تلفیق بیزی مدل‌های قطعی و کریگیدن برای تحلیل داده‌های وابستهٔ فضایی

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

دریافت ۹۱/۱۱/۷ پذیرش ۹۲/٤/۹

چکیده

برقراری پیوند میان سامانه اطلاعات مکانی و روش‌های تصمیم‌گیری مرهون ابداع و تکامل روش تلفیق داده‌های مکانی است. روش‌های تلفیق اطلاعات، داده‌ها را با هدف دست‌یابی به نتیجهٔ بهتر برای افزایش قابلیت تفسیر اطلاعات با یك‌دیگر 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 اُزن شهر تهران تحلیل شد و با ملاك میانگین توان‌های دوم خطا نتایج حاصل از آن‌ها ارزیابی و مقایسه شدند.

واژه‌های کلیدی: تلفیق بیزی، مدل قطعی، کریگیدن.

مقدمه

منظور از تلفیق داده‌ها[1]، 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[2] 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 اُزن روی توری‌های مشبکه‌ای[3] را شبیه‌سازی می‌کنند و اندازه‌ها[4] در هر یك از ایستگاه‌های واقع در منطقهٔ 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 براین، می‌توان خروجی مدل‌های قطعی را برای پیش‌گویی کالیبره[5] کرد.

پوول و رفتری (۲۰۰۰) از ترکیب پیشین‌های مختلف برای کاهش عدم قطعیت خروجی‌های یك مدل بیزی استفاده کردند. کسیبتلا و چمید (۲۰۰۰) و هاگرف و همکاران (a,b۲۰۰۱) 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

[Downloaded from c4i2016.khu.ac.ir on 2026-11-04]

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

کریگیدن (ماترون، ۱۹٦۳) روشی برای پیش‌گویی فضایی در موقعیت‌های مشخص و بهترین پیش‌گوی خطی ناااریب است. هر چند محاسبات کریگیدن سرراست 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 پوول و رفتری (۲۰۰۰) تلفیق بیزی را پیشنهاد کردند. سپس سانسو و گوین (۲۰۰۲) و فوئنتز و رفتری (۲۰۰۵) آن را توسعه دادند. برلینر (۲۰۰۳) و برلینر و ویکال (۲۰۰۵) نیز نحوهٔ استفاده از مدل‌های تلفیق بیزی برای ترکیب اطلاعات از منابع مختلف را بیان کرده‌اند. مدل تلفیق بیزی، سطح واقعی پنهان یك 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 یك توری درشت‌تر از توری اندازه‌های اصلی هستند. به‌عنوان مثال داده‌های ساعتی غلظت اُزن در ناحیهٔ $[٦ \times ٦]$ 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، فوئنتز و رفتری (۲۰۰۲ و ۲۰۰۵) و فوئنتز و همکاران (۲۰۰۳) مدل تلفیق را مطرح کردند. در این مقاله مدل تلفیق بیزی و نحوهٔ پیش‌گویی با آن معرفی می‌شود. سپس نحوهٔ کاربست مدل تلفیق بیزی برای تحلیل داده‌های آلاینده اُزن شهر تهران پیشنهاد و دقت نتایج حاصل از دو روش کریگیدن و تلفیق بیزی تحلیل و ارزیابی شدند.

مدل تلفیق بیزی

فرض کنید ناحیهٔ بررسی شده $D \subset R^d$، $d \geq 1$ را مشبکه کرده و هر یك 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 فرآیند واقعی با $\{Z(s): s \in D \subset R^d\}$، فرآیند 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 فرآیندهای خروجی مدل‌های قطعی $\{\tilde{Z}(s): s \in D\}$ موجود و به موقعیت‌های s وابسته‌اند. هدف مرتبط ساختن خروجی مدل‌های قطعی و اندازه‌ها با مقدار واقعی $Z(s)$ در مشبکه‌های کوچك مقیاس از طریق $\hat{Z}(s)$ و $\tilde{Z}(s)$ است. بنا بر این فرآیندهای واقعی فصل مشترك دو فرآیند $\hat{Z}(s)$ و $\tilde{Z}(s)$ هستند، که لزوماً مقیاس یك‌سانی ندارند. برای غلبه بر مشکل عدم تطابق مقیاس‌ها، می‌توان $\hat{Z}(s)$ و $\tilde{Z}(s)$ را از طریق فرآیندهای 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

[Downloaded from c4i2016.khu.ac.ir on 2026-11-04]

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

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

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:

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

که در آن $a(s)$ اریبی جمعی[6] و $b(s)$ اریبی ضربی[7] نامیده می‌شوند و $\delta(s)$ فرآیند خطا با توزیع $N \sim (0, \sigma_\delta^2)$ است. فرض کنید خروجی مدل قطعی در مشبکهٔ $B_1, \dots, B_J$ فراهم شده است، فرآیندهای در سطح نقطه با استفاده از (۱) در مدل (۲) بدین‌صورت به سطح مشبکه تبدیل می‌شوند:

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

اگر برای $\hat{Z}$ و $\tilde{Z}$ توزیع نرمال چندمتغیره در نظر گرفته شود آن‌گاه، پیش‌گویی فضایی فرایند 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

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

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 ، ۲۰۰٤) برای محاسبهٔ آن‌ها استفاده می‌شود. اکنون می‌توان مؤلفه‌های مدل تلفیق بیزی را بدین‌صورت خلاصه کرد:

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

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

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

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

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

$$\mu(s) = X(s)\beta \qquad (۴)$$

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 $\boldsymbol{\Sigma_\theta}$ هستند، که درآن $\boldsymbol{\theta}$ بردار پارامتر کوواریانس است. هر یك از $Z(B)$ و $\tilde{Z}(B)$ انتگرال‌هایی بر حسب $Z(B)$ و $\tilde{Z}(B)$ روی سلول B هستند و درعمل فقط فرآیندهای

6. Additive
7. Multiplicative

[Downloaded from c4i2016.khu.ac.ir on 2026-11-04]

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

میانگین فرآیند واقعی به‌صورت $\mu(s) = X(s)\beta$ 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. فوئنتز و رفتری (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[8] نامیده می‌شوند. سپس مقادیر فرآیندها محاسبه می‌شوند.آن‌گاه مقدار تقریبی $Z(B)$ به‌صورت $Z(B) \approx \frac{1}{L}\sum_{j=1}^{L} Z(s_{j,B})$ حاصل می‌شود.

برازش مدل تلفیق بیزی

در این بخش نحوهٔ برازش مدل تلفیق بیزی با استفاده از الگوریتم‌های نمونه‌گیری گیبز معرفی خواهد شد. اصولاً رهیافت بیزی با در نظر گرفتن توزیع‌های پیشین مناسب و تعیین توزیع‌های پسین پارامترها بر اساس داده‌ها، به‌دنبال لحاظ کردن عدم قطعیت برآورد پارامترها است. در این‌جا پارامترها شامل، واریانس‌های خطاهای تصادفی σ_δ^2 وσ_e^2، بردار ضرایب $\boldsymbol{\beta}$، فرآیندهای واقعی مورد نظر $\boldsymbol{Z}$، پارامترهای کالیبراسیون a و b و بردارپارامترهای کوواریانس $\boldsymbol{\theta}$ است. از بردار Z برای مشخص کردن مقادیر فرآیند واقعی مورد نظر در ایستگاه‌ها و نقاط نمونه‌گیری درون سلول مشبکه‌ای استفاده می‌شود. اگر تعداد ایستگاه‌ها n و تعداد سلول‌های مشبکه‌ای m باشد، آن‌گاه بعد بردار $\boldsymbol{Z}$ برابر $m + n \times L$ است. ابعاد اندازه‌های $\widehat{\boldsymbol{Z}}$ و خروجی مدل $\widetilde{\boldsymbol{Z}}$ نیز به‌ترتیب n و m هستند.

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

$$p(\boldsymbol{\theta}, \boldsymbol{\beta}, a, b, \sigma_\delta^2, \sigma_e^2, \boldsymbol{Z}, \widehat{\boldsymbol{Z}}, \widetilde{\boldsymbol{Z}}) = p(\widehat{\boldsymbol{Z}}|\boldsymbol{Z}, \sigma_e^2)p(\widetilde{\boldsymbol{Z}}|\boldsymbol{Z}, a, b, \sigma_e^2)p(\boldsymbol{Z}|\boldsymbol{\beta}, \boldsymbol{\theta})p(\sigma_\delta^2, \sigma_e^2\boldsymbol{\theta}, \boldsymbol{\beta})$$
$$= \phi_{\boldsymbol{\Sigma}_1}(\widehat{\boldsymbol{Z}} - \boldsymbol{A_0 Z})\phi_{\boldsymbol{\Sigma}_2}(\widetilde{\boldsymbol{Z}} - a - b\boldsymbol{A_1 Z})\phi_{\boldsymbol{\Sigma}_3}(\boldsymbol{Z} - \boldsymbol{X\beta})p(\sigma_\delta^2)p(\sigma_e^2)p(\boldsymbol{\beta})p(\boldsymbol{\theta})$$

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

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

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

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

8. Sampling points

[Downloaded from c4i2016.khu.ac.ir on 2026-11-04]

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

$$p\left(\boldsymbol{Z}, \widehat{\boldsymbol{Z}}, \widetilde{\boldsymbol{Z}} | \boldsymbol{\theta}, \boldsymbol{\beta}, a, b, \sigma_\delta^2, \sigma_e^2\right)$$
$$\propto exp\{-\frac{1}{2}\big[-\boldsymbol{Z}^T\left(\boldsymbol{A_0^T\Sigma_1^{-1}}\widehat{\boldsymbol{Z}} + b\boldsymbol{A_1^T\Sigma_2^{-1}}\left(\widetilde{\boldsymbol{Z}} - a\right) + \boldsymbol{\Sigma_3^{-1}\mu}\right)$$
$$- \left(\widehat{\boldsymbol{Z}}^{\boldsymbol{T}}\boldsymbol{\Sigma_1^{-1}A_0} + b\left(\widetilde{\boldsymbol{Z}} - a\right)\boldsymbol{\Sigma_2^{-1}A_1} + \boldsymbol{\mu^T\Sigma_3^{-1}}\right)\boldsymbol{Z}$$
$$+ \boldsymbol{Z}^T\left(\boldsymbol{A_0^T\Sigma_1^{-1}A_0} + b^2\boldsymbol{A_1^T\Sigma_2^{-1}A_1} + \boldsymbol{\Sigma_3^{-1}}\right)\boldsymbol{Z}\big]\} + \boldsymbol{C}$$

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

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

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

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

که در آن $\boldsymbol{C}$ ماتریس فاصلهٔ اقلیدسی[9] بین ایستگاه‌ها و نقاط نمونه‌گیری و $g(.)$ تابع هم‌بستگی فضایی است. بنا بر این توزیع شرطی کامل $\boldsymbol{\theta}$ بدین‌صورت به‌دست می‌آید:

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

در نتیجه توزیع شرطی کامل σ ، گامای واون با پارامترهای

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

است. پارامتر ρ دارای توزیع شرطی کاملی که بتوان به‌طور 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 به‌ترتیب 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 به‌ترتیب بدین‌صورت هستند:

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

9. Euclidean distance

[Downloaded from c4i2016.khu.ac.ir on 2026-11-04]

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

که در آن‌ها $\tilde{\lambda}=\left(\widetilde{\boldsymbol{Z}}-a-b\boldsymbol{A_1Z}\right)^T\left(\widetilde{\boldsymbol{Z}}-a-b\boldsymbol{A_1Z}\right)$ و $\gamma=\left(\widehat{\boldsymbol{Z}}-\boldsymbol{A_0Z}\right)^T\left(\widehat{\boldsymbol{Z}}-\boldsymbol{A_0Z}\right)$. با در نظر گرفتن توزیع پیشین $(a,b)^{'}\sim MVN(\overline{\boldsymbol{\beta}},\overline{\boldsymbol{F}})$ توزیع شرطی توأم کامل پارامترهای کالیبراسیون به‌صورت $(a,b)^{'}|(\tilde{Z},A_1Z,\Sigma_2)\sim MVN(\overline{\boldsymbol{B}}\overline{\boldsymbol{C}},\overline{\boldsymbol{B}})$ حاصل می‌شوند، که در آن $\overline{\boldsymbol{B}}=\begin{pmatrix}\underline{\mathbf{1}}^{\boldsymbol{T}}\\(\boldsymbol{A_1Z})^{\boldsymbol{T}}\end{pmatrix}\boldsymbol{\Sigma_2^{-1}}(\underline{\mathbf{1}}\quad\boldsymbol{A_1Z})$، $\overline{\boldsymbol{C}}=\begin{pmatrix}\underline{\mathbf{1}}^{\boldsymbol{T}}\\(\boldsymbol{A_1Z})^{\boldsymbol{T}}\end{pmatrix}\boldsymbol{\Sigma_2^{-1}}(\underline{\mathbf{1}}\quad\boldsymbol{Z})+\overline{\boldsymbol{F}}^{-1}$ و بردار ستونی $\underline{1}$ دارای بعد مشابه بردار $\widetilde{\boldsymbol{Z}}$ است.

برای پیش‌گویی فرآیندهای مورد نظر در ایستگاه‌های فاقد مشاهده فرض کنید تحقق‌های فرآیندهای واقعی در ایستگاه‌های فاقد مشاهده با نماد $\boldsymbol{Z_u}$ نشان داده شوند. درون‌یابی فرآیندها مستلزم پیدا کردن توزیع پسین $(\boldsymbol{Z_u}|\boldsymbol{Z_g})$ است، که در آن $\boldsymbol{Z_g}$ تحقق‌های فرآیندهای واقعی در ایستگاه‌های مانیتورینگ و نقاط نمونه‌گیری در درون سلول‌های مشبکه است. ملاحظه می‌شود که $E(\boldsymbol{Z_u})=\boldsymbol{\mu_u}$ و $E(\boldsymbol{Z_g})=\boldsymbol{\mu_g}$. بنا بر این میانگین و واریانس توزیع شرطی $(\boldsymbol{Z_u}|\boldsymbol{Z_g},\boldsymbol{\beta},\boldsymbol{\theta})$ به‌ترتیب بدین‌صورتند:

$$E(\boldsymbol{Z_u}|\boldsymbol{Z_g},\boldsymbol{\beta},\boldsymbol{\theta})=\boldsymbol{\mu_u}+\boldsymbol{\Sigma_{gu}^{-1}\Sigma_g^{-1}}(\boldsymbol{Z_g}-\boldsymbol{\mu_g})$$
$$Var(\boldsymbol{Z_u}|\boldsymbol{Z_g},\boldsymbol{\beta},\boldsymbol{\theta})=\boldsymbol{\Sigma_u}-\boldsymbol{\Sigma_{gu}\Sigma_g^{-1}\Sigma_g}$$

که در آن $\boldsymbol{\Sigma_u}$ ماتریس کوواریانس $\boldsymbol{Z_u}$ و $\boldsymbol{\Sigma_{gu}}$ ماتریس کوواریانس بین $\boldsymbol{Z_u}$ و $\boldsymbol{Z_g}$ است. فرض کنید $\boldsymbol{\Sigma_{ug}}$ نشان دهنده ترانهاده $\boldsymbol{\Sigma_{gu}}$ باشد. با توجه به آن‌که در مدل تلفیق بیزی (٤) داریم $\widehat{\boldsymbol{Z}}(\boldsymbol{s})=\boldsymbol{Z}(\boldsymbol{s})+\boldsymbol{e}(\boldsymbol{s})$، میانگین و واریانس شرطی $(\widehat{\boldsymbol{Z}}_u|\boldsymbol{Z_g},\boldsymbol{\beta},\boldsymbol{\theta})$ به این‌صورت به‌دست می‌آید:

$$E(\widehat{\boldsymbol{Z}}_u|\boldsymbol{Z_g},\boldsymbol{\beta},\boldsymbol{\theta})=\boldsymbol{\mu_u}+\boldsymbol{\Sigma_{gu}^{-1}\Sigma_g^{-1}}(\boldsymbol{Z_g}-\boldsymbol{\mu_g})$$
$$Var(\widehat{\boldsymbol{Z}}_u|\boldsymbol{Z_g},\boldsymbol{\beta},\boldsymbol{\theta})=\boldsymbol{\Sigma_u}-\boldsymbol{\Sigma_{gu}\Sigma_g^{-1}\Sigma_g}+\sigma_e^2\boldsymbol{I}$$

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

مطالعه شبیه‌سازی

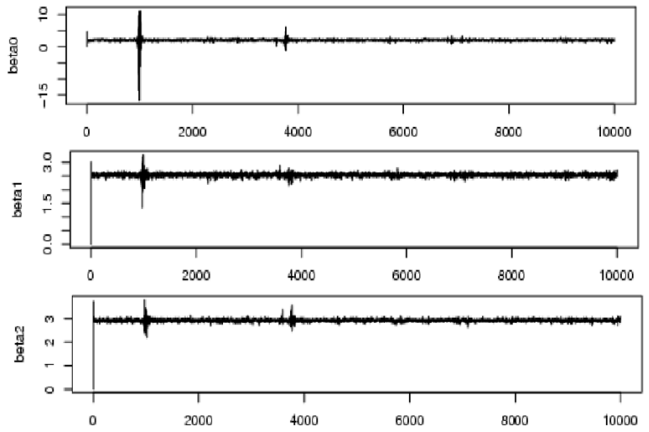
با فرض آن‌که فرآیندهای واقعی، مانا هستند، ابتدا تحلیل مدل تلفیق بیزی با استفاده از الگوریتم MCMC ارزیابی می‌شود. سپس دقت پیش‌گویی‌ها با دو روش کریگیدن و تلفیق بیزی با هم مقایسه می‌شوند. برای شبیه‌سازی، ۱۵ ایستگاه مانیتورینگ و ۸۰ ایستگاه برای پیش‌گویی و تعداد سلول‌های مشبکه ۲، ۱۰، ۲۰، ٤۰ و٦۰ در نظر گرفته شده است. برای کاهش محاسبات، از درون هر سلول مشبکه‌ای فقط یك 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 یك‌نواخت در ناحیهٔ $[٦\times٦]\times[-٦\times٦]$ تولید شده‌اند. برای هر یك از ترکیبات ایستگاه‌ها و نقاط نمونه‌گیری، ۵۰ مجموعه داده‌ی مستقل از اندازه‌ها و خروجی‌ها طبق مدل (٤) تولید شده‌اند. میانگین فرآیندهای واقعی به‌صورت

$$E[Z(\boldsymbol{s})]=\beta_0+\beta_1\boldsymbol{s}_1+\beta_2\boldsymbol{s}_2 \qquad (۸)$$

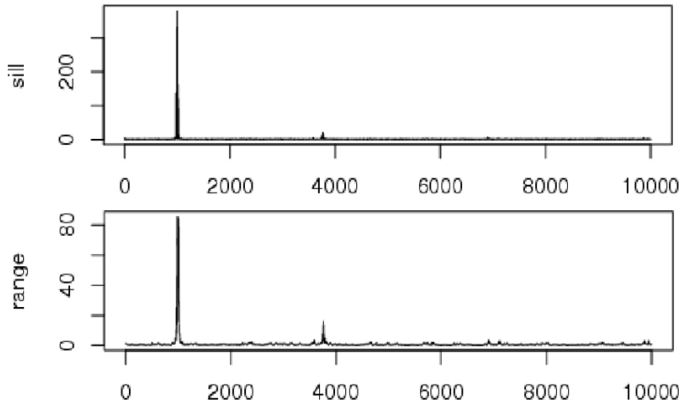
در نظر گرفته شده است، که در آن $\boldsymbol{s}_1$ و $\boldsymbol{s}_2$ مختصات موقعیت‌ها و $Z(\boldsymbol{s})$ دارای کوواریانس فضایی

[Downloaded from c4i2016.khu.ac.ir on 2026-11-04]

نمایی $C_{\boldsymbol{\theta}}(h) = \sigma\exp\{-|\mathrm{h}|/\rho\}$ 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

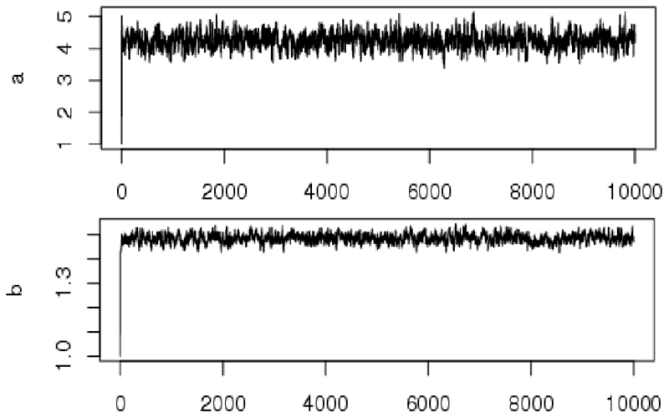


شکل ۲. نمودارهای همگرایی شبیه‌سازی پارامترهای کوواریانس

برآورد پارامترهای مدل تلفیق بیزی در جدول ۱ آورده شده است. ستون‌های ٤ تا ۸، متوسط میانگین‌ها و انحراف استاندارد توزیع‌های شرطی کامل پارامترهای مدل تلفیق بیزی هستند، که بر اساس۵۰ 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 نزدیك هستند و با افزایش تعداد سلول‌های مشبکه‌ای دقت برآوردها افزایش می‌یابد. برآورد پارامترهای کوواریانس $\boldsymbol{\theta} = (\sigma, \rho)$ تقریباً دقیق است. اما با این‌که میانگین پسین با افزایش تعداد سلول‌های مشبکه‌ای از ۲ تا ٦۰، به مقدار واقعی آن نزدیك می‌شود، انحراف استاندارد حتی با ٦۰ سلول مشبکه‌ای نسبتاً بزرگ باقی می‌ماند، که ممکن است ناشی از تعداد نقاط ناکافی نزدیك به یکدیگر باشد. به عبارت دیگر، از فرآیندهای کوچك 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 گریگیدن۲/۲۰ و ۵/٤۲ است

[Downloaded from c4i2016.khu.ac.ir on 2026-11-04]

در حالی‌که با افزایش تعداد سلول‌های مشبکه‌ای، 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 نسبت‌به کریگیدن دارد. در هریك از مجموعه داده‌ها، لزومی ندارد که با افزایش تعداد سلول‌های مشبکه‌ای، 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 ٤، سه نمودار مقدارهای واقعی اُزن، پیش‌گویی تلفیق بیزی و پیش‌گویی کریگیدن در حالتی که ٦۰ سلول مشبکه‌ای وجود دارد برای ۱۰ ایستگاه



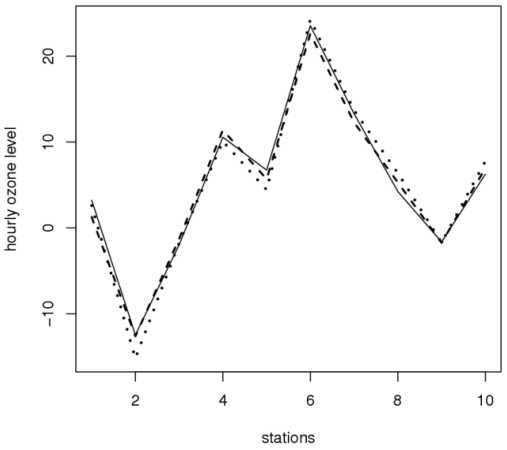
شکل ۳. نمودارهای همگرایی شبیه‌سازی پارامترهای کالیبراسیون

جدول ۱. برآورد پارامترهای مدل تلفیق بیزی

منبع	پارامتر	مقدار واقعی	تعداد سلول ۲	۱۰	۲۰	۴۰	۶۰
		۴/۰۰	۴/۳۰	۳/۹۲	۴/۰۱	۳/۹۹	۴/۰۲
کالیبراسیون			۲/۴۲	۰/۵۹	۰/۴۴	۰/۳۱	۰/۲۴
		۱/۵۰	۱/۵۶	۱/۵۱	۱/۴۹	۱/۴۹	۱/۴۹
			۰/۳۲	۰/۰۴	۰/۰۳	۰/۰۲	۰/۰۱
		۲/۲۰	۲/۱۷	۲/۲۰	۲/۲۱	۲/۲۱	۲/۱۹
			۰/۶۹	۰/۶۴	۰/۵۷	۰/۳۴	۰/۲۵
روند		۲/۵۰	۲/۴۴	۲/۴۸	۲/۴۹	۲/۵۰	۲/۴۹
			۰/۱۷	۰/۱۱	۰/۰۹	۰/۰۶	۰/۰۵
		۳/۰۰	۲/۹۳	۲/۹۷	۳/۰۰	۲/۹۸	۲/۹۹
			۰/۹۱	۰/۱۱	۰/۱۵	۰/۰۷	۰/۰۵
		۱/۲۰	۱/۶۸	۱/۵۸	۱/۷۳	۱/۲۱	۱/۲۹
کوواریانس			۲/۳۵	۳/۸۵	۳/۸۸	۱/۹۶	۲/۷۳
		۲/۵۰	۲/۵۷	۲/۲۰	۱/۸۸	۰/۹۵	۰/۷۳
			۲/۷۶	۲/۹۴	۲/۸۷	۱/۲۲	۱/۰۰

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

[Downloaded from c4i2016.khu.ac.ir on 2026-11-04]



شکل ٤. پیش‌گویی فضایی در ۱۰ ایستگاه فاقد مشاهده با مدل تلفیق بیزی بر اساس ٦۰ سلول مشبکه‌ای

مثال کاربردی

در این بخش داده‌های آلاینده اُزن 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 اُزن بر اساس آزمون اندرسون- دارلینگ (اندرسون و دارلینگ، ۱۹٥٤)، با p-مقدار $۲/۲\times۱۰^{-۱۶}$ نرمال نیست، ولی جذر داده‌ها استفاده شد که فرض نرمال بودن آن‌ها با p-مقدار ۱/٦۸ پذیرفته شد. برای پیش‌گویی بر اساس مدل تلفیق بیزی داده‌ها از دو منبع اطلاعاتی جمع‌آوری می‌شوند. منبع اول اندازه‌های غلظت اُزن ایستگاه‌های مانیتورینگ در سطوح منطقه‌ای و منبع دوم خروجی‌های مدل کریکینگ است. به هر یك 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 ٥. موقعیت ایستگاه‌های مانیتورینگ (•)، نقاط نمونه‌گیری (×) و ایستگاه‌های فاقد مشاهده (Δ)

سطح غلظت اُزن در شب و نزدیك به صبح نسبت به روز خیلی کم‌تر است. برای تحلیل میانگین سطوح غلظت اُزن به‌صورت روزانه و هفتگی از مدل تلفیق بیزی استفاده شده است. سطوح اُزن 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 ٥ 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 زمانt، که می‌تواند ساعت، روز یا هفته باشد به‌صورت

10. Paets per billion

[Downloaded from c4i2016.khu.ac.ir on 2026-11-04]

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					
			۲	۱۰	۲۰	٤۰	٦۰	
۱	۱/۷۸	٥	۱/۳۸	۱/۰۳	۰/۹۹	۰/۷۷	۰/۸۸	۱٦
۲	۱/٦٤	٤	۱/۰٦	۰/۹۲	۰/۸۸	۰/٦٦	۰/٦٦	۱٥
۳	۱/۷۰	٤	۱/٤۰	۱/۳۰	۱/۰۸	۰/۹٦	۰/٦۹	۱۷
٤	۲/۱۷	۷	۱/۱۸	۱/۰۸	۰/۹۹	۰/٦۸	۰/۹۱	۱٦
٥	۱/۰۷	٦	۱/۳٤	۱/۰۷	۰/۹۷	۰/٦۷	۱/۰٤	۱٦
٦	۳/۳٤	٤	۱/٥۷	۱/۰۱	۱/۰۳	۱/۲٤	۰/۹۷	۲۰
۷	۱/۳٥	٤	۰/۸۱	۰/۷۷	۰/۹۲	۰/۷۹	۰/۷٤	۱٤
۸	۱/۱۲	٦	۱/۳۹	۱/٥٥	۰/۹۲	۰/۸۸	۰/۸۱	۱۲
۹	۸/۹۱	٦	۱/۳۱	۱/۳۲	۰/۸٦	۰/۹۷	۰/۸۱	۱۹
۱۰	۲/۱۳	۸	۱/۳۸	۱/۲۷	۱/۰٦	۱/۰۹	۱/۲۸	۱۸
۱۱	۱/٥٦	۳	۰/۹۸	۱/۰۹	۰/۹۲	۰/۸٤	۰/۸٤	۱۸
۱۲	۱/٦٥	٤	۰/۸٤	۰/۹۸	۰/۸٥	۰/۹٥	۱/۰۹	۱٤
۱۳	۱/٥۱	٤	۰/۹	۰/۸٤	۱/۰۲	۰/۹۲	۰/۹۲	۱٥
۱٤	۱/٦۲	۳	۱/۰۰	۰/۹	۰/۹۸	۱/۰۲	۰/۸۱	۱۹
۱٥	۱/٦٤	٥	۱/۳٤	۱/٤۱	۱/۰۱	۱/۲۹	۱/۲٥	۱۹
۱٦	۱/٦۰	۷	۰/۹۹	۱/۳۳	۰/۸۸	۰/٦٥	۰/۷۱	۲۲
۱۷	۱/٦٦	٤	۱/۲۰	۰/۹	۱/۰۷	۰/۹۸	۱/۱۳	۲۰
۱۸	۱/۹۲	۳	۱/۰٤	۰/۸٥	۰/۸	۰/۹۳	۰/۹	۱۹
۱۹	۱/٦۰	۸	۱/۱٤	۰/۹۱	۱/۰۹	۱/۰۱	۰/۸٦	۱٤
۲۰	۲/۱۱	۷	۱/۰٦	۰/۹۱	۰/۷۳	۰/۸۳	۰/۹۲	۱٥
۲۱	۱/٥۰	۹	۱/۱۱	۱/۱۱	۱/۲۳	۱/۰۱	۰/٦۷	۱۸
۲۲	۳/۳۷	۷	۱/۲٦	۰/۹۱	۰/۸٤	۱/۲۰	۱/۱۰	۱٤
۲۳	٦/٥۱	٥	۱/۱۱	۱/۰۳	۰/۸۹	۰/۸۰	۰/۸۷	۱٦
۲٤	۱/٦٦	٤	۰/۹۱	۱/۲۱	۱/۰۳	۰/۸۸	۰/۸۰	۱٦
۲٥	۱/٥۳	٥	۱/۰۸	۱/۱۲	۱/۱۰	۰/۸٥	۰/۸۰	۱۸
۲٦	۲/۱۲	٤	۰/۹۷	۰/٦٥	۰/۷٤	۱/۰٥	۰/۹۹	۱٥
۲۷	۱/٤۳	۸	۱/۱۸	۰/۹۸	۰/۸۷	۰/۸٤	۱/۰۰	۱۸
۲۸	۳/۷٦	٥	۱/۲٥	۱/۰۲	۰/۷٦	۰/٥۰	۰/٦۰	۱۳
۲۹	۱/٥۳	۳	۱/۰٤	۱/۱۲	۱/۱۲	۱/۰۸	۱/۱۰	۱۹
۳۰	۱/۹٦	٦	۱/۱٦	۰/۹٥	۱/۰۲	۰/۹۷	۰/۸٥	۲۲
۳۱	۲/۹٤	۸	۰/۸٥	۱/۲۷	۱/۱۳	۱/۲۹	۰/۸٦	۱۷
۳۲	٦/۳٤	٤	۰/۹٤	۱/٥٦	۰/۹۸	۱/۱۸	۰/۷۳	۱۸
۳۳	۲/۲۳	۹	۰/۷۲	۱/۲۱	۱/۱۰	۰/۸٥	۰/۷٥	۱۸
۳٤	۲/٥۸	۸	۰/۹٥	۱/۲۷	۰/۸۷	۱/٦۲	۱/۰۹	۱٤
۳٥	۲/۱٦	۷	۰/۹٤	۰/۷۸	۱/۳۰	۰/۷۸	۰/٦٦	۱٦
۳٦	۲/۰۳	٥	۱/۰۳	۰/۸۳	۰/۷٤	۰/۸٤	۰/۸۰	۱۷
۳۷	۲/۳۹	٤	۱/۲۲	۰/۸٤	۰/۳۹	۱/۲۱	۰/٦۸	۱٥
۳۸	۲/۰٦	۳	۰/۸۹	۰/٦۲	۰/٦۷	۰/۷۳	۰/۷٦	۱٤
۳۹	۱/۷۰	۳	۰/٦٥	۰/۹٦	۱/۱۱	۰/۸٥	۰/٦٦	۱۷
٤۰	۱/۷۹	٤	۰/۹٦	۰/۹۰	۰/۸۳	۰/۸۱	۰/۷۸	۱۹
٤۱	۱/٤٥	٥	۱/۰۱	۰/۷٤	۰/٦۹	۰/۷۸	۰/۷۱	۱۹
٤۲	۱/۲۸	٥	۱/۰۲	۱/۰۰	۰/۸۳	۰/۸٥	۰/۸۰	۱۷
٤۳	۲/۲۰	٤	۱/۱۲	۰/٦٥	۰/۸٦	۰/٦٤	۰/۸۸	۱۸
٤٤	۱/٥۱	۳	۱/۰۸	۰/۸۷	۰/۸٥	۰/۸٦	۰/۷۹	۱۷
٤٥	۱/٦۰	۸	۱/۳۹	۰/٦۷	۱/۰۰	۰/۸۳	۰/۸۰	۱٦
٤٦	۱/۲۰	۷	۱/۳۱	۰/۸۳	۰/۸۰	۰/۷۹	۰/٦٥	۱۸
٤۷	۱/۲۸	٥	۱/۰۰	۰/۹٦	۰/۹٤	۰/۷٦	۰/٦۹	۱۹
٤۸	۱/۳٥	٤	۱/۲۱	۰/۸۸	۱/۰۲	۰/۸۳	۰/۸٥	۱۷
٤۹	۱/٦٦	٦	۱/۰۷	۰/۹۰	۰/۷۰	۰/۷۹	۰/۸٥	۱٦
٥۰	۲/۳۰	۸	۱/٥۱	۱/۰۱	۱/۳۲	۱/۲۷	۰/۸۱	۱٦
میانگین	**۲/۲۰**	**٥/٤۲**	**۱/۰۹**	**۱/۰۱**	**۰/۹٤**	**۰/۹۱**	**۰/۸٥**	**۱٦/۹**

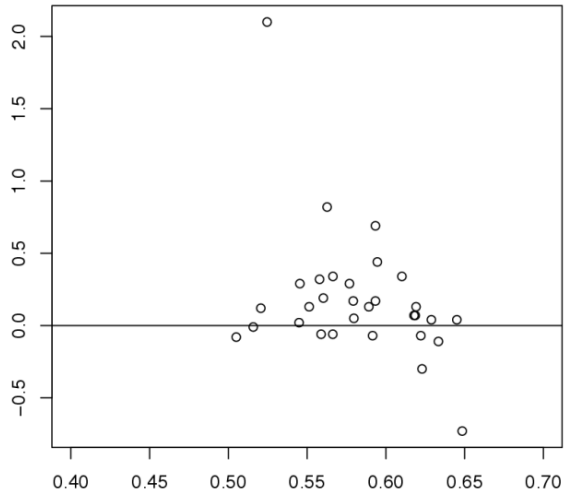
[Downloaded from c4i2016.khu.ac.ir on 2026-11-04]

$$\mathrm{RMSPE} = [\frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}}(\mathrm{O_i} - \widehat{\mathrm{O}}_\mathrm{i})^2]^{\frac{1}{2}}$$

محاسبه و عمل‌کرد پیش‌گویی‌ها را اندازه‌گیری می‌کند، که در آن n تعداد ایستگاه‌های فاقد مشاهده برای پیش‌گویی و O_i اندازه در ایستگاه iام و $\widehat{O}_i$ 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 به‌دیگر 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 ٦. تفاوت میانگین RMSPE پیش‌گویی‌های کریگیدن با استفاده از اندازه‌ها و مدل تلفیق بیزی در مقابل هم‌بستگی بین اندازه‌ها و خروجی‌های مدل

شکل ٦ 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

[Downloaded from c4i2016.khu.ac.ir on 2026-11-04]

جدول ۳. میانگین روزانه RMSPE پیش‌گویی اُزن در ۱۲ ایستگاه

روز	تلفیق بیزی	کریگیدن با اندازه‌ها	کریگیدن با خروجی‌های مد
۱	۱/٥۸*	۱/۷٥	۲/٦۷
۲	۱/۸٦	۱/۱۳*	۱/۸۸
۳	۱/۹۲*	۲/۰٥	۲/۱٦
٤	۱/٦۳	۱/۳۳*	۲/۲۸
٥	۲/۳۹*	۲/٤۳	۲/٤۰
٦	۱/۸٦*	۲/۲۰	۳/۲۰
۷	۱/۸۰	۱/۷۳*	۲/٥٥
۸	۱/۸۱*	۲/۲٥	۲/٤۱
۹	۲/۱۰*	۲/۹۲	۲/۷٤
۱۰	۱/۲۲*	۱/۳٥	۱/۳۹
۱۱	۱/٤٤	۱/٥٦	۱/۰۱*
۱۲	۱/٥۸*	۱/۷۷	۲/۷۷
۱۳	۱/٦۹*	۱/۸٦	۳/۳٥
۱٤	۱/٦۱*	۱/٦٥	۱/٦۲
۱٥	۱/۳۰	۱/۲٤*	۱/٤۹
۱٦	۲/۱٤	۲/۰۷*	۳/۳۲
۱۷	۱/۹۲	۱/۹۱*	٤/٤۸
۱۸	۱/۹۲*	۱/۹۹	۲/۰٥
۱۹	۱/۸۷	۱/۸۱*	۲/٥۲
۲۰	۲/۰۸	۲/۰۰*	۲/٥۰
۲۱	۱/۷۲	۱/۱٥	۱/۰۷*
۲۲	۱/٦۹*	۱/۷۱	۲/۹۲
۲۳	۱/۸٥*	۲/٥٤	۲/۹٦
۲٤	۱/۱۰*	۱/٤۲	۲/۸٤
۲٥	۱/٤۷*	۱/۷٦	۳/٦۸
۲٦	۱/٦۷	۱/٥٦*	۱/۸۸
۲۷	۱/۲۸*	۱/٤۱	۱/۹٤
۲۸	۱/۷٥*	۱/۸۰	۱/۹۱
۲۹	۱/٤٥	۱/۸٦	۱/۰۱*
۳۰	۲/۲۲*	۲/۲۹	۲/۳۹
۳۱	۱/۱۱*	۱/٤۰	۲/۰۹
میانگین	۱/۷۱*	۱/۸۰	۲/۳٥

تحلیل میانگین هفتگی داده‌ها

ایستگاه‌های مانیتورینگ، ایستگاه‌های فاقد مشاهده و سلول‌های مشبکه‌ای استفاده شده در تحلیل هفتگی داده‌ها همان ایستگاه‌ها و سلول‌های مشبکه‌ای استفاده شده در تحلیل روزانه داده‌ها هستند. به‌طور متوسط در مقیاس‌های زمانی طولانی‌تر، انتظار می‌رود که پیش‌گویی‌های هر دو مدل تلفیق بیزی و کریگیدن 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 کریکینگ در زمان‌های طولانی‌تر می‌شود. جدول ٤، میانگین هفتگی جذر میانگین توان‌های دوم خطای پیش‌گویی مدل تلفیق بیزی، کریگیدن با استفاده از اندازه‌ها، کریگیدن با استفاده از خروجی مدل قطعی را نشان می‌دهد.

جدول ٤. میانگین هفتگی RMSPE پیش‌گویی اُزن در ۱۲ ایستگاه

هفته	تلفیق بیزی	کریگیدن با اندازه‌ها	کریگیدن با خروجی‌های مد
اول	۱/٥۷	۱/۷٥	۱/٥٥
دوم	۱/٤۱*	۱/۸٥	۳/٦۳
سوم	۱/٥٥*	۱/٦٥	۱/۷۲
چهارم	۱/۰٥	۱/۰۳	۱/٦۹
میانگین	۱/۳۹	۱/٥۷	۲/۱٤

در تحلیل میانگین هفتگی داده‌ها با توجه به جدول ٤، مدل تلفیق بیزی در هفته دوم و سوم، میانگین جذر

[Downloaded from c4i2016.khu.ac.ir on 2026-11-04]

میانگین توان‌های دوم خطای کوچک‌تری دارد که به‌ترتیب برابر ۱/٤۱ و ۱/۵۵ است. میانگین جذر میانگین توان‌های دوم خطای کل هفته‌ها نیز برای مدل تلفیق بیزی از همه کم‌تر و برابر ۱/۳۹ است. چنان‌که انتظار داشتیم، میانگین جذر میانگین توان‌های دوم خطای همهٔ 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 و٤ 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 مشبکه‌ای 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 نزدیك به‌هم 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 یك 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 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

[Downloaded from c4i2016.khu.ac.ir on 2026-11-04]

Research, 108 (2003) 80-91.

5. Fuentes M., Raftery A., "Model Validation and Spatial Interpolation by Bayesion 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 cheins 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, Atoms. 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 Repoblic of Iran, 24(2013) 171-178.
15. Poole D., Reftery A., "Inference for Deterministic Simulation Mels, the Bayesian Melding Approch, 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).

[Downloaded from c4i2016.khu.ac.ir on 2026-11-04]