

تلفیق بیزی مدل‌های قطعی و کریگیدن برای تحلیل داده‌های وابسته فضایی

نسرین مهدیان‌فرد، محسن محمدزاده*؛ دانشگاه تربیت مدرس، گروه آمار

پذیرش ۹۲/۴/۹

دریافت ۹۱/۱۱/۷

چکیده

برقراری پیوند میان سامانه اطلاعات مکانی و روش‌های تصمیم‌گیری مرسوم ابداع و تکامل روش تلفیق داده‌های مکانی است. روش‌های تلفیق اطلاعات، داده‌ها را با هدف دستیابی به نتیجه بهتر برای افزایش قابلیت تفسیر اطلاعات با یکدیگر ترکیب می‌کنند. در این مقاله روش تلفیق بیزی برای ترکیب اندازه‌ها و خروجی مدل‌های قطعی و کریگیدن بررسی شدند. سپس با استفاده از آن‌ها داده‌های آژن شهر تهران تحلیل شد و با ملاک میانگین توان‌های دوم خطا نتایج حاصل از آن‌ها ارزیابی و مقایسه شدند.
واژه‌های کلیدی: تلفیق بیزی، مدل قطعی، کریگیدن.

مقدمه

منظور از تلفیق داده‌ها^۱، فرآیندی است که با ترکیب داده‌های موجود در ابعاد یا منابع چندگانه، استنباط‌ها بر اساس اطلاعات فراهم آمده که نسبت به هر منبع تنها، مؤثرتر و دقیق‌ترند انجام می‌شود. مدل قطعی^۲ مدلی ریاضی برای بررسی روابط علت و معلولی است، که به‌ازای مجموعه‌های مشخص از ورودی‌ها، به‌عنوان جواب مدل، خروجی معین منحصر به‌فرد ارائه می‌دهد. به‌عنوان مثال مدل‌های قطعی، در نواحی معین، آلودگی آژن روی توری‌های شبکه‌ای^۳ را شبیه‌سازی می‌کنند و اندازه‌ها^۴ در هر یک از ایستگاه‌های واقع در منطقه بررسی شده به‌دست می‌آیند. معمولاً نمی‌توان به‌طور مستقیم خروجی‌های مدل و مشاهدات اندازه‌گیری شده را با هم مقایسه کرد. با این حال، وجود هم‌بستگی بین اندازه‌ها و خروجی‌های مدل را می‌توان بررسی کرد. اگر چنین هم‌بستگی وجود داشته باشد می‌توان از مدل‌های آماری برای ترکیب اندازه‌ها با خروجی‌های مدل برای به‌دست آوردن پیش‌گویی‌های بهتر استفاده کرد. علاوه بر این، می‌توان خروجی مدل‌های قطعی را برای پیش‌گویی کالیبره^۵ کرد. پوول و رفتری (۲۰۰۰) از ترکیب پیشین‌های مختلف برای کاهش عدم قطعیت خروجی‌های یک مدل بیزی استفاده کردند. کسیتلا و چمید (۲۰۰۰) و هاگرف و همکاران (a,b2001) خروجی‌های مدلی قطعی در هواشناسی و علوم محیطی را با استفاده از نمودارهای پراکنش و تحلیل کمترین توان‌های دوم اندازه‌ها بررسی کردند و هم‌بستگی بین آن‌ها را بررسی کردند. رمزی و همکاران (۲۰۰۷) از روش‌های ناپارامتری برای برآورد

*نویسنده مسئول Mohsen_m@modares.ac.ir

1. Data fusion
2. Deterministic models
3. Grid cells
4. Measurment
5. Calibrate

پارامترهای مدل‌های قطعی در معادلات مختلف استفاده کردند. گاتورپ و همکاران (۱۹۸۷) ساختار آماری مدل‌های قطعی را بررسی کردند.

کریگیدن (ماترون، ۱۹۶۳) روشی برای پیش‌گویی فضایی در موقعیت‌های مشخص و بهترین پیش‌گویی خطی نارایب است. هر چند محاسبات کریگیدن سراسر است، اما مستلزم معلوم بودن هم‌تغییرنگار و ثابت بودن پارامترها است. از این رو میزان عدم قطعیت واقعی پیش‌گویی‌ها کم لحاظ می‌شود و لازم است در به‌کارگیری نتایج به آن توجه شود. برای کاهش عدم قطعیت در آمار فضایی می‌توان مجموعه داده‌های اندازه‌گیری شده در ایستگاه‌های مانیتورینگ را با خروجی مدلی قطعی موجود برای آن پدیده تلفیق کرد. اولین بار پول و رفتی (۲۰۰۰) تلفیق بیزی را پیشنهاد کردند. سپس سانسو و گوین (۲۰۰۲) و فوننتز و رفتی (۲۰۰۵) آن را توسعه دادند. برلینر (۲۰۰۳) و برلینر و ویکال (۲۰۰۵) نیز نحوه استفاده از مدل‌های تلفیق بیزی برای ترکیب اطلاعات از منابع مختلف را بیان کرده‌اند. مدل تلفیق بیزی، سطح واقعی پنهان Y پدیده که با داده‌های ایستگاه‌ها و مدل قطعی ارائه می‌شوند را به نوعی با هم ترکیب می‌کند. گاهی داده‌های حاصل از مدلی قطعی بر اساس Y توری درشت‌تر از توری اندازه‌های اصلی هستند. به‌عنوان مثال داده‌های ساعتی غلظت Y در ناحیه $[6 \times 6]$ کیلومتر مربعی را در نظر بگیرید. وقتی برای درون‌یابی، اطلاعات با خروجی مدل‌های قطعی ترکیب می‌شوند، عدم تطابق مقیاس‌ها مستلزم کالیبره کردن اطلاعات شبیه‌سازی شده از مدل‌ها است. برای رفع این مشکل، فوننتز و رفتی (۲۰۰۲ و ۲۰۰۵) و فوننتز و همکاران (۲۰۰۳) مدل تلفیق را مطرح کردند. در این مقاله مدل تلفیق بیزی و نحوه پیش‌گویی با آن معرفی می‌شود. سپس نحوه کاربست مدل تلفیق بیزی برای تحلیل داده‌های آلاینده Y در شهر تهران پیشنهاد و دقت نتایج حاصل از دو روش کریگیدن و تلفیق بیزی تحلیل و ارزیابی شدند.

مدل تلفیق بیزی

فرض کنید ناحیه بررسی شده $d \geq 1, D \subset R^d$ را شبکه کرده و هر Y از سلول‌های آن با B و هر عضو آن با s نمایش داده شود. هدف مرتبط ساختن مقادیر اندازه‌گیری شده در موقعیت $s \in D$ و خروجی مدل‌های قطعی از طریق فرایندی واقعی است. بدین منظور فرایند واقعی با $\{Z(s): s \in D \subset R^d\}$ ، فرایند اندازه‌گیری با $\{\hat{Z}(s): s \in D\}$ و خروجی مدل‌های قطعی در سلول B با $\tilde{Z}(B)$ نشان داده می‌شوند. فرایندهای واقعی قابل مشاهده نیستند و باید پیش‌گویی شوند. برای مطابقت با فرایند واقعی $Z(s)$ ، فرض می‌شود فرایندهای خروجی مدل‌های قطعی $\{\hat{Z}(s): s \in D\}$ موجود و به موقعیت‌های s وابسته‌اند. هدف مرتبط ساختن خروجی مدل‌های قطعی و اندازه‌ها با مقدار واقعی $Z(s)$ در شبکه‌های کوچک مقیاس از طریق $\hat{Z}(s)$ و $\tilde{Z}(s)$ است. بنا بر این فرایندهای واقعی فصل مشترک دو فرایند $\hat{Z}(s)$ و $\tilde{Z}(s)$ هستند، که لزوماً مقیاس یکسانی ندارند. برای غلبه بر مشکل عدم تطابق مقیاس‌ها، می‌توان $\hat{Z}(s)$ و $\tilde{Z}(s)$ را از طریق فرایندهای واقعی ادغام کرد، از این رو «تلفیق» نام گرفته است.

در مدل تلفیق، سطحی از پدیده‌ای واقعی در نظر گرفته می‌شود و اندازه‌گیری‌ها در ایستگاه‌های مانیتورینگ و خروجی مدل‌های قطعی با روشی مناسب در اطراف سطح واقعی تلفیق می‌شوند. میانگین پدیده در هر سلول

مشبکه‌ای B بر حسب سطح نقطه‌ای $Z(s)$ بدین‌صورت بدست می‌آید:

$$Z(B) = \frac{1}{|B|} \int_B Z(s) ds \quad (۱)$$

که در آن $|B|$ مساحت مشبکه B است. فرآیند واقعی $Z(s)$ مرتبط با $\tilde{Z}(s)$ در موقعیت s را می‌توان به‌صورت $\hat{Z}(s) = Z(s) + e(s)$ در نظر گرفت. از طرفی فرآیند واقعی به‌صورت $Z(s) = \mu(s) + \epsilon(s)$ قابل تجزیه است، که در آن $\mu(s)$ روند فضایی است و اغلب با توابعی معلوم از موقعیت‌ها تعیین می‌شود. خروجی مدل قطعی را می‌توان بدین‌صورت مدل‌بندی کرد:

$$\tilde{Z}(s) = a(s) + b(s)Z(s) + \delta(s) \quad (۲)$$

که در آن $a(s)$ اریبی جمعی^۶ و $b(s)$ اریبی ضربی^۷ نامیده می‌شوند و $\delta(s)$ فرآیند خطا با توزیع $N(0, \sigma_\delta^2)$ است. فرض کنید خروجی مدل قطعی در مشبکه $B_1, \dots, B_J$ فراهم شده است، فرآیندهای در سطح نقطه با استفاده از (۱) در مدل (۲) بدین‌صورت به سطح مشبکه تبدیل می‌شوند:

$$\tilde{Z}(B_j) = \int_{B_j} a(s) ds + \int_{B_j} b(s)Z(s) ds + \int_{B_j} \delta(s) ds \quad (۳)$$

اگر برای $\hat{Z}$ و $\tilde{Z}$ توزیع نرمال چندمتغیره در نظر گرفته شود آن‌گاه، پیش‌گویی فضایی فرآیند واقعی در موقعیت جدید s' را می‌توان از طریق توزیع پیش‌گوی پسین $p(Z(s') | \hat{Z}, \tilde{Z})$ که نرمال چندمتغیره است انجام داد. چنان‌چه $\hat{Z}(s_i)$ مشاهده‌های از ایستگاه واقع در مکان s_i ، $\tilde{Z}(B_j)$ خروجی مدل قطعی در سلول B_j و $Z(s)$ فرآیند واقعی باشد، داده‌های ایستگاه از مدل

$$\hat{Z}(s_i) = Z(s_i) + e(s_i), \quad i = 1, \dots, n$$

پیروی می‌کنند و خروجی مدل قطعی برای کالیبره کردن به‌صورت (۳) مدل‌بندی می‌شود، که در آن $a(s)$ و $b(s)$ فرآیندهای تصادفی گاوسی هستند. چون محاسبه تحلیلی انتگرال‌های موجود در (۳) میسر نیست از روش‌های مونت کارلویی (بانرجی و همکاران، ۲۰۰۴) برای محاسبه آن‌ها استفاده می‌شود. اکنون می‌توان مؤلفه‌های مدل تلفیق بیزی را بدین‌صورت خلاصه کرد:

$$\hat{Z}(s) = Z(s) + e(s)$$

$$Z(s) = \mu(s) + \epsilon(s)$$

$$\tilde{Z}(s) = a(s) + b(s)Z(s) + \delta(s)$$

$$Z(B) = \frac{1}{|B|} \int_B Z(s) ds$$

$$\tilde{Z}(B_j) = \frac{1}{|B|} \int_B a(s) ds + \frac{1}{|B|} \int_B b(s)Z(s) ds + \frac{1}{|B|} \int_B \delta(s) ds$$

$$\mu(s) = X(s)\beta \quad (۴)$$

در این مدل خطای اندازه‌گیری $e(s)$ و خطای خروجی مدل $\delta(s)$ از هم مستقل هستند و مانده‌ها، $\epsilon(s)$ ، هم‌بسته فضایی با میانگین صفر و ماتریس کوواریانس Σ_θ هستند، که در آن θ بردار پارامتر کوواریانس است. هر يك از $Z(B)$ و $\tilde{Z}(B)$ انتگرال‌هایی بر حسب $Z(B)$ و $\tilde{Z}(B)$ روی سلول B هستند و در عمل فقط فرآیندهای

6. Additive

7. Multiplicative

$\hat{Z}(s)$ و $\tilde{Z}(B)$ برای خروجی‌های مدل در ایستگاه‌های اندازه‌گیری واقع در موقعیت‌های $s_1, \dots, s_n$ و سلول‌های شبکه‌ای قابل مشاهده هستند.

میانگین فرآیند واقعی به صورت $\mu(s) = X(s)\beta$ قابل بیان است، که در آن $X(s)$ متشکل از توابعی چند جمله‌ای از مختصات موقعیت‌های s و β بردار ضرایب متناظر با آن است. برای سادگی، فعلاً فرض می‌شود a و b ثابت هستند، ولی بعداً در تحلیل داده‌ها، a تابعی از موقعیت‌ها و b همچنان ثابت در نظر گرفته می‌شود. فوئنتر و رفتزی (۲۰۰۵) نشان دادند که فقط وقتی می‌توان مدلی مناسب برازش داد که $b(s) = b$ باشد. برای تقریب انتگرال‌ها با روش‌های $MCMC$ ، L موقعیت $s_{1,B}, \dots, s_{L,B}$ از درون سلول B نمونه‌گیری می‌شود، که نقاط نمونه‌گیری^۸ نامیده می‌شوند. سپس مقادیر فرآیندها محاسبه می‌شوند. آن‌گاه مقدار تقریبی $Z(B)$ به صورت $Z(B) \approx \frac{1}{L} \sum_{j=1}^L Z(s_{j,B})$ حاصل می‌شود.

برازش مدل تلفیق بیزی

در این بخش نحوه برازش مدل تلفیق بیزی با استفاده از الگوریتم‌های نمونه‌گیری گیزر معرفی خواهد شد. اصولاً رهیافت بیزی با در نظر گرفتن توزیع‌های پیشین مناسب و تعیین توزیع‌های پسین پارامترها بر اساس داده‌ها، به دنبال لحاظ کردن عدم قطعیت برآورد پارامترها است. در اینجا پارامترها شامل، واریانس‌های خط‌های تصادفی σ_δ^2 و σ_ϵ^2 ، بردار ضرایب β ، فرآیندهای واقعی مورد نظر Z ، پارامترهای کالیبراسیون a و b و بردار پارامترهای کوواریانس θ است. از بردار Z برای مشخص کردن مقادیر فرآیند واقعی مورد نظر در ایستگاه‌ها و نقاط نمونه‌گیری درون سلول شبکه‌ای استفاده می‌شود. اگر تعداد ایستگاه‌ها n و تعداد سلول‌های شبکه‌ای m باشد، آن‌گاه بعد بردار Z برابر $m + n \times L$ است. ابعاد اندازه‌های $\tilde{Z}$ و خروجی مدل $\tilde{Z}$ نیز به ترتیب n و m هستند.

فرض کنید $H = \{X, \tilde{Z}, \tilde{Z}\}$ تمام داده‌ها را نشان دهد، که در آن X ماتریس متغیرهای تبیینی در مدل (۴) است. با فرض آن‌که پارامترهای $(\sigma_\epsilon^2, \sigma_\delta^2, \beta, \theta)$ دارای توزیع‌های پیشین مستقل هستند، تابع درست‌نمایی را می‌توان به صورت

$$\begin{aligned} p(\theta, \beta, a, b, \sigma_\delta^2, \sigma_\epsilon^2, Z, \tilde{Z}, \tilde{Z}) &= p(\tilde{Z}|Z, \sigma_\epsilon^2) p(\tilde{Z}|Z, a, b, \sigma_\delta^2) p(Z|\beta, \theta) p(\sigma_\delta^2, \sigma_\epsilon^2, \theta, \beta) \\ &= \phi_{\Sigma_1}(\tilde{Z} - A_0 Z) \phi_{\Sigma_2}(\tilde{Z} - a - b A_1 Z) \phi_{\Sigma_3}(Z - X\beta) p(\sigma_\delta^2) p(\sigma_\epsilon^2) p(\beta) p(\theta) \end{aligned}$$

تجزیه کرد، که در آن $\phi_{\Sigma}(\mu)$ چگالی نرمال چندمتغیره با میانگین μ و کوواریانس Σ ،

$$A_0 = \begin{pmatrix} 1 & \dots & 00 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & 10 & \dots & 0 \end{pmatrix}_{n \times (mL+n)}$$

$$A_1 = \begin{pmatrix} 0 & \dots & 0 & 1/L & \dots & 1/L & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 & \dots & 1/L & \dots & 1/L \end{pmatrix}_{m \times (mL+n)}$$

در ماتریس A_0 ، n ستون اول ماتریسی همانی و بقیه درایه‌ها صفر هستند. در ماتریس A_1 ، درایه‌های سطریام از ستون $n + 1 + (i - 1) \times L$ تا $n \times i \times L$ ، برابر $\frac{1}{L}$ و بقیه درایه‌ها صفر هستند. $\Sigma_1 = \sigma_\epsilon^2 I$ ماتریس کوواریانس بردار خطای اندازه‌ها $\Sigma_2 = \sigma_\delta^2 I$ ، $e = [e(s_1), \dots, e(s_n)]^T$ ماتریس کوواریانس بردار خطای

8. Sampling points

خروجی مدل $\delta = [\delta(B_1), \dots, \delta(B_n)]^T$ و $\Sigma_3 = \Sigma(\theta)$ ماتریس کوواریانس Z است. توجه شود که برای Z داده شده $\tilde{Z}$ و $\tilde{Z}$ مستقل هستند و چگالی توأم شرطی $(Z, \tilde{Z} | \theta, \beta, a, b, \sigma_\delta^2, \sigma_e^2)$ بدین صورت است:

$$\begin{aligned} p(Z, \tilde{Z} | \theta, \beta, a, b, \sigma_\delta^2, \sigma_e^2) \\ \propto \exp\left\{-\frac{1}{2}\left[-Z^T(A_0^T \Sigma_1^{-1} \tilde{Z} + bA_1^T \Sigma_2^{-1}(\tilde{Z} - a) + \Sigma_3^{-1}\mu) \right. \right. \\ \left. \left. - (\tilde{Z}^T \Sigma_1^{-1} A_0 + b(\tilde{Z} - a)\Sigma_2^{-1} A_1 + \mu^T \Sigma_3^{-1})Z \right. \right. \\ \left. \left. + Z^T(A_0^T \Sigma_1^{-1} A_0 + b^2 A_1^T \Sigma_2^{-1} A_1 + \Sigma_3^{-1})Z\right]\right\} + C \end{aligned}$$

که در آن $\mu = X\beta$ بردار میانگین Z و C سایر عبارات بدون Z است. چون توزیع پیشین نرمال، مزدوجی از توزیع نمونه‌ای نرمال است، توزیع شرطی کامل Z باید بدین صورت باشد:

$$p(Z | \theta, \beta, a, b, \sigma_\delta^2, \sigma_e^2, H) \propto \exp\left\{-\frac{1}{2}[Z^T \tilde{\Sigma}^{-1} Z - \tilde{\mu}^T \tilde{\Sigma}^{-1} Z]\right\}$$

اگر توزیع پیشین β نرمال چند متغیره به صورت $MVN(\beta_0, F)$ باشد، آنگاه توزیع شرطی کامل آن نیز $MVN(Bb, B)$ خواهد شد، که در آن $B^{-1} = X^T \Sigma_3^{-1} X + F^{-1}$ و $b = X^T \Sigma_3^{-1} Z + F^{-1} \beta_0$ (فونتز و رفتی، ۲۰۰۵). فرض کنید در تابع کوواریانس نمایی $C_\theta(h) = \sigma \exp\{-|h|/\rho\}$ دو مؤلفه بردار $\theta = (\sigma, \rho)$ مستقل‌اند و ρ دارای توزیع پیشین گاما و σ دارای توزیع پیشین گامای وارون هستند. در حالت کلی داریم:

$$\Sigma_3 = \sigma \exp\{-C/\rho\},$$

که در آن C ماتریس فاصله اقلیدسی^۹ بین ایستگاه‌ها و نقاط نمونه‌گیری و $g(\cdot)$ تابع همبستگی فضایی است. بنا بر این توزیع شرطی کامل θ بدین صورت به دست می‌آید:

$$\begin{aligned} p(\theta | \beta, Z) \propto p(\rho) \sigma^{-\alpha-1-\frac{n}{2}} \left| e^{-\frac{C}{\rho}} \right|^{-\frac{1}{2}} \\ \times \exp\left[-\frac{1}{\sigma} \left(\frac{1}{\gamma} + \frac{1}{2} (Z - X\beta)^T \left(e^{-\frac{C}{\rho}} \right)^{-1} (Z - X\beta) \right)\right] \end{aligned}$$

در نتیجه توزیع شرطی کامل σ ، گامای واون با پارامترهای

$$\tilde{\alpha} = \alpha + n/2, \tilde{\gamma} = \left(\frac{1}{\gamma} + \frac{1}{2} (Z - X\beta)^T \left(e^{-\frac{C}{\rho}} \right)^{-1} (Z - X\beta) \right)^{-1}$$

است. پارامتر ρ دارای توزیع شرطی کاملی که بتوان به طور مستقیم از آن مقادیر تصادفی تولید کرد نیست، بنا بر این از الگوریتم متروپولیس هستینگس (هستینگس، ۱۹۷۰) برای نمونه‌گیری از ρ با σ ثابت استفاده می‌شود. اگر σ_e^2 و σ_δ^2 به ترتیب دارای توزیع پیشین $IG \sim (\alpha_1, \frac{1}{\lambda_1})$ و $IG \sim (\alpha_2, \frac{1}{\lambda_2})$ باشند، آنگاه توزیع‌های

شرطی کامل σ_e^2 و σ_δ^2 به ترتیب بدین صورت هستند:

$$\sigma_\delta^2 | (a, b, \tilde{Z}, Z) \sim IG\left(\alpha_1 + \frac{m}{2}, \left(\lambda_1 + \frac{1}{2} \tilde{\lambda}\right)^{-1}\right)$$

9. Euclidean distance

$$\sigma_e^2 | (\tilde{\mathbf{Z}}, \mathbf{Z}) \sim IG \left(\alpha_2 + \frac{n}{2}, \left(\lambda_2 + \frac{1}{2} \gamma \right)^{-1} \right)$$

که در آن‌ها $\tilde{\lambda} = (\tilde{\mathbf{Z}} - \mathbf{a} - b\mathbf{A}_1\mathbf{Z})^T (\tilde{\mathbf{Z}} - \mathbf{a} - b\mathbf{A}_1\mathbf{Z})$ و $\gamma = (\tilde{\mathbf{Z}} - \mathbf{A}_0\mathbf{Z})^T (\tilde{\mathbf{Z}} - \mathbf{A}_0\mathbf{Z})$ با در نظر گرفتن توزیع پیشین $(a, b)' \sim MVN(\bar{\mathbf{B}}, \bar{\mathbf{F}})$ توزیع شرطی توأم کامل پارامترهای کالیبراسیون به صورت $(a, b)' | (\tilde{\mathbf{Z}}, \mathbf{A}_1\mathbf{Z}, \Sigma_2) \sim MVN(\bar{\mathbf{B}}\bar{\mathbf{C}}, \bar{\mathbf{B}})$ حاصل می‌شوند، که در آن $\bar{\mathbf{B}} = \left(\frac{\mathbf{1}^T}{(\mathbf{A}_1\mathbf{Z})^T} \right) \Sigma_2^{-1} (\mathbf{1} \quad \mathbf{A}_1\mathbf{Z})$ و $\bar{\mathbf{C}} = \left(\frac{\mathbf{1}^T}{(\mathbf{A}_1\mathbf{Z})^T} \right) \Sigma_2^{-1} (\mathbf{1} \quad \mathbf{Z}) + \bar{\mathbf{F}}^{-1}$ دارای بعد مشابه بردار $\tilde{\mathbf{Z}}$ است.

برای پیش‌گویی فرایندهای مورد نظر در ایستگاه‌های فاقد مشاهده فرض کنید تحقق‌های فرایندهای واقعی در ایستگاه‌های فاقد مشاهده با نماد $\mathbf{Z}_u$ نشان داده شوند. درون‌یابی فرایندها مستلزم پیدا کردن توزیع پسین $(\mathbf{Z}_u | \mathbf{Z}_g)$ است، که در آن $\mathbf{Z}_g$ تحقق‌های فرایندهای واقعی در ایستگاه‌های مانیتورینگ و نقاط نمونه‌گیری در درون سلول‌های شبکه است. ملاحظه می‌شود که $E(\mathbf{Z}_u) = \mu_u$ و $E(\mathbf{Z}_g) = \mu_g$. بنا بر این میانگین و واریانس توزیع شرطی $(\mathbf{Z}_u | \mathbf{Z}_g, \boldsymbol{\beta}, \boldsymbol{\theta})$ به ترتیب بدین‌صورتند:

$$E(\mathbf{Z}_u | \mathbf{Z}_g, \boldsymbol{\beta}, \boldsymbol{\theta}) = \mu_u + \Sigma_{gu}^{-1} \Sigma_g^{-1} (\mathbf{Z}_g - \mu_g)$$

$$Var(\mathbf{Z}_u | \mathbf{Z}_g, \boldsymbol{\beta}, \boldsymbol{\theta}) = \Sigma_u - \Sigma_{gu} \Sigma_g^{-1} \Sigma_g$$

که در آن Σ_u ماتریس کوواریانس $\mathbf{Z}_u$ و Σ_{gu} ماتریس کوواریانس بین $\mathbf{Z}_g$ و $\mathbf{Z}_u$ است. فرض کنید Σ_{ug} نشان دهنده ترانهاد Σ_{gu} باشد. با توجه به آن‌که در مدل تلفیق بیزی (۴) داریم $\hat{\mathbf{Z}}(s) = \mathbf{Z}(s) + \mathbf{e}(s)$ ، میانگین و واریانس شرطی $(\hat{\mathbf{Z}}_u | \mathbf{Z}_g, \boldsymbol{\beta}, \boldsymbol{\theta})$ به این‌صورت به‌دست می‌آید:

$$E(\hat{\mathbf{Z}}_u | \mathbf{Z}_g, \boldsymbol{\beta}, \boldsymbol{\theta}) = \mu_u + \Sigma_{gu}^{-1} \Sigma_g^{-1} (\mathbf{Z}_g - \mu_g)$$

$$Var(\hat{\mathbf{Z}}_u | \mathbf{Z}_g, \boldsymbol{\beta}, \boldsymbol{\theta}) = \Sigma_u - \Sigma_{gu} \Sigma_g^{-1} \Sigma_g + \sigma_e^2 \mathbf{I}$$

که در آن بعد ماتریس همانی $\mathbf{I}$ تعداد ایستگاه‌هایی است که در آن‌ها پیش‌گویی انجام می‌شود.

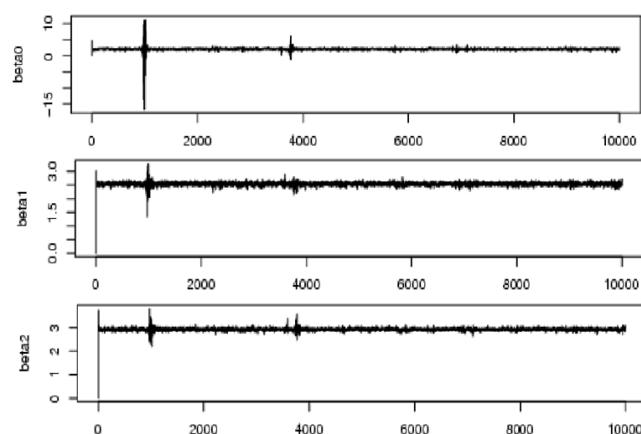
مطالعه شبیه‌سازی

با فرض آن‌که فرایندهای واقعی، مانا هستند، ابتدا تحلیل مدل تلفیق بیزی با استفاده از الگوریتم MCMC ارزیابی می‌شود. سپس دقت پیش‌گویی‌ها با دو روش کریگیدن و تلفیق بیزی با هم مقایسه می‌شوند. برای شبیه‌سازی، ۱۵ ایستگاه مانیتورینگ و ۸۰ ایستگاه برای پیش‌گویی و تعداد سلول‌های شبکه ۲، ۱۰، ۲۰، ۴۰ و ۶۰ در نظر گرفته شده است. برای کاهش محاسبات، از درون هر سلول شبکه‌ای فقط یک نقطه نمونه‌برداری شده است. در حالی که می‌توان آن را برای هر تعداد تعمیم داد. مختصات ایستگاه‌های مانیتورینگ و ایستگاه‌های بدون مشاهده و نقاط نمونه‌گیری به‌طور یک‌نواخت در ناحیه $[-6 \times 6] \times [-6 \times 6]$ تولید شده‌اند. برای هر یک از ترکیبات ایستگاه‌ها و نقاط نمونه‌گیری، ۵۰ مجموعه داده‌ی مستقل از اندازه‌ها و خروجی‌ها طبق مدل (۴) تولید شده‌اند. میانگین فرایندهای واقعی به‌صورت

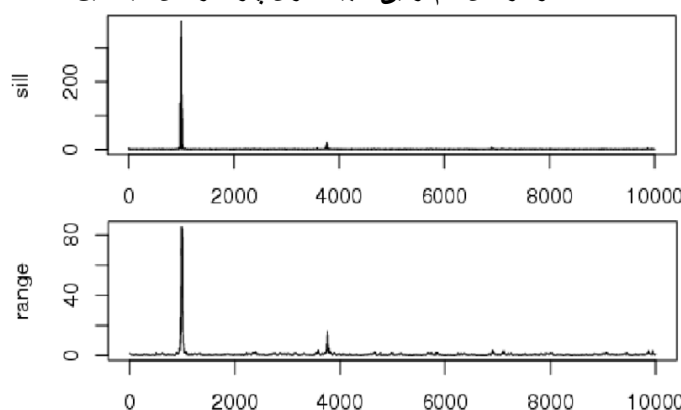
$$E[Z(s)] = \beta_0 + \beta_1 s_1 + \beta_2 s_2 \quad (۸)$$

در نظر گرفته شده است، که در آن s_1 و s_2 مختصات موقعیت‌ها و $Z(s)$ دارای کوواریانس فضایی

نمایی $C_\theta(h) = \sigma \exp\{-|h|/\rho\}$ است. با استفاده از الگوریتم نمونه‌گیری گیبز از توزیع‌های شرطی کامل همه پارامترهای ارائه شده نمونه تولید شده است، تعداد تکرارهای نمونه‌گیری گیبز پس از ۱۰۰۰ داغیدن ۱۰۰۰۰ در نظر گرفته شده است. شکل‌های ۱ تا ۳ این تعداد تکرار را برای رسیدن به همگرایی زنجیر مارکوفی تأیید می‌کند.



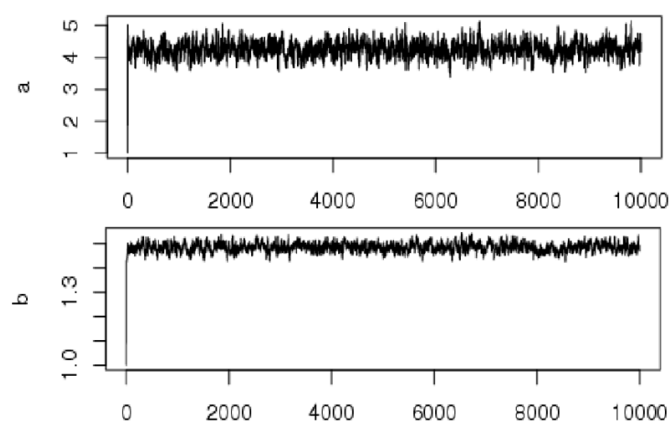
شکل ۱. نمودارهای همگرایی شبیه‌سازی پارامترهای میانگین



شکل ۲. نمودارهای همگرایی شبیه‌سازی پارامترهای کواریانس

برآورد پارامترهای مدل تلفیق بیزی در جدول ۱ آورده شده است. ستون‌های ۴ تا ۸، متوسط میانگین‌ها و انحراف استاندارد توزیع‌های شرطی کامل پارامترهای مدل تلفیق بیزی هستند، که بر اساس ۵۰ مجموعه داده تولید شده مستقل محاسبه شده‌اند. چنان‌که در جدول ۱ ملاحظه می‌شود برآورد پارامترهای کالیبراسیون و ضرایب روند نسبتاً دقیق است و به مقادیر واقعی آن‌ها خیلی نزدیک هستند و با افزایش تعداد سلول‌های شبکه‌ای دقت برآوردها افزایش می‌یابد. برآورد پارامترهای کواریانس $\theta = (\sigma, \rho)$ تقریباً دقیق است. اما با این‌که میانگین پسین با افزایش تعداد سلول‌های شبکه‌ای از ۲ تا ۶۰، به مقدار واقعی آن نزدیک می‌شود، انحراف استاندارد حتی با ۶۰ سلول شبکه‌ای نسبتاً بزرگ باقی می‌ماند، که ممکن است ناشی از تعداد نقاط ناکافی نزدیک به یکدیگر باشد. به عبارت دیگر، از فرآیندهای کوچک مقیاس مختلف که برای برآورد هم‌تغییرنگار لازم است، اطلاع کافی در دست نیست. نمودارهای توزیع‌های پسین بیان‌گر آن هستند که ρ بسیار پراکنده است. در جدول ۲ میانگین توان‌های دوم خطای پیش‌گویی ($MSPE$) و مدت زمان انجام محاسبات برای دو روش کریگیدن و تلفیق بیزی آورده شده است. مقدار $MSPE$ و میانگین مدت زمان انجام محاسبات با روش کریگیدن ۲/۲۰ و ۵/۴۲ است.

در حالی که با افزایش تعداد سلول‌های شبکه‌ای، مقدار $MSPE$ برای مدل تلفیق به $۰/۸۵$ کاهش پیدا می‌کند اما مدت زمان انجام این محاسبات افزایش پیدا کرده است. یعنی روش تلفیق بیزی با وجود زمان طولانی‌تر برای انجام محاسبات، عملکرد بهتری نسبت به روش کریگیدن دارد و پیش‌گویی با آن دقت بیشتری نسبت به کریگیدن دارد. در هر یک از مجموعه داده‌ها، لزومی ندارد که با افزایش تعداد سلول‌های شبکه‌ای، مقدار میانگین توانهای دوم خطای پیش‌گویی کاهش یابد. از این رو، الگوهای غیرنزولی متنوع مقدار میانگین توان‌های دوم خطای پیش‌گویی ممکن است ناشی از نوع نمونه‌گیری از مجموعه داده‌ها باشد. در شکل ۴، سه نمودار مقدارهای واقعی ازن، پیش‌گویی تلفیق بیزی و پیش‌گویی کریگیدن در حالتی که ۶۰ سلول شبکه‌ای وجود دارد برای ۱۰ ایستگاه

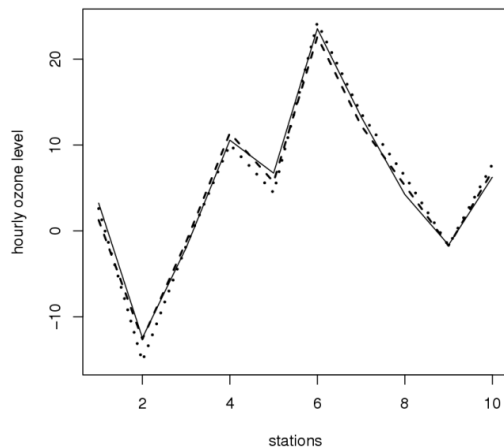


شکل ۳. نمودارهای همگرایی شبیه‌سازی پارامترهای کالیبراسیون

جدول ۱. برآورد پارامترهای مدل تلفیق بیزی

منبع	پارامتر مقدار واقعی	تعداد سلول				
		۶۰	۴۰	۲۰	۱۰	۲
کالیبراسیون	۴/۰۰	۴/۰۲	۳/۹۹	۴/۰۱	۳/۹۲	۴/۳۰
	۲/۴۲	۰/۲۴	۰/۳۱	۰/۴۴	۰/۵۹	۲/۴۲
	۱/۵۰	۱/۴۹	۱/۴۹	۱/۴۹	۱/۵۱	۱/۵۶
	۰/۳۲	۰/۰۱	۰/۰۲	۰/۰۳	۰/۰۴	۰/۳۲
روند	۲/۲۰	۲/۱۹	۲/۲۱	۲/۲۱	۲/۲۰	۲/۱۷
	۰/۶۹	۰/۲۵	۰/۳۴	۰/۵۷	۰/۶۴	۰/۶۹
	۲/۵۰	۲/۴۹	۲/۵۰	۲/۴۹	۲/۴۸	۲/۴۴
	۳/۰۰	۲/۹۹	۲/۹۸	۳/۰۰	۲/۹۷	۲/۹۳
کواریانس	۱/۲۰	۱/۲۹	۱/۲۱	۱/۲۳	۱/۵۸	۱/۶۸
	۲/۳۵	۲/۷۳	۱/۹۶	۳/۸۸	۳/۸۵	۲/۳۵
	۲/۵۰	۰/۷۳	۰/۹۵	۱/۸۸	۲/۲۰	۲/۵۷
	۲/۷۶	۱/۰۰	۱/۲۲	۲/۸۷	۲/۹۴	۲/۷۶

بدون مشاهده نشان داده شده است. مقادیر واقعی با خط توپر، پیش‌گویی‌های تلفیق بیزی با خط چین و پیش‌گویی‌های کریگیدن با نقطه‌چین مشخص شده‌اند. چنان‌که ملاحظه می‌شود پیش‌گویی‌های تلفیق کاملاً نزدیک به مقدارهای واقعی هستند. در حالت کلی، افزایش تعداد سلول‌های شبکه‌ای دقت پیش‌گویی‌های فضایی را برای ایستگاه‌های فاقد مشاهده بهبود می‌بخشد. در حالت کلی، افزایش تعداد سلول‌های شبکه‌ای دقت پیش‌گویی‌های فضایی را برای ایستگاه‌های فاقد مشاهده بهبود می‌بخشد. پیش‌گویی فضایی تلفیق، برآوردهای واقعی‌تری از پیش‌بینی عدم قطعیت نسبت به روش کریگیدن ارائه می‌دهد.



شکل ۴. پیش‌گویی فضایی در ۱۰ ایستگاه فاقد مشاهده
با مدل تلفیق بیزی بر اساس ۶۰ سلول مشبکه‌ای

مثال کاربردی

در این بخش داده‌های آلاینده ازن در مناطق ۲۲ گانه شهر تهران تحلیل و بررسی شدند. داده‌های بررسی شده مقادیر متوسط روزانه میزان غلظت O_3 برحسب ppb^{10} (تعداد در میلیون) مربوط به سال ۱۳۸۸ شمسی (موسوی و محمدزاده، ۲۰۱۱ و ۲۰۱۳) است که شرکت کنترل کیفیت هوای شهر تهران در ایستگاه‌های سنجش آلودگی هوا اندازه‌گیری و ثبت کرده است. از آنجا که توزیع داده‌های ازن بر اساس آزمون اندرسون-دارلینگ (اندرسون و دارلینگ، ۱۹۵۴)، با p -مقدار $2/2 \times 10^{-16}$ نرمال نیست، ولی جذر داده‌ها استفاده شد که فرض نرمال بودن آن‌ها با p -مقدار $1/68$ پذیرفته شد. برای پیش‌گویی بر اساس مدل تلفیق بیزی داده‌ها از دو منبع اطلاعاتی جمع‌آوری می‌شوند. منبع اول اندازه‌های غلظت ازن ایستگاه‌های مانیتورینگ در سطوح منطقه‌ای و منبع دوم خروجی‌های مدل کریکینگ است. به هر یک از ۲۲ ناحیه شهر تهران مدل کریکینگ برازش داده شده و از خروجی‌های این مدل به‌عنوان خروجی‌های مدل قطعی در روش تلفیق بیزی استفاده شده است. در حقیقت اندازه‌ها و خروجی مدل تکیه‌گاه‌های متفاوتی دارند، یعنی، اندازه‌ها مقادیر نقطه‌ای هستند در حالی که خروجی‌های مدل متوسط توری‌ها در شهر تهران هستند.



شکل ۵. موقعیت ایستگاه‌های مانیتورینگ (•)،
نقاط نمونه‌گیری (×) و ایستگاه‌های فاقد مشاهده (Δ)

سطح غلظت ازن در شب و نزدیک به صبح نسبت به روز خیلی کمتر است. برای تحلیل میانگین سطوح غلظت ازن به‌صورت روزانه و هفتگی از مدل تلفیق بیزی استفاده شده است. سطوح ازن در ۲۲ ناحیه و اندازه‌ها در ۹ ایستگاه برای برازش مدل تلفیق بیزی استفاده می‌شود. اندازه‌ها در ۱۲ ایستگاه باقی‌مانده برای ارزیابی داده‌ها به‌عنوان ایستگاه‌های فاقد مشاهده استفاده می‌شوند. شکل ۵ نقشه ایستگاه‌های مانیتورینگ و فاقد مشاهده و ناحیه‌ها را نشان می‌دهد که در آن (•) ایستگاه‌های مانیتورینگ، (Δ) ایستگاه‌های بدون مشاهده و (×) نقاط نمونه‌گیری در هر ناحیه است. مقدار جذر میانگین توان‌های دوم خطای پیش‌گو در زمان t ، که می‌تواند ساعت، روز یا هفته باشد به‌صورت

10. Paets per billion

جدول ۲. مجموع توان‌های دوم خطای پیش‌گویی با کریگیدن و تلفیق بیزی برای تعداد سلول‌های مختلف

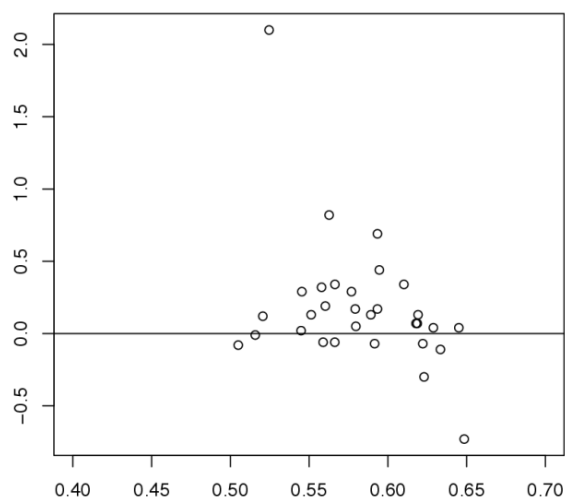
مجموعه داده‌ها	کریگیدن	زمان محاسبه کریگیدن	تلفیق بیزی				زمان محاسبه تلفیق بیزی	
			تعداد سلول	۱۰	۲۰	۴۰		۶۰
۱	۱/۷۸	۵	۱/۳۸	۱/۰۳	۰/۹۹	۰/۷۷	۰/۸۸	۱۶
۲	۱/۶۴	۴	۱/۰۶	۰/۹۲	۰/۸۸	۰/۶۶	۰/۶۶	۱۵
۳	۱/۷۰	۴	۱/۴۰	۱/۳۰	۱/۰۸	۰/۹۶	۰/۶۹	۱۷
۴	۲/۱۷	۷	۱/۱۸	۱/۰۸	۰/۹۹	۰/۶۸	۰/۹۱	۱۶
۵	۱/۰۷	۶	۱/۳۴	۱/۰۷	۰/۹۷	۰/۶۷	۱/۰۴	۱۶
۶	۳/۳۴	۴	۱/۵۷	۱/۰۱	۱/۰۳	۱/۲۴	۰/۹۷	۲۰
۷	۱/۳۵	۴	۰/۸۱	۰/۷۷	۰/۹۲	۰/۷۹	۰/۷۴	۱۴
۸	۱/۱۲	۶	۱/۳۹	۱/۵۵	۰/۹۲	۰/۸۸	۰/۸۱	۱۲
۹	۸/۹۱	۶	۱/۳۱	۱/۳۲	۰/۸۶	۰/۹۷	۰/۸۱	۱۹
۱۰	۲/۱۳	۸	۱/۳۸	۱/۲۷	۱/۰۶	۱/۰۹	۱/۲۸	۱۸
۱۱	۱/۵۶	۳	۰/۹۸	۱/۰۹	۰/۹۲	۰/۸۴	۰/۸۴	۱۸
۱۲	۱/۶۵	۴	۰/۸۴	۰/۹۸	۰/۸۵	۰/۹۵	۱/۰۹	۱۴
۱۳	۱/۵۱	۴	۰/۹	۰/۸۴	۱/۰۲	۰/۹۲	۰/۹۲	۱۵
۱۴	۱/۶۲	۳	۱/۰۰	۰/۹	۰/۹۸	۱/۰۲	۰/۸۱	۱۹
۱۵	۱/۶۴	۵	۱/۳۴	۱/۴۱	۱/۰۱	۱/۲۹	۱/۲۵	۱۹
۱۶	۱/۶۰	۷	۰/۹۹	۱/۳۳	۱/۸۸	۰/۶۵	۰/۷۱	۲۲
۱۷	۱/۶۶	۴	۱/۲۰	۰/۹	۱/۰۷	۰/۹۸	۱/۱۳	۲۰
۱۸	۱/۹۲	۳	۱/۰۴	۰/۸۵	۰/۸	۰/۹۳	۰/۹	۱۹
۱۹	۱/۶۰	۸	۱/۱۴	۰/۹۱	۱/۰۹	۱/۰۱	۰/۸۶	۱۴
۲۰	۲/۱۱	۷	۱/۰۶	۰/۹۱	۰/۷۳	۰/۸۳	۰/۹۲	۱۵
۲۱	۱/۵۰	۹	۱/۱۱	۱/۱۱	۱/۲۳	۱/۰۱	۰/۶۷	۱۸
۲۲	۳/۳۷	۷	۱/۲۶	۰/۹۱	۰/۸۴	۱/۲۰	۱/۱۰	۱۴
۲۳	۶/۵۱	۵	۱/۱۱	۱/۰۳	۰/۸۹	۰/۸۰	۰/۸۷	۱۶
۲۴	۱/۶۶	۴	۰/۹۱	۱/۲۱	۱/۰۳	۱/۸۸	۰/۸۰	۱۶
۲۵	۱/۵۳	۵	۱/۰۸	۱/۰۲	۱/۱۰	۱/۸۵	۰/۸۰	۱۸
۲۶	۲/۱۲	۴	۰/۹۷	۰/۶۵	۰/۷۴	۱/۰۵	۰/۹۹	۱۵
۲۷	۱/۴۳	۸	۱/۱۸	۰/۹۸	۰/۸۷	۰/۸۴	۱/۰۰	۱۸
۲۸	۳/۷۶	۵	۱/۲۵	۱/۰۲	۱/۰۲	۰/۷۶	۰/۶۰	۱۳
۲۹	۱/۵۳	۳	۱/۰۴	۱/۰۲	۱/۱۲	۱/۰۸	۱/۱۰	۱۹
۳۰	۱/۹۶	۶	۱/۱۶	۰/۹۵	۰/۹۲	۰/۹۷	۰/۸۵	۲۲
۳۱	۲/۹۴	۸	۰/۸۵	۱/۲۷	۱/۱۳	۱/۲۹	۱/۸۶	۱۷
۳۲	۶/۳۴	۴	۰/۹۴	۱/۵۶	۰/۹۸	۱/۱۸	۰/۷۳	۱۸
۳۳	۲/۲۳	۹	۰/۷۲	۱/۲۱	۱/۱۰	۰/۸۵	۰/۷۵	۱۸
۳۴	۲/۵۸	۸	۰/۹۵	۱/۲۷	۱/۸۷	۱/۶۲	۱/۰۹	۱۴
۳۵	۲/۱۶	۷	۰/۹۴	۰/۷۸	۱/۳۰	۰/۷۸	۰/۶۶	۱۶
۳۶	۲/۰۳	۵	۱/۰۳	۰/۸۳	۰/۷۴	۰/۸۴	۰/۸۰	۱۷
۳۷	۲/۳۹	۴	۱/۲۲	۰/۸۴	۰/۳۹	۱/۲۱	۰/۶۸	۱۵
۳۸	۲/۰۶	۳	۰/۸۹	۰/۶۲	۰/۶۷	۰/۷۳	۰/۷۶	۱۴
۳۹	۱/۷۰	۳	۰/۶۵	۰/۹۶	۱/۱۱	۰/۸۵	۰/۶۶	۱۷
۴۰	۱/۷۹	۴	۰/۹۶	۰/۹۰	۰/۸۳	۰/۸۱	۰/۷۸	۱۹
۴۱	۱/۴۵	۵	۱/۰۱	۰/۷۴	۰/۶۹	۰/۷۸	۰/۷۱	۱۹
۴۲	۱/۲۸	۵	۱/۰۲	۱/۰۰	۰/۸۳	۰/۸۵	۰/۸۰	۱۷
۴۳	۲/۲۰	۴	۱/۱۲	۰/۶۵	۰/۸۶	۰/۶۴	۰/۸۸	۱۸
۴۴	۱/۵۱	۳	۱/۰۸	۰/۸۷	۰/۸۵	۰/۸۶	۰/۷۹	۱۷
۴۵	۱/۶۰	۸	۱/۳۹	۰/۶۷	۱/۰۰	۰/۸۳	۰/۸۰	۱۶
۴۶	۱/۲۰	۷	۱/۳۱	۰/۸۳	۰/۸۰	۰/۷۹	۰/۶۵	۱۸
۴۷	۱/۲۸	۵	۱/۰۰	۰/۹۶	۰/۹۴	۰/۷۶	۰/۶۹	۱۹
۴۸	۱/۳۵	۴	۱/۲۱	۰/۸۸	۱/۰۲	۰/۸۳	۰/۸۵	۱۷
۴۹	۱/۶۶	۶	۱/۰۷	۰/۹۰	۰/۷۰	۰/۷۹	۰/۸۵	۱۶
۵۰	۲/۳۰	۸	۱/۵۱	۱/۰۱	۱/۳۲	۱/۲۷	۰/۸۱	۱۶
میانگین	۲/۲۰	۵/۴۲	۱/۰۹	۱/۰۱	۰/۹۴	۰/۹۱	۰/۸۵	۱۶/۹

$$\text{RMSPE} = \left[\frac{1}{n} \sum_{i=1}^n (O_i - \hat{O}_i)^2 \right]^{\frac{1}{2}}$$

محاسبه و عملکرد پیش‌گویی‌ها را اندازه‌گیری می‌کند، که در آن n تعداد ایستگاه‌های فاقد مشاهده برای پیش‌گویی و O_i اندازه در ایستگاه i ام و $\hat{O}_i$ مقدار پیش‌گویی است.

تحلیل میانگین روزانه داده‌ها

برای ۳۱ روز خرداد سال ۱۳۸۸، از میانگین روزانه اندازه‌ها و خروجی مدل‌های قطعی استفاده شده است. چنان‌که در جدول ۳ ملاحظه می‌شود در تحلیل میانگین روزانه جذر میانگین توان‌های دوم خطای پیش‌گویی با مدل تلفیق بیزی نسبت به دیگر روش‌های پیش‌گویی کریگیدن کوچک‌تر است. در این جدول تعداد روزهایی که روش تلفیق بیزی عملکرد بهتری نسبت به کریگیدن دارد بیش‌تر است که با علامت * مشخص شده‌اند و همچنین تعداد روزهایی که کریگیدن با اندازه‌ها استفاده شده بیش‌تر از تعداد روزهایی است که کریگیدن با خروجی‌های مدل استفاده شده است. بنا بر این پیش‌گویی بر اساس تلفیق بیزی بهتر از کریگیدن است که فقط با استفاده از اندازه‌ها محاسبه می‌شود. مقدار میانگین جذر میانگین توان‌های دوم خطا برای روش تلفیق بیزی ۱/۷۱، برای کریگیدن با استفاده از اندازه‌ها ۱/۸۰ و برای کریگیدن با استفاده از خروجی‌های مدل ۲/۳۵ محاسبه شده است. کریگیدن با استفاده از خروجی مدل میانگین جذر میانگین توان‌های دوم خطای بزرگ‌تری دارد و بیان‌گر این است که خروجی‌های مدل باید کالیبره شوند.



شکل ۶. تفاوت میانگین RMSPE پیش‌گویی‌های کریگیدن با استفاده از اندازه‌ها و مدل تلفیق بیزی در مقابل همبستگی بین اندازه‌ها و خروجی‌های مدل

شکل ۶ تفاوت جذر توان‌های دوم خطای بین پیش‌گویی‌های مدل تلفیق بیزی و کریگیدن با استفاده از اندازه‌ها در مقابل همبستگی بین اندازه‌ها و خروجی‌های مدل را نشان می‌دهد. ملاحظه می‌شود که عملکرد پیش‌گویی‌های مدل تلفیق بیزی نسبت به کریگیدن با اندازه‌ها و کریگیدن با خروجی‌های مدل بهتر است، چون با استفاده از تلفیق بیزی همبستگی بین اندازه‌ها و خروجی‌های مدل در حال افزایش است و همبستگی برای تمام روزها بیش‌تر از ۰/۵ است.

جدول ۳. میانگین روزانه RMSPE پیش‌گویی ازن در ۱۲ ایستگاه

روز	تلفیق بیزی	کریگین با اندازه‌ها	کریگین با خروجی‌های مد
۱	۱/۵۸*	۱/۷۵	۲/۶۷
۲	۱/۸۶	۱/۱۳*	۱/۸۸
۳	۱/۹۲*	۲/۰۵	۲/۱۶
۴	۱/۶۳	۱/۳۳*	۲/۲۸
۵	۲/۳۹*	۲/۴۳	۲/۴۰
۶	۱/۸۶*	۲/۲۰	۳/۲۰
۷	۱/۸۰	۱/۷۳*	۲/۵۵
۸	۱/۸۱*	۲/۲۵	۲/۴۱
۹	۲/۱۰*	۲/۹۲	۲/۷۴
۱۰	۱/۲۲*	۱/۳۵	۱/۳۹
۱۱	۱/۴۴	۱/۵۶	۱/۰۱*
۱۲	۱/۵۸*	۱/۷۷	۲/۷۷
۱۳	۱/۶۹*	۱/۸۶	۳/۳۵
۱۴	۱/۶۱*	۱/۶۵	۱/۶۲
۱۵	۱/۳۰	۱/۲۴*	۱/۴۹
۱۶	۲/۱۴	۲/۰۷*	۳/۳۲
۱۷	۱/۹۲	۱/۹۱*	۴/۴۸
۱۸	۱/۹۲*	۱/۹۹	۲/۰۵
۱۹	۱/۸۷	۱/۸۱*	۲/۵۲
۲۰	۲/۰۸	۲/۰۰*	۲/۵۰
۲۱	۱/۷۲	۱/۱۵	۱/۰۷*
۲۲	۱/۶۹*	۱/۷۱	۲/۹۲
۲۳	۱/۸۵*	۲/۵۴	۲/۹۶
۲۴	۱/۱۰*	۱/۴۲	۲/۸۴
۲۵	۱/۴۷*	۱/۷۶	۳/۶۸
۲۶	۱/۶۷	۱/۵۶*	۱/۸۸
۲۷	۱/۲۸*	۱/۴۱	۱/۹۴
۲۸	۱/۷۵*	۱/۸۰	۱/۹۱
۲۹	۱/۴۵	۱/۸۶	۱/۰۱*
۳۰	۲/۲۲*	۲/۲۹	۲/۳۹
۳۱	۱/۱۱*	۱/۴۰	۲/۰۹
میانگین	۱/۷۱*	۱/۸۰	۲/۳۵

تحلیل میانگین هفتگی داده‌ها

ایستگاه‌های مانیتورینگ، ایستگاه‌های فاقد مشاهده و سلول‌های شبکه‌ای استفاده شده در تحلیل هفتگی داده‌ها همان ایستگاه‌ها و سلول‌های شبکه‌ای استفاده شده در تحلیل روزانه داده‌ها هستند. به‌طور متوسط در مقیاس‌های زمانی طولانی‌تر، انتظار می‌رود که پیش‌گویی‌های هر دو مدل تلفیق بیزی و کریگین بهبود پیدا کنند، چون توزیع داده‌های تبدیل‌یافته نرمال است و تغییرپذیری میانگین داده‌ها در مقیاس‌های زمانی طولانی‌تر، کمتر است. بهبود پیش‌گویی‌های مدل تلفیق بیزی همچنین منجر به عملکرد بهتر مدل کریگین در زمان‌های طولانی‌تر می‌شود. جدول ۴، میانگین هفتگی جذر میانگین توان‌های دوم خطای پیش‌گویی مدل تلفیق بیزی، کریگین با استفاده از اندازه‌ها، کریگین با استفاده از خروجی مدل قطعی را نشان می‌دهد.

جدول ۴. میانگین هفتگی RMSPE پیش‌گویی ازن در ۱۲ ایستگاه

هفته	تلفیق بیزی	کریگین با اندازه‌ها	کریگین با خروجی‌های مد
اول	۱/۵۷	۱/۷۵	۱/۵۵
دوم	۱/۴۱*	۱/۸۵	۳/۶۳
سوم	۱/۵۵*	۱/۶۵	۱/۷۲
چهارم	۱/۰۵	۱/۰۳	۱/۶۹
میانگین	۱/۳۹	۱/۵۷	۲/۱۴

در تحلیل میانگین هفتگی داده‌ها با توجه به جدول ۴، مدل تلفیق بیزی در هفته دوم و سوم، میانگین جذر

میانگین توان‌های دوم خطای کوچک‌تری دارد که به‌ترتیب برابر $1/41$ و $1/55$ است. میانگین جذر میانگین توان‌های دوم خطای کل هفته‌ها نیز برای مدل تلفیق بیزی از همه کمتر و برابر $1/39$ است. چنان‌که انتظار داشتیم، میانگین جذر میانگین توان‌های دوم خطای همه پیش‌گویی‌ها در تحلیل میانگین هفتگی داده‌ها کمتر از میانگین جذر میانگین توان‌های دوم خطای پیش‌گویی‌ها در تحلیل میانگین روزانه داده‌ها است. میانگین جذر میانگین توان‌های دوم خطای مدل تلفیق بیزی در جداول ۳ و ۴ بیان‌گر آن است که به‌طور متوسط پیش‌گویی‌های مدل تلفیق بیزی با افزایش مقیاس زمانی بهبود یافته‌اند، یعنی در مقیاس‌های زمانی طولانی، خروجی‌های مدل، پیش‌گویی‌های بهتری از ایستگاه‌ها ارائه می‌دهند.

بحث و نتیجه‌گیری

معمولاً در تحلیل مدل تلفیق بیزی، توزیع پسینی بعضی از پارامترها فرم بسته‌ای ندارند و استفاده از روش‌های MCMC بسیار زمان‌بر هستند. در مطالعه شبیه‌سازی مشخص شد این روش در عین حالی که دقیق‌تر از روش کریگیدن عمل می‌کند، اما محاسبات پرهزینه‌تری دارد. تعداد نقاط نمونه‌گیری در درون توری‌های شبکه‌ای روی بعد ماتریس همبستگی فضایی تأثیر دارد. تلفیق پارامترهای کوواریانس و در شرایطی خطاهای استانداردشان را به خوبی برآورد نمی‌کند، مگر این‌که تعداد محسوسی از نقاط نمونه‌گیری خیلی نزدیک به هم باشند. فرض نرمال بودن در مدل تلفیق بیزی می‌تواند به‌عنوان مسئله‌ای پژوهشی مطرح شود. تلفیق، پیش‌گویی ایستگاه‌های بدون اندازه، ارزیابی خروجی مدل قطعی، تشخیص روند فضایی و برآورد همبستگی فضایی را ممکن می‌سازد. این مدل عدم قطعیت پیش‌گویی‌ها را بهتر از کریگیدن لحاظ می‌کند. با وجود این‌که مدل تلفیق بیزی برای ترکیب اندازه‌ها با خروجی مدل‌های قطعی استفاده می‌شود، به‌عنوان یک مدل فضایی طراحی شده است و نمی‌تواند همبستگی زمانی داده‌ها را بررسی کند. برای این‌گونه داده‌ها لازم است مدل‌های فضایی-زمانی جای‌گزین شود.

تشکر و قدردانی

بدین‌وسیله از داوران محترم مجله که نظرات ارزنده آن‌ها موجب بهبود کیفیت مقاله شد و از حمایت قطب علمی داده‌های ترتیبی و فضایی دانشگاه فردوسی مشهد قدردانی می‌کنیم.

منابع

۱. موسوی س.، ارزیابی روش‌های مختلف مدل‌بندی روند فضایی زمانی بر برآورد تابع کوواریانس، پایان‌نامه کارشناسی ارشد، گروه آمار، دانشگاه تربیت مدرس (۱۳۹۰).
2. Banerjee S., Carlin B.P. Gelfand A.E., "Hierarchical Modeling and Analysis for Spatial Data", Chapman and Hall/CRC (2004).
3. Berrocal V.J., Gelfand A.E., Holland, D.M., "A Spatio-Temporal Downscaler for Output from Numerical Models", Journal of Agricultural, Biological and Environmental Statistics, 15 (2010) 176-197.
4. Berliner L., "Physical-Statistical Modeling in Geophysics", Journal of Geophysical

- Research, 108 (2003) 80-91.
5. Fuentes M., Raftery A., "Model Validation and Spatial Interpolation by Bayesian Combination of Observations with Outputs from Numerical Models", *Biometrics*, 61 (2005) 36-45.
 6. Fuentes M., Guttorp P., Challenor P., "Statistical Assessment of Numerical Models, *International Statistical review*, 71 (2003) 201-221.
 7. Fuentes M., Raftery A., "Model Validation and Spatial Interpolation by Combining Observations with Outputs from Numerical Models via Bayesian Melding", *Technical Report 403*, Department of Statistics, University of Washington.
 7. Gelfand A., Smith A., "Sampling-Based Approaches to Calculating Marginal Densities", *Journal of American Statistical Association*, 85 (1990) 398-409.
 9. Hastings W., "Monte Carlo Sampling Method Using Markov chains and their Application", *Biometrika*, 57 (1970) 97-109.
 10. Hogrefe C. Rao S., Kasibhatla P., Hao W., Sistla G. R. M., McHenry J., "Evaluating the performance of regional-Scale Photochemical Modeling Systems", Part II-ozone Predictions, *Atmos, Environ*, 35 (2001a) 4175-4188.
 11. Hogrefe C., Rao S., Kasibhatla P., Kallos G., Trembagk C., Hao W., Olerud D., Xiu A., McHenry J., Alapaty K., "Evaluating the performance of Regional-Scale Photochemical Modeling Systems", part I-meteorological Predictions, *Atmos. Environ*, 35 (2001b) 4150-4174.
 12. Matheron G., "Traite de Geostatistique Appliquee, Technical Report", *Memoires du Bureau de Recerches Geologiques et Minieres*.
 13. Mousavi S.S., Mohammadzadeh M., "Spatial-Temporal Trend Modeling for Ozone Concentration in Tehran City", *Journal of Statistical Research of Iran*, 8 (2011) 149-163.
 14. Mousavi S.S., Mohammadzadeh M., "Determination of Spatial-Temporal Correlation Structure of Troposphere Ozone Data in Tehran City", *Journal of Sciences Islamic Republic of Iran*, 24(2013) 171-178.
 15. Poole D., Raftery A., "Inference for Deterministic Simulation Models, the Bayesian Melding Approach, *Journal of American Statistical Association*, 95 (2000) 1244-1255.
 16. Sanso B., Guenni L., "Combining Observed Rainfall and Deterministic prediction Using a Bayesian Approach, *Technical Report*", CESMA (2002).