف متفاوتی از ارائه هر شبیه‌سازی می‌توان دنبال نمود، که در ادامه به چند مورد اشاره نموده و در مقایسه‌ها دنبال می‌کنیم. قبل از ارائه اهداف کلی مورد نظر از انجام شبیه‌سازی، تذکر این نکته ضروری به نظر می‌رسد که در این شبیه‌سازی اساس کار بر مقایسه دو مدل در برآورد پارامترها و پیش‌گویی‌ها متمرکز شده و لذا بحث آتی به سمت بررسی عواقب نادیده گرفتن عامل خطای اندازه‌گیری (از جمله اریبی برآوردها و تورم واریانس) هدایت نخواهد شد.

• ارزیابی برآورد

در این شبیه‌سازی برآورد پارامترها تحت هر یک از مدل‌های ناپخته و خطای اندازه‌گیری محاسبه و با مقادیر واقعی مقایسه خواهند شد. معیار مقایسه را (با توجه به تک بعدی بودن پارامترها در مدلی که در ادامه معرفی و از آن شبیه‌سازی می‌شود) می‌توان به صورت قدر مطلق تفاضل مقادیر برآورد شده و مقادیر واقعی در نظر گرفت.

• ارزیابی پیش‌گویی

ارزیابی پیش‌گویی در این مطالعه به روش اعتبارسنجی متقابل انجام می‌شود که در آن بجای این‌که در n موقعیت شبیه‌سازی صورت گیرد، در $n + m$ موقعیت از میدان تصادفی مورد نظر شبیه‌سازی شده و با کنار گذاشتن m موقعیت اضافی، هر یک از مدل‌ها برازش داده می‌شوند. سرانجام مقادیر میدان تصادفی در این m موقعیت پیش‌گویی شده و با مقادیر واقعی مقایسه می‌شوند. در این حالت، بر اساس معیار کمترین توان‌دوم‌های خطا، مقایسه دو مدل صورت می‌گیرد.

• تحلیل حساسیت مدل

با توجه به این‌که تحلیل در چارچوب بیزی انجام می‌شود این سؤال مطرح می‌شود که نتایج تحلیل تا چه حد وابسته (یا حساس) نسبت به اعمال تغییرات در مقادیر ابرپارامترهاست؟ راهکار پیشنهادی آن است که ابتدا با جایگزینی مقادیر مورد نظر، نمونه‌گیری‌های مجدد از توزیع پسین مربوطه انجام می‌شود. در ادامه، حاصل تقسیم میانگین توزیع پسین، بر انحراف استاندارد محاسبه‌شده بر اساس پیشین مبنای، را معیار قرار داده و آن را معیار "تغییرات نسبی" می‌نامیم. چنانچه پارامتر مربوطه به شکل برداری در نظر گرفته و یا مقادیر جایگزین متفاوتی برای پارامتری خاص لحاظ شود، پس از محاسبه معیار مذکور برای تک تک درایه‌های آن، بیشینه تغییرات نسبی را به عنوان تغییرات ناشی از تعویض مقدار ابرپارامتر در نظر می‌گیریم. جنبه دیگری که در ارزیابی حساسیت می‌توان مطرح نمود، حساسیت تحلیل‌ها نسبت به تغییرات در مقادیر پایه‌ای است که برای خود پارامترها به منظور اجرای الگوریتم در نظر گرفته می‌شود، که این ارزیابی را نیز می‌توان به طریق مشابه با آن‌چه که در بالا تشریح شد، مورد مطالعه قرار داد.

• ارزیابی شناسایی‌پذیر بودن پارامترهای مدل

بررسی شناسایی‌پذیر بودن پارامترهای بکار رفته در مدل از دیگر اهداف این شبیه‌سازی است. به بیان دیگر، از آن‌جا که تعداد پارامترهای بکار رفته در مدل به نسبت زیاد بوده و به‌علاوه تعدادی از این پارامترها به‌طور توأم ویژگی خاصی از مدل را (مانند واریانس یا وابستگی) کنترل می‌کنند، بدین ترتیب رفتار ویژگی مذکور را می‌توان مؤثر از تغییرات در هر یک از پارامترهای مذکور (به تفکیک یا توأم) دانست. با این توضیح، به نحوی می‌توان یک افراز روی پارامترهای مدل تصور نمود. حال طرح این سؤال دور از ذهن نخواهد بود که مدل مورد مطالعه به چه میزان قادر است بین مؤلفه‌های این دسته از افراز تمایز قائل شده یا به زبان ساده‌تر آن‌ها را از یکدیگر تمیز دهد؟ شیوه اجرای این ارزیابی بدین ترتیب پیشنهاد می‌شود که ابتدا آن دسته از پارامترهایی را که انتظار می‌رود مدل در شناسایی آن‌ها با

مشکل مواجه شود را تعیین نموده و سپس مقادیر متفاوتی برای هر یک از آن‌ها به تفکیک در نظر می‌گیریم (بدون اعمال هیچ‌گونه تغییر اضافی در سایر مقادیر اولیه و پارامترها). در ادامه با توجه به تعداد مقادیر اختیاریافته، از مدل مورد نظر به همان تعداد دسته‌های نمونه‌ای متفاوتی تولید شده و روند برآورد پارامترها را به شیوه سابق و به تفکیک برای هر دسته از نمونه‌ها اجرا می‌شود و نهایتاً برآورد حاصل شده هر پارامتر را با مقدار واقعی لحاظ شده مقایسه خواهند شد. در این‌جا نیز معیار مقایسه را می‌توان قدرمطلق تفاضل مقادیر واقعی و برآورد شده در نظر گرفت.

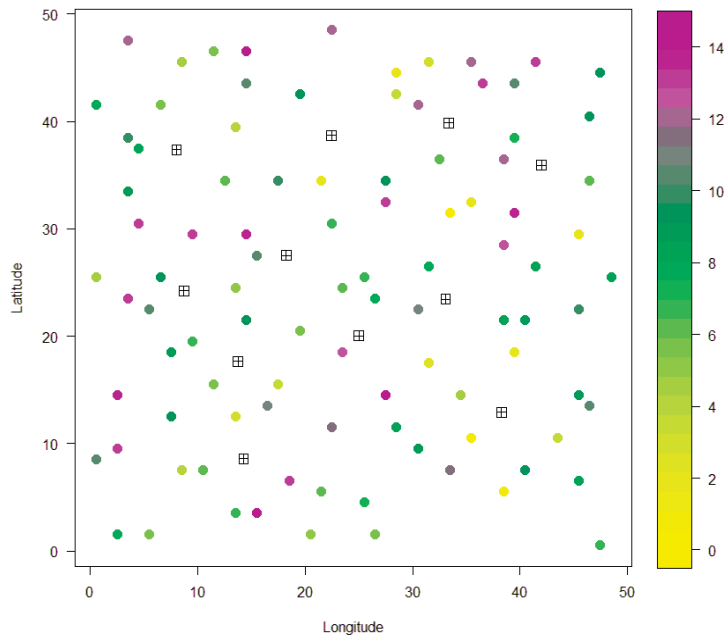
به منظور ارائه شبیه‌سازی، نیازمند تعیین مختصات موقعیت‌های نمونه‌ای خواهیم بود. در این مثال، از مختصات داده‌های 97data.dat موجود در نرم‌افزار GSLIB (دویچ و جرنل، ۱۹۹۲) که حاکی از ۹۷ موقعیت جغرافیایی در ناحیه‌ای $5^\circ \times 5^\circ$ مایلی است، استفاده خواهیم نمود. در ادامه با به‌کارگیری بسته geoR در محیط نرم‌افزار R، از یک میدان تصادفی گاوسی با ساختار وابستگی نمایی به عنوان حالتی خاص از مدل مترن (به ازای $\theta_2 = 0.5$)، در موقعیت‌های تعیین شده شبیه‌سازی می‌شود. در ادامه به شرح جزئیات شبیه‌سازی خواهیم پرداخت. لازم به ذکر است که در اکثر محاسبات آتی، مقادیر حاصل ابتدا تا چهار رقم اعشار گرد شده و سپس تحت عنوان نتایج حاصل از تحلیل در قالب جداولی منفک ارائه خواهند شد. ابتدا با رسم مختصات ۹۷ موقعیت مورد مطالعه در یک صفحه دکارتی و تعیین یازده موقعیت جدید مابین مکان‌های قبلی، یک نمونه تصادفی 108 تایی از یک میدان تصادفی گاوسی با میانگین ثابت $2/5$ ، واریانس واحد و تابع همبستگی نمایی با پارامتر همبستگی یک، تولید و مقادیر حاصل را درون یک بردار به نام x ذخیره نموده و از این به بعد به آن‌ها به چشم یک مقدار ثابت (و نه تصادفی) خواهیم نگریست. جدول ۱ مختصات این یازده موقعیت را نشان می‌دهد. سپس مقادیر متغیر مورد مطالعه را با ساختار وابستگی فضایی و با تابع همبستگی نمایی با پارامتر $\theta = 1/25$ از مدل

$$Y(s_i) = \beta_0 + \beta_1 x(s_i) + \sigma \varepsilon(s_i) + \sigma \omega \rho(s_i) \quad (4)$$

بر اساس مقادیر $\beta_0 = 0.75$ ، $\beta_1 = 1/6$ ، $\sigma^2 = 0.9$ و $\omega^2 = 1/15$ تولید می‌شود. شکل ۱ مکان قرارگیری موقعیت‌های نمونه‌ای و بزرگی مقادیر تولیدشده را بر اساس طیف‌های رنگی نمایش می‌دهد. در این شکل، موقعیت‌های یازده موقعیت اضافه‌شده با نماد \boxplus نمایش داده شده‌اند. در ادامه به منظور لحاظ کردن خطای اندازه‌گیری در متغیر پیش‌گوی x ، با کنار گذاشتن یازده موقعیت مذکور، و تولید یک نمونه تصادفی

جدول ۱: مختصات یازده موقعیت افزوده‌شده با هدف انجام پیشگویی.

| ردیف | طول جغرافیایی | عرض جغرافیایی | ردیف | طول جغرافیایی | عرض جغرافیایی |
|------|---------------|---------------|------|---------------|---------------|
| ۱ | ۱۴/۱۴۳۵۷۸ | ۸/۴۴۹۵۲۸ | ۷ | ۲۲/۶۷۷۱۸۸ | ۳۸/۸۱۵۲۸۶ |
| ۲ | ۱۳/۶۱۰۷۹۱ | ۱۷/۷۸۲۷۲۶ | ۸ | ۳۳/۷۸۳۷۵۰ | ۳۹/۸۶۶۹۱۳ |
| ۳ | ۹/۰۰۴۲۳۱ | ۲۴/۲۲۳۹۴۸ | ۹ | ۳۳/۱۵۱۷۸۷ | ۲۳/۰۵۴۷۶۲ |
| ۴ | ۸/۵۰۷۵۰۹ | ۳۷/۳۶۹۲۹۷ | ۱۰ | ۴۳/۵۳۵۳۹۰ | ۳۶/۴۳۵۲۲۷ |
| ۵ | ۱۸/۰۳۴۵۶۳ | ۲۷/۳۷۸۸۳۲ | ۱۱ | ۳۸/۹۲۳۰۹۷ | ۱۳/۰۵۰۴۰۱ |
| ۶ | ۲۴/۸۲۹۶۴۸ | ۲۰/۱۴۸۸۸۹ | | | |



شکل ۱: نمایش موقعیت‌های نمونه‌ای و مکان‌های اضافه‌شده (⊠).

۹۷ تایی از توزیع نرمال $N(0, 1)$ به نام v و قرار دادن v در $\mu = x_{97 \times 1} - v$ ، مدل (۴) به شکل

$$Y(s_i) = \beta_0 + \beta_1 \mu(s_i) + \sigma \varepsilon(s_i) + \sigma \omega \rho(s_i) + \beta_1 \sigma \tau v(s_i) \quad (5)$$

بازنویسی می‌شود، که در آن $\tau^2 = 0.15$ در نظر گرفته شده است. لذا ادامه کار، متشکل از مقایسه نتایج تحلیل مدل خطای اندازه‌گیری (۵) و مدل ناپخته

$$Y(s_i) = \beta_0 + \beta_1 \mu(s_i) + \sigma \varepsilon(s_i) + \sigma \omega \rho(s_i) \quad (6)$$

خواهد بود. لازم به تأکید است در ارائه تحلیل‌ها بر اساس مدل (۵)، فرض می‌شود که هیچ اطلاعی از مقادیر مؤلفه‌های بردارهای x و v در اختیار نبوده و تنها بردار μ را مشاهده شده است. در این بخش، توزیع‌های پیشین پیشنهادی را به صورت $(\beta \sim N_2(0, 12I_2), \sigma^{-2} \sim Inv-Gamma(1.25, 0.09))$ ، $\omega^2 \sim GIG(0.05, 0.03, 2)$ ، $\tau^2 \sim GIG(0, 0.09, 2)$ و $\theta \sim Exp(1)$ در نظر گرفته می‌شود. مقادیر ابرپارامترها به گونه‌ای اختیار شده‌اند که میانگین مقادیر توزیع‌های متناظر، چندان از مقادیر واقعی پارامترها فاصله نداشته باشند (در اکثر موارد، حداکثر این اختلاف تقریباً سه برابر انحراف معیار است). به منظور ارائه تحلیل مدل (۶)، نیازمند تعیین توزیع‌های شرطی کامل پارامترها و متغیرهای افزوده شده تحت این مدل خواهیم بود. از آن‌جا که محاسبات لازم، مشابه محاسبات ارائه شده در پیوست خواهد بود، از بیان آن اجتناب می‌شود.

با توجه به مطالب شرح داده شده، به تحلیل داده‌های تولید شده توسط دو مدل (۵) و (۶) پرداخته خواهد شد. اجرای کدهای دستوری، نیازمند تعیین مقادیر پایه‌ای برای هر یک از پارامترها و متغیرهای پنهان افزوده شده است. به منظور مقایسه دقیق عملکرد دو مدل با یکدیگر، این مقادیر پایه‌ای یکسان اختیار می‌شوند. مقادیر پایه‌ای اختیار شده برای پارامترها به صورت $\beta_0 = 1.25$ ، $\beta_1 = 3$ ، $\sigma^2 = 2.5$ ، $\omega^2 = 2.75$ ، $\tau^2 = 0.3$ و $\theta = 2/6$ بوده و مقادیر پایه‌ای برای متغیرهای پنهان افزوده شده، یک نمونه تصادفی ۹۷ تایی از توزیع‌های نرمال به صورت $\varepsilon_i \sim N(0, 0.33)$ و $v_i \sim N(0, 0.3)$ برای $i = 1, 2, \dots, 97$ در نظر گرفته شده است، با این توضیح که مقادیر پایه‌ای v_i ها و τ صرفاً برای تحلیل داده‌ها تحت مدل (۵) مورد نیاز خواهند بود. لازم به ذکر است که به منظور اطمینان از حساس نبودن تحلیل‌ها به مقادیر پایه‌ای، با اعمال تغییراتی جزئی در مقادیر پایه‌ای پارامترها و همچنین ابرپارامترها، الگوریتم را مجدداً اجرا نموده و تغییرات حاصل در نتایج را مقایسه نمودیم، که نتایج به تفکیک در ادامه بیان می‌شوند. با تکرار نمونه‌گیری و بررسی اجمالی در خصوص همگرایی نتایج، دوره داغیدن برابر $10,000$ در نظر گرفته شد و نتایج با میانگین‌گیری از 1000 مشاهده (هر پنجمین مشاهده پس از دوره داغیدن، با هدف کاهش وابستگی موجود بین نمونه‌های حاصل از الگوریتم MCMC)، بیان خواهند شد.

جدول ۲ خطا و واریانس برآورد هر یک از پارامترها را تحت دو مدل مورد مطالعه نمایش می‌دهد. با استفاده از معیار قدم‌مطلق تفاضل مقادیر برآورد شده و محاسبه شده، عملکرد بهتر مدل خطای اندازه‌گیری مورد تأیید قرار می‌گیرد. در تفسیر این جدول همچنین می‌توان اظهار داشت، با توجه به این‌که مدل (۶) قسمتی از پراکندگی موجود در داده‌ها را به مؤلفه σ^2 و مابقی را به مؤلفه ω^2 تخصیص داده و در مقابل مدل (۵) این پراکندگی را علاوه بر این دو پارامتر، در قالب پارامتر سوم τ^2 نیز تفسیر می‌کند، بنابراین، رویداد این پدیده که برآورد σ^2 یا ω^2 (یا هر دو) تحت مدل (۶) توأم با مسئله بیش‌برآوردی باشد (که افزایش خطای برآورد را در پی خواهد داشت)، چندان دور از انتظار نبوده و نتایج نیز مؤید این موضوع است.

جدول ۲: خطا و واریانس برآورد پارامترها تحت دو مدل خطای اندازه‌گیری و ناپخته.

| پارامتر | مقدار واقعی | مدل ناپخته | | مدل خطای اندازه‌گیری | |
|------------|-------------|------------|---------|----------------------|---------|
| | | خطا | واریانس | خطا | واریانس |
| β_0 | ۰/۷۵ | ۰/۷ | ۱/۱۵ | ۰/۱۴ | ۰/۴ |
| β_1 | ۱/۶ | ۰/۶۱ | ۰/۹۴ | ۰/۱ | ۰/۳۷ |
| σ^2 | ۰/۹ | ۱/۸۶ | ۱/۹۷ | ۰/۴۶ | ۰/۷۱ |
| ω^2 | ۱/۱۵ | ۰/۹۹ | ۱/۶۴ | ۰/۳۴ | ۰/۶۵ |
| θ | ۱/۲۵ | ۰/۸۴ | ۱/۵۴ | ۰/۳۸ | ۰/۵۲ |
| τ^2 | ۰/۱۵ | - | - | ۰/۳۵ | ۰/۳۹ |

جدول ۳ نتایج بررسی حساسیت مدل‌های مورد مطالعه نسبت به اعمال تغییرات جزئی در مقادیر اولیه لازم برای اجرای الگوریتم را نمایش می‌دهد. این نتایج علاوه بر تأیید حساسیت کمتر مدل خطای اندازه‌گیری، نشان دهنده ضعف نسبتاً بارز عملکرد مدل ناپخته (در این موضوع مورد بررسی) به‌ویژه نسبت به تغییرات در مقادیر پایه‌ای پارامتر σ^2 است. با دقت نظر در مقادیر ستون آخر این جدول، ملاحظه می‌شود که بیشینه تغییرات نسبی محاسبه شده برای پارامتر τ^2 در مقایسه با سایر پارامترها از اندازه بزرگ‌تری برخوردار بوده، که البته این موضوع از ابتدا نیز دور از انتظار نبود، زیرا که τ^2 در واقع پارامتر متغیر پنهان V بوده و بدیهی است که مدل در تفسیر رفتار این پارامتر نسبت به تفسیر رفتارهای سایر پارامترها ناتوان‌تر خواهد بود.

جدول ۳: تحلیل حساسیت مدل به ازای تغییرات اعمال شده در مقادیر مبنا برای اجرای الگوریتم.

| بیشینه تغییرات نسبی | | | | | |
|---------------------|-------------|------------|----------------|------------|----------------------|
| پارامتر | مقدار واقعی | مقدار مبنا | مقادیر جایگزین | مدل ناپخته | مدل خطای اندازه‌گیری |
| β_0 | ۰/۷۵ | ۱/۲۵ | ۰/۱ ، ۱/۵ | ۰/۲۱ | ۰/۱۷ |
| β_1 | ۱/۶ | ۳ | ۱/۴ ، ۴ | ۰/۱۶ | ۰/۰۱۴ |
| σ^2 | ۰/۹ | ۲/۵ | ۰/۳۵ ، ۱/۴۵ | ۰/۳۹ | ۰/۱۹ |
| ω^2 | ۱/۱۵ | ۲/۷۵ | ۰/۷۶ ، ۲/۱ | ۰/۳۴ | ۰/۲ |
| θ | ۱/۲۵ | ۲/۶ | ۱/۱ ، ۳/۵ | ۰/۳ | ۰/۱۸ |
| τ^2 | ۰/۱۵ | ۰/۳ | ۰/۰۵ ، ۰/۴ | - | ۰/۲۸ |

جدول ۴ نیز نتایج تحلیل حساسیت مدل‌های ناپخته و خطای اندازه‌گیری را نسبت به اعمال تغییرات در مقادیر پایه‌ای ابرپارامترهای مدل ارائه می‌کند. با مقایسه مقادیر تغییرات نسبی دو مدل مشاهده می‌شود که در غالب موارد مدل شامل خطای اندازه‌گیری حساسیت کمتری از خود نشان داده و لذا نتایج حاصل از تحلیل داده‌ها بر اساس این مدل از ثبات و پایداری بیشتری برخوردارند.

جدول ۵ نتایج پیش‌گویی بیزی برای مقادیر میدان تصادفی مورد مطالعه در یازده موقعیت افزوده شده (جدول ۱) را به نمایش می‌گذارد. به منظور مقایسه عملکرد دو مدل مذکور، میانگین توان دوم خطای پیش‌گویی (MSPE) این موقعیت‌ها ارائه شده است، که دقت چشمگیر مدل خطای اندازه‌گیری را در مقایسه با مدل ناپخته تأیید می‌کند.

در این قسمت به بررسی شناسایی‌پذیری بودن پارامترهای مدل خواهیم پرداخت. با توجه به توضیحات ارائه شده در قسمت اهداف شبیه‌سازی، لزوم ارزیابی شناسایی‌پذیری بودن پارامترهای کنترل‌کننده پراکنندگی داده‌ها یعنی $(\sigma^2, \omega^2, \tau^2)$ در مدل (۵) آشکار می‌شود، زیرا که هر یک از پارامترها به تفکیک و همچنین هر سه به‌طور توأم کنترل‌کننده واریانس متغیر مورد مطالعه خواهند بود، و احتمال آن می‌رود که مدل مذکور قادر به تمیز آن‌ها از یکدیگر نباشد. بر این اساس مقادیر متفاوتی برای هر یک از پارامترها اختیار شده و پس از برازش مدل، مجدداً مقدار خطای برآورد در جدول ۶ گزارش شده است. در این‌جا تذکر یک نکته خارج از

جدول ۴: تحلیل حساسیت مدل نسبت به تغییرات اعمال شده در مقادیر پایه‌ای ابرپارامترها.

| تغییرات نسبی در مدل | | | | | |
|---------------------|----------------------|------------------|--------------------------------|--------|------------------|
| پارامتر | ابریارامتر | مقدار مبنا | مقادیر جایگزین | ناپخته | خطای اندازه‌گیری |
| β_0 | c_1 | ۱۲ | ۱۰/۷۵ | ۰/۱۹ | ۰/۰۷ |
| β_1 | c_1 | ۱۲ | ۱۱/۶ | ۰/۲۶ | ۰/۰۸ |
| σ^2 | (c_2, c_3) | (۱/۲۵, ۰/۰۹) | (۰/۲, ۰/۷) | ۰/۸۷ | ۰/۲۲ |
| ω^2 | (γ, c_4, c_5) | (۰/۰۰۵, ۰/۰۳, ۲) | (۱۰ ^{-۳} , ۰/۰۱, ۱/۵) | ۰/۴ | ۰/۲۱ |
| θ | c_8 | ۱/۱ | ۰/۹۸ | ۰/۳ | ۰/۱۸ |
| τ^2 | (c_6, c_7) | (۰/۰۹, ۲) | (۰/۰۵, ۱/۴) | - | ۰/۳۲ |

جدول ۵: میانگین توان دوم خطای پیش‌گویی بر اساس موقعیت‌های جدول ۱.

| ملاک ارزیابی | دل ناپخته | مدل خطای اندازه‌گیری |
|--------------|-----------|----------------------|
| MSPE | ۱/۹۲ | ۰/۶۷ |

بحث ضروری به نظر می‌رسد و آن این‌که در ستون چهارم جدول ۶، مقادیر واقعی اختیار شده همان مقادیری است که در جدول ۲ به منظور ارزیابی مدل‌های مورد مطالعه در برآورد پارامترها اتخاذ شده بود و لذا از نتایج ستون مذکور می‌توان به عنوان تکرار یک شبیه‌سازی مجدد با هدف بررسی دقیق‌تر عملکرد مدل (۵) در برآورد این پارامترها استفاده نمود، که این نتایج نیز عملکرد مناسب مدل را نشان می‌دهد. با بازگشت به موضوع اصلی مورد بحث و تحلیل جدول ۶، می‌توان اظهار داشت که مدل مورد مطالعه ضعف چشمگیری در خصوص شناسایی پارامترها از خود نشان نمی‌دهد.

جدول ۶: ارزیابی شناسایی‌پذیر بودن پارامترهای کنترل‌کننده واریانس در مدل خطای اندازه‌گیری.

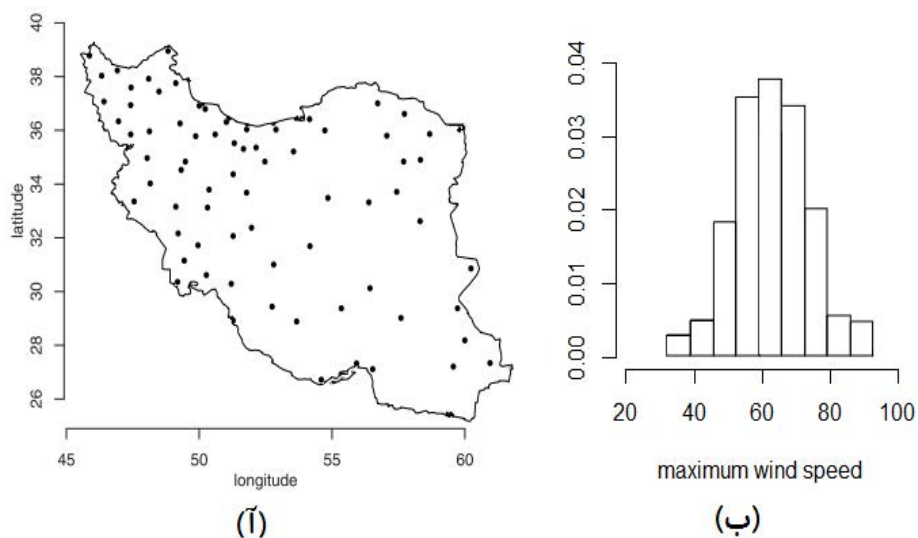
| τ^2 | | ω^2 | | σ^2 | |
|-------------|-------------|-------------|-------------|-------------|-------------|
| مقدار واقعی | خطای برآورد | مقدار واقعی | خطای برآورد | مقدار واقعی | خطای برآورد |
| ۰/۴۱ | ۰/۰۸ | ۰/۴ | ۰/۷۴ | ۰/۳۵ | ۰/۸۵ |
| ۰/۳ | ۰/۱۵ | ۰/۳۳ | ۱/۱۵ | ۰/۴۷ | ۰/۹ |
| ۰/۳۹ | ۰/۱۹ | ۰/۴۴ | ۱/۲۸ | ۰/۴۹ | ۱/۳۳ |

۶ داده‌های سرعت باد

در این بخش تحلیل داده‌های بیشینه سرعت باد^۴ در ۷۲ موقعیت جغرافیایی در سرتاسر ایران به عنوان کاربردی از مدل پیشنهادی ارائه می‌شود. این داده‌ها علاوه بر مقادیر بیشینه سرعت باد (برحسب کیلومتر بر ساعت)، به عنوان متغیر پاسخ، شامل میانگین دمای اندازه‌گیری شده (در طول مدت زمان ثبت شده برای بیشینه سرعت باد و بر حسب سانتیگراد)، به عنوان متغیر پیشگو و همچنین طول و عرض جغرافیایی موقعیت ایستگاه‌های هواشناسی نیز است. شکل ۲ یک نمای کلی از موقعیت جغرافیایی ایستگاه‌ها و همچنین بافت‌نگار داده‌ها را نمایش می‌دهد که در آن طول و عرض جغرافیایی نمایش داده‌شده (برحسب درجه)، تبدیل یافته مختصات ارائه‌شده برحسب UTM بوده که در مقیاس 10^{-5} گزارش شده‌اند.

بدیهی است که متغیر پیشگو در این مطالعه (میانگین دما) با خطا در اندازه‌گیری توأم بوده و به‌طور دقیق مشاهده نمی‌شود و از آن‌جا که ماهیت متغیر پاسخ (سرعت باد) نیز فضایی است، تحلیل این داده‌ها به عنوان کاربردی از هدف این مطالعه مورد توجه قرار می‌گیرد. به‌منظور اطمینان از صحت فرض گاوسی بودن داده‌ها، ابتدا در قالب تحلیل اکتشافی، مقدار آماره آزمون شاپیرو-ویلک محاسبه (جدول ۷) و از صحت نرمال بودن داده‌ها (در سطح ۰/۰۵) اطمینان حاصل شد. به‌علاوه، نمودار Q-Q داده‌ها نیز در شکل ۳ نمایش داده شده است. در گام بعد، با هدف برازش یک مدل تغییرنگار تجربی، از بسته automap در محیط نرم‌افزار R استفاده نموده و یک مدل کروی با اثر قطعه‌ای ۱۲، دامنه ۱/۸ و مقدار آستانه ۴۳ انتخاب می‌شود. در نهایت مدل

^۴ این داده‌ها مربوط به بازه ۱۵ آبان تا ۱۵ بهمن سال ۱۳۷۹ بوده و از طریق درگاه سازمان هواشناسی کشور به نشانی



شکل ۲: (آ) موقعیت جغرافیایی ایستگاه‌ها، (ب) بافت‌نگار داده‌های بیشینه سرعت باد

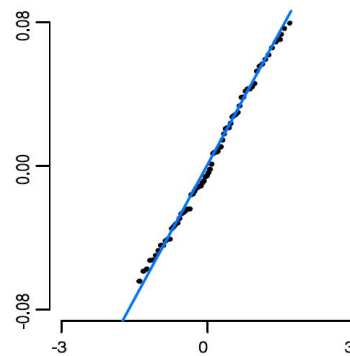
پیشنهادی به داده‌ها برازش داده شد که نتایج در جدول ۸ ارائه شده است. شکل ۴ نیز نقشه پیشگویی بیزی را تحت مدل خطای اندازه‌گیری نمایش می‌دهد.

بحث و نتیجه‌گیری

در این مقاله تحلیل بیزی یک مدل فضایی گاوسی با خطای اندازه‌گیری در پیش‌گوها مورد بررسی و مطالعه قرار گرفت. به‌منظور ارائه برآوردها، الگوریتم متروپولیس-هستینگز درون‌گنبد مورد استفاده قرار گرفت و همچنین عملکرد رویکرد پیشنهادی در مقایسه با مدل ناپخته به کمک شبیه‌سازی بررسی شد. در این شبیه‌سازی اهداف مختلفی دنبال شد که از جمله آن‌ها می‌توان به تحلیل حساسیت و شناسایی‌پذیری پارامترها اشاره نمود. نتایج

جدول ۷: آزمون نرمال بودن شاپیرو-ویلک برای داده‌های سرعت باد.

| Shapiro-Wilk | p_value |
|--------------|---------|
| ۰٫۹۸۱ | ۰٫۶۷۶ |



شکل ۳: نمودار Q-Q داده‌های سرعت باد.

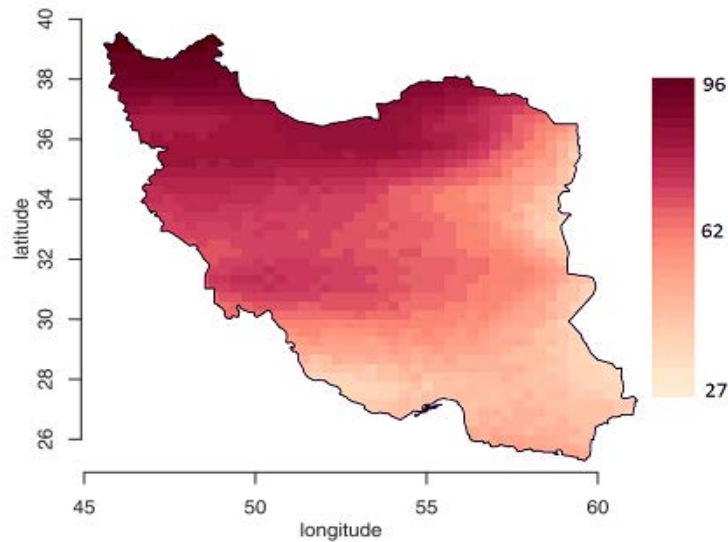
جدول ۸: برآورد و واریانس برآوردهای پارامترها در تحلیل داده‌های سرعت باد تحت مدل خطای اندازه‌گیری.

| پارامتر | β_0 | β_1 | σ^2 | ω^2 | θ | τ^2 |
|---------|-----------|-----------|------------|------------|----------|----------|
| برآورد | ۵۹٫۶ | -۱٫۸۲ | ۱٫۴۲ | ۱٫۲۳ | ۳٫۵۴ | ۰٫۱۷ |
| واریانس | ۰٫۵۶ | ۰٫۴۹ | ۰٫۶۸ | ۰٫۹۴ | ۰٫۸۸ | ۱٫۱۴ |

گویای عملکرد بهتر رویکرد پیشنهادی در مقایسه با مدل ناپخته در هر یک از این دو مورد بود. در پایان نیز تحلیل داده‌های بیشینه سرعت باد به عنوان کاربردی از مدل پیشنهادی ارائه شد. حال از آن‌جا که عامل خطای اندازه‌گیری الزاماً در اندازه‌گیری همه پیش‌گوها دخیل نیست، ارائه تحلیل مدل بر اساس تفکیک پیش‌گوها به دو دسته آمیخته با خطا و مشاهدات دقیق به عنوان زمینه مطالعاتی جدید پیشنهاد می‌شود. به‌علاوه می‌توان به مدل اثرات تصادفی نیز اضافه نمود تا تعمیم حاصل را توانمندتر نمود.

تقدیر و تشکر

نویسندگان کمال تشکر و قدردانی از داوران محترم مقاله و ویراستار مجله که با پیشنهادات ارزنده خود باعث بهبود مقاله شدند را دارند.



شکل ۴: نقشه پیشگویی بیزی داده‌های سرعت باد تحت مدل خطای اندازه‌گیری.

مراجع

- [۱] سینی، ف. و همایون فال، الف. (۱۳۹۵)، تحلیل بیزی تقریبی داده‌های فضایی-زمانی با استفاده از یک میدان تصادفی مارکوفی گاوسی، مجله علوم آماری، ۱۰، ۲۳۳-۲۶۰.
- [2] Berger, J. O., De Oliveira, V. and Sansó, B. (2001), Objective Bayesian Analysis of Spatially Correlated Data, *Journal of the American Statistical Association*, **96**, 1361-1374.
- [3] Bernardinelli, L., Pascutto, C., Best, N. G. and Gilks, W. R. (1997), Disease Mapping with Errors in Covariates, *Statistics in Medicine*, **16**, 741-752.
- [4] Bibby, B. M. and Sørensen, M. (2003), "Hyperbolic Processes in Finance" in *Handbook of Heavy-Tailed Distributions in Finance*, Elsevier, New York.
- [5] Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006), *Measurement Error in Nonlinear Models: A Modern Perspective*, CRC Press.

- [6] Cressie, N. (1993), *Statistics for Spatial Data*, Revised Edition, John Wiley, New York.
- [7] Deardon, R., Habibzadeh, B. and Chung, H. (2012), Spatial Measurement Error in Infectious Disease Models, *Journal of Applied Statistics*, **39**, 1139–1150.
- [8] Deutsch, C. V. and Journel, A. G. (1992), *GSLIB: Geostatistical Software Library and User's Guide*, Oxford University Press, New York.
- [9] Gelman, A. (2006), Prior Distributions for Variance Parameters in Hierarchical Models (Comment on Article by Browne and Draper), *Bayesian Analysis*, **1**, 515–534.
- [10] Griffin, J. and Brown, P. (2010), Inference with Normal-Gamma Prior Distributions in Regression Problems, *Bayesian Analysis*, **5**, 171–188.
- [11] Gryparis, A., Paciorek, C. J., Zeka, A., Schwartz, J. and Coull, B. A. (2009), Measurement Error Caused by Spatial Misalignment in Environmental Epidemiology, *Biostatistics*, **10**, 258–274.
- [12] Hamzeshpour, N., Eghbal, M. K., Bogaert, P., Toomanian N. and Sokouti, R. S. (2013), Spatial Prediction of Soil Salinity Using Kriging with Measurement Errors and Probabilistic Soft Data, *Arid Land Research and Management*, **27**, 128-139.
- [13] Huque, M. H., Bondell, H. D., Carroll, R. J. and Ryan, L. M. (2016), Spatial Regression with Covariate Measurement Error: A Semiparametric Approach, *Biometrics*, **72**, 678-686
- [14] Huque, M. H., Bondell, H. D. and Ryan, L. (2014), On the Impact of Covariate Measurement Error on Spatial Regression Modelling, *Environmetrics*, **25**, 560-570.

- [15] Ibrahim, J. and Chen, H. (2000), Power Prior Distributions for Regression Models, *Statistical Science*, **15**, 46–60.
- [16] Le Gallo, J. and Fingleton, B. (2012), Measurement Errors in a Spatial Context, *Regional Science and Urban Economics*, **42**, 114-125.
- [17] Militino, A. F., Ugarte, M. D., Iribas, J. and Lizarraga-Garcia, E. (2013), Mapping GPS Positional Errors Using Spatial Linear Mixed Models, *Journal of Geodesy*, **87**, 675–685.
- [18] Sheppard, L., Burnett, R. T., Szpiro, A. A., Kim, S. Y., Jerrett, M., Pope III, C. A. and Brunekreef, B. (2012), Confounding and Exposure Measurement Error in Air Pollution Epidemiology, *Air Quality, Atmosphere & Health*, **5**, 203-216.
- [19] Stein, M. L. (1999), *Interpolation of Spatial Data: Some Theory of Kriging*, Springer-Verlag, New York.
- [20] Tadayon V. (2017), Bayesian Analysis of Censored Spatial Data Based on a Non-Gaussian Model, *Journal of Statistical Research of Iran*, **13**(2), 155-180.
- [21] Tadayon V. Khaledi, M. J. (2015), Bayesian Analysis of Skew Gaussian Spatial Models Based on Censored Data, *Communications in Statistics-Simulation and Computation*, **44**(9), 2431-2441.
- [22] Thomas, D. C. (2013), Measurement Error in Spatial Exposure Models: Study Design Implications, *Environmetrics*, **24**, 518-520.
- [23] West, O. R., Siegrist, R. L., Mitchell, T. J. and Jenkins, R. A. (1995), Measurement Error and Spatial Variability Effects on Characterization of Volatile Organics in the Subsurface, *Environmental Science & Technology*, **29**, 647-656.

-
- [24] Xia, H. and Carlin, B. P. (1998), Spatio-temporal Models with Errors in Covariates: Mapping Ohio lung Cancer Mortality, *Statistics in Medicine*, **17**, 2025-2043.
- [25] Yi, L., Tang, H. and Lin, X. (2009), Spatial Linear Mixed Models with Covariate Measurement Errors, *Statistica Sinica*, **19**, 1077-1093.

Analysis of Gaussian Spatial Models with Covariate Measurement Error

Tadayon, V., Rasekh, A.

Department of Statistics, Shahid Chamran University, Ahvaz , Iran.

Abstract:

Uncertainty is an inherent characteristic of biological and geospatial data which is almost made by measurement error in the observed values of the quantity of interest. Ignoring measurement error can lead to biased estimates and inflated variances and so an inappropriate inference. In this paper, the Gaussian spatial model is fitted based on covariate measurement error. For this purpose, we adopt the Bayesian approach and utilize the Markov chain Monte Carlo algorithms and data augmentations to carry out calculations. The methodology is illustrated using simulated data.

Keywords: Gaussian Spatial Model, Measurement Error, Bayesian Analysis.

Mathematics Subject Classification (2010): 62F15, 62H11.