سیزدهمین کنفرانس
آمار ایران
۲-۴ شهریور ۱۳۹۵

مدل‌بندی فضایی-زمانی انعطاف‌پذیر با توابع پایه زمانی هموار و جانهی مقادیر گمشده

میثم اسلامی[1]، محسن محمدزاده[2]

[1,2]گروه آمار دانشگاه تربیت‌مدرس

چکیده: هرچند در مدل‌های فضایی-زمانی روش‌های مختلفی برای برآورد روند فضایی وجود دارد، اما با اضافه شدن متغیر زمان، پیچیدگی‌ها و تعداد انتخاب‌های بسیاری به‌وجود می‌آید. مدل‌بندی روند فضایی-زمانی با استفاده از توابع پایه زمانی هموار تا حدی این مشکل را برطرف ساخت و تعمیم آن رویکرد جدیدی در مدل‌بندی فضایی-زمانی به وجود آورد. اما با توجه به اینکه در عصر بزرگ داده به سر می‌بریم و گاهی وجود داده‌های گمشده اجتناب ناپذیر است، ناگزیر در مدل‌بندی فضایی-زمانی، مسئله جانهی مقادیر گمشده اهمیت پیدا می‌کند. جانهی مقادیر گمشده با الگوریتم فوئنتس همچون سایر الگوریتم‌ها دارای حلقه‌های تکراری است و امکان عدم‌همگرایی برای آن وجود دارد. در این مقاله ضمن معرفی مدل‌بندی مذکور، روشی برای جانهی مقادیر گمشده ارائه می‌شود که نه تنها از امکان همگرایی زیادی برخوردار است، بلکه وابستگی بین مشاهدات را نیز در نظر می‌گیرد. سپس روش ارائه شده برای تحلیل داده‌های آلودگی هوا مورد استفاده قرار می‌گیرد و در انتها بحث و نتیجه گیری خواهد شد.

واژه‌های کلیدی: مدل‌بندی فضایی-زمانی، کریگیدن عادی، مقادیر گمشده فضایی، توابع پایه زمانی و الگوریتم فوئنتس

کد موضوع‌بندی ریاضی (۲۰۱۰): 62M30

۱ مقدمه

یکی از مباحث مهم در مدل‌بندی فضایی-زمانی داده‌ها، چگونگی مدل‌بندی روند آن‌ها است. کرسی (۱۹۹۳) نشان داد وجود روند در داده‌ها موجب اریبی در براورد تابع کوواریانس می‌شود. برای کاهش اریبی لازم است براورد تابع کوواریانس بر اساس داده‌های روند زدوده صورت پذیرد و برای این منظور باید روند داده‌ها مدل‌بندی شود و با کسر روند از مشاهدات، مانده‌ها حساب شوند. سپس از مانده‌های روندزوده به جای داده‌های واقعی برای برازش تابع کوواریانس استفاده نمود. برای مدل‌بندی روند داده‌های فضایی-زمانی روش‌های متعددی را می‌توان به کار گرفت. ساده‌ترین حالت ترکیبی خطی از موقعیت‌های فضایی و زمانی یا متغیرهای تبیینی است. تامسون و همکاران (۲۰۰۱) تکنیک‌های مختلف آمار را در سه بخش تحلیل رگرسیونی، تحلیل مقادیر فرین و تحلیل فضایی-زمانی داده‌ها به کار

[1]میثم اسلامی : m.islami@modares.ac.ir

گرفتند. استراد و همکاران (۲۰۰۱) و هورتا و همکاران (۲۰۰۴) داده‌های فضایی-زمانی را با مدل خطی پویا با رهیافت بیزی تحلیل کردند.

گایتن و گریگولتو (۲۰۰۴) یک مدل پویای غیرخطی معرفی نمودند. مک میلان و همکاران (۲۰۰۵) یک مدل خطی نزدیکترین همسایگی[۱] را به کار بردند. از آنجا که توابع کوواریانس فضایی-زمانی معمولاً یک تابع برداری نامعلوم است و برازش یک مدل معتبر به داده‌های کوواریانس تجربی معمولاً امری دشوار است، توابع کوواریانس فضایی-زمانی تفکیک‌پذیر و تفکیک‌ناپذیر توسط گنتیگ و گوتورپ (۲۰۱۰) مطرح و نحوه برآورد آنها ارائه شد. گنتینگ (۲۰۰۲) مدلی کلی برای ساخت توابع کواریانس فضایی-زمانی تفکیک‌ناپذیر ارائه داد اما کنت و محمدزاده (۲۰۱۱) نشان دادند که این مدل برای برخی از توابع کاملاً یکنوا و برنشتاین دارای گودی[۲] است. گودی وقتی رخ می‌دهد که تابع کواریانس فضایی-زمانی به‌ازای افزایش برخی از تأخیرهای فضایی یا زمانی روند نزولی بودن خود را از دست دهد. زیرا در تحلیل فضایی و فضایی-زمانی با افزایش تأخیرها انتظار داریم همبستگی فضایی یا فضایی-زمان کم شود و در نتیجه تابع کورایانس فضایی یا کواریانس فضایی-زمانی بایستی روند نزولی داشته باشد. وجودگودی در تابع کواریانس فضایی-زمانی استفاده از خانواده توابع کواریانس گنیتینگ را محدود می‌سازد که لازم است در مسائل کاربردی به آن توجه ویژه شود. امیدی و محمدزاده (۲۰۱۵) با استفاده از تابع مفصل خانواده‌ای از توابع فضایی-زمانی تفکیک‌پذیر ساختند و نشان دادند که برای همه تاخیرهای فضای و زمانی فاقد هر گونه گودی هستند.

ازپیرو و همکاران (۲۰۱۰) با استفاده از توابع پایه زمانی هموار توانستند یک مدل فضایی-زمانی معرفی کنند که در آن عبارت روند توسط توابع مذکور بیان می‌شود و هدف از تعریف این توابع، درنظرگرفتن تغییرات زمانی در داده‌ها است. (لیندستروم و همکاران، ۲۰۱۳) برای انعطاف‌پذیری عبارت روند درکنار توابع پایه زمانی، متغیرهای کمکی فضایی-زمانی را نیز افزوده و مدل مذکور را به مدلی انعطاف‌پذیرتر تعمیم دادند که رویکرد جدیدی در مدل‌بندی فضایی-زمانی به وجود آورد. برای کاهش زمان محاسبه برآورد پارامترهای مدل، از روش درستنمایی نیمرخی استفاده کردند، که در آن از ساختار ماتریس کوواریانس قطری بلوکی نیز استفاده می‌شود. برای تعیین تعداد توابع پایه زمانی هموار بایستی تمام مشاهدات در دسترس باشند و به دلیل حجم بالای ابعاد ماتریس مشاهدات فضایی-زمانی و وجود مقادیر گمشده، یکی از بخش‌های مهم و گریز ناپذیر، چگونگی جانهی مقادیر گمشده است. برای جانهی مقادیر گمشده در این مدل‌بندی‌ها تا به حال از الگوریتم فوئنتس و همکاران (۲۰۰۶) استفاده شده است. این الگوریتم به‌دلیل تکراری بودن، احتمال عدم‌همگرایی در آن وجود دارد و از معایب دیگر آن، عدم‌توجه به همبستگی زمانی و مکانی بین مشاهدات است. در این بخش ضمن معرفی مدل‌بندی مذکور، روش کریگیدن عادی برای جانهی مقادیر گمشده ارائه می‌شود که نه تنها همگرا است بلکه همبستگی مکانی مشاهدات را نیز در نظر می‌گیرد. در آخر برای داده‌های آلودگی هوا نشان داده می‌شود که جانهی با استفاده از روش کریگیدن عادی، نسبت به الگوریتم فوئنتس موجب پیشگویی با دقت بالاتر می‌شود.

۲ مدل

مدل فضایی-زمانی $y(s,t) = \mu(s,t) + v(s,t)$ را در نظر بگیرید، که در آن $y(s,t)$ مشاهده در موقعیت s و زمان t، $\mu(s,t)$ روند یا میانگین میدان و $v(s,t)$ مانده فضایی-زمانی است. فرض کنید روند به‌صورت $\mu(s,t) = \sum_{\ell=1}^{L} \gamma_\ell \mathcal{M}_\ell(s,t) +$

[۱] Nearest-neighbor

[۲] Dimple

$\sum_{i=1}^{m}\beta_i(s)f_i(t)$ قابل مدل‌بندی باشد، که در آن $\mathcal{M}_\ell(s,t)$ متغیرهای تبیینی فضایی-زمانی، γ_ℓ ضرایبی برای متغیرهای تبیینی، $\{f_i(t)\}_{i=1}^{m}$ مجموعه‌ای از توابع پایه زمانی(هموار) است به‌طوری که $f_1(t)\equiv 1$ و $f_1(t),...,f_m(t)$ دارای میانگین صفر هستند، $\beta_i(s)$ ضرایبی برای توابع پایه زمانی با میانگین $X_i\alpha_i$ و تابع کوواریانس $\Sigma_{\beta_i}(\theta_i)$ برای iهای $1,...,m$ هستند که در آن X_iماتریس طرح با ابعاد $n\times p_i$، α_i بردار ضرایب رگرسیونی p_iبعدی و $\sum_{\beta_i}(\theta_i)$ ماتریس کوواریانس با ابعاد $n\times n$ هستند. فرض کنید میدان فضایی-زمانی مانده $\nu(s,t)$ مانا با میانگین صفر و کوواریانس فضایی $\Sigma_\nu^t(\theta_\nu)$ برای tهای $1,...,T$ باشد، که در آن $t_1\neq t_2,\ \nu(s_1,t_1)\perp\nu(s_2,t_2)$. اندازه هر بلوک از ماتریس $\Sigma_\upsilon^t(\theta_\upsilon)$ تعداد مشاهدات n_t در هر نقطه زمان t است. هریک از ماتریس‌های کوواریانس $\Sigma_{\beta_i}(\theta_i)$ و $\Sigma_\nu^t(\theta_\nu)$ می‌توانندتوابع کوواریانس متفاوتی داشته باشند که این امر موجب انعطاف‌پذیری مدل می‌شود. میدان ν متشکل از یک مولفه همبسته ν^* و یک اثرقطعه‌ای ناهمبسته شامل تغییرات کوچک مقیاس و خطاهای اندازه‌گیری ν_{nugget} است، یعنی $\nu(s,t)=\nu^*(s,t)+\nu_{nugget}(s,t)$ و $\Sigma_\nu=\Sigma_{\nu^*}+\Sigma_{\nu,nugget}$ که در آن $\sum_{\nu,nugget}$ ماتریس قطری است. بعلاوه یک ماتریس تُنُک[۳] $F=(f_{st,is'})$ به ابعاد $N\times nm$ با درایه‌های

$$f_{st,is'}=\begin{cases} f_i(t) & s=s' \\ 0 & \text{در غیر این صورت}\end{cases}$$

معرفی شده است. ماتریس‌های $Y=y(s,t)$ و $V=\nu(s,t)$ را به صورت بردارهای $N\times 1$ بعدی می‌نویسیم که ابتدا مکان و سپس زمان‌ها به‌صورت $Y=[y(s_1,1)\ \ldots\ y(s_n,1)\ \ldots\ y(s_1,T)\ \ldots\ y(s_n,T)]^T$ تغییر می‌کنند و فرض می‌شود که مشاهدات متناظر برای متغیرهای تبیینی نیز موجود باشند. پارامترهای رگرسیونی و کوواریانس مدل به‌صورت بردارهای ستونی

$$\gamma=[\gamma_1\ldots\gamma_\ell]^T\quad,\quad \alpha=[\alpha_1^T\ldots\alpha_m^T]^T\quad,\quad \theta_B=\{\theta_1(t),...,\theta_m(t)\}\quad,\quad \psi=\{\theta_B,\theta_\upsilon\}$$

و تمام متغیرهای تبیینی فضایی-زمانی $\mathcal{M}=[\mathcal{M}_1\ldots\mathcal{M}_\ell]$ به‌صورت ماتریس $N\times L$ و مولفه‌های میدان β_i نیز به صورت ماتریس‌های بلوکی بازنویسی شده است. با استفاده از آن‌ها می‌توان مدل فضایی-زمانی را به‌صورت

$$Y=\mathcal{M}\gamma+FB+V\qquad,\qquad B\sim N(X\alpha,\Sigma_B(\theta_B))\quad,\quad V\sim N(0,\Sigma_\nu(\theta_\nu))$$

بازنویسی کرد و از آنجا که این مدل ترکیب خطی از متغیرهای گاوسی مستقل است، با تعریف ماتریس‌های $\tilde{X}=[\mathcal{M}\ \ FX]$ و $\tilde{\Sigma}(\Psi)=\Sigma_\nu(\theta_\nu)+F\Sigma_B(\theta_B)F^T$، توزیع بردار Y به‌صورت

$$[Y|\Psi,\eta]\sim N(\tilde{X}\eta^T,\tilde{\Sigma}(\Psi)) \tag{۱.۲}$$

حاصل می‌شود، که در آن $\eta=[\ \gamma\quad\alpha\ ]$.

۳ توابع پایه زمانی هموار

برای بدست آوردن توابع زمانی ابتدا ماتریس داده‌ها با ابعاد $T\times n$ به‌صورت

$$D(t,s)=\begin{cases} \text{y(t،s)} & \text{وجود مشاهده y(t،s)} \\ NA & \text{در غیر این صورت}\end{cases} \tag{۱.۳}$$

[۳] Sparse

ایجاد می‌شوند. برای جانهی مقادیر گمشده در (۱.۳) با توابع پایه زمانی هموار از الگوریتم فوئنتس استفاده می‌شود که دارای حلقه‌های تکراری است و احتمال عدم‌همگرایی برای آن وجود دارد.

الف - الگوریتم فوئنتس

گام ۱ : ستون‌های ماتریس $D(t,s)$ را استاندارد نموده و میانگین تمامی مشاهدات در هر نقطه زمانی $u(t)$ محاسبه شوند. هر ستون از $D(t,s)$ بر روی بردار $u(t)$ رگرسیو شده و برای مقادیر گمشده پیش‌گویی صورت گیرد. در این گام برای خوش تعریف بودن ماتریس داده‌ها بایستی حداقل یک مشاهده در هر سطر و ستون باشد.

گام ۲ : تجزیه ویژه مقدار [۴] (SVD) ماتریس داده‌های جدید با درنظرگرفتن مقادیر جایگزین برای مقادیر گمشده محاسبه شود.

گام ۳ : هرستون از ماتریس داده‌های جدید بر روی $m-1$تابع پایه متعامد که از گام ۲ بدست آمده است رگرسیو شود. سپس داده‌های گمشده توسط مقادیر برازش داده شده از این رگرسیون، جایگزین شوند.

گام ۴ : گام‌های ۲ و ۳ تا همگرایی تکرار شوند. همگرایی توسط تغییر در مقادیر جایگزین شده بین تکرارها اندازه‌گیری می‌شود.

حال مقادیر حاصل برای داده‌های گمشده در $D(t,s)$ جایگزین شود و با روش اسپلاین $m-1$ بردار ویژه هموارسازی می‌شود. یعنی$m-1$ ستون اول ماتریس U در تجزیه $D=USV^T$ به عنوان توابع پایه زمانی هموار $f_1(s),\ldots,f_m(s)$ در نظر گرفته شود.

ب - کریگیدن عادی

الگوریتم فوئنتس برای جانهی مقادیر گمشده در (۱.۳) دارای حلقه‌های تکراری است که در عمل ممکن است همگرا نباشند. برای رفع این مشکل می‌توان مقادیر گمشده را با استفاده از کریگینگ عادی [۵] که توسط ماترون (۱۹۷۱) برای پیش‌گویی مقدار میدان تصادفی در موقعیت مشخص s_0 استفاده نمود. اگر میدان تصادفی به صورت $Y(s)=\mu+V(s)$ باشد که در آن $\mu\in R$ نامعلوم و $V(s)$ میدان تصادفی مانای ذاتی، با میانگین صفر است، بهترین پیشگوی نااریب خطی برای $Y(s_0)$ به صورت $\hat{Y}(s_0)=\sum_{i=1}^{n}\lambda_i Y(s_i)$ است که میانگین توان دوم خطاها را مینیمم می‌کند (محمدزاده، ۱۳۹۴). در این روش برای هر زمان (در هر سطر از ماتریسD) کریگیدن عادی انجام می‌شود و برای موقعیت‌های فاقد مشاهده، پیش‌گویی صورت می‌گیرد. در این روش نیازی به تکرار نداریم ولی باید این مسئله را در نظر داشته باشیم که برای استفاده از کریگیدن عادی باید مشاهدات فاقد روند فضایی باشند.

برای تعیین تعداد توابع پایه زمانی مناسب از اعتبارسنجی متقابل استفاده می‌شود که در آن jامین ستون از $D(t,s)$ خارج می‌شود و سپس توابع زمانی هموار برای ماتریس فروکاسته محاسبه می‌شود و هریک از توابع با توجه به اینکه باحذف ستون jام تا چه اندازه تغییرات زمانی در داده‌ها را تبیین می‌کند ارزیابی می‌شوند. با تکرار برای تمامی ستون‌های $D(t,s)$، با ملاک‌های ارزیابی از جمله میانگین توان دوم خطاها [۶] (MSE)، ضریب تعیین[۷] (R^2)، ملاک اطلاع آکائیک[۸] (AIC) و ملاک اطلاع بیزی[۹] (BIC) تعداد توابع پایه زمانی مناسب محاسبه و قابل ارائه است.

[۴] Singular Value Decomposition

[۵] Ordinary Kriging

[۶] Mean Squared Errors

[۷] Coefficient of Determination

[۸] Akaike Information Criterion

[۹] Bayesian Information Criterion

۴ برآورد پارامترها و پیشگویی فضایی-زمانی

لگاریتم تابع درستنمائی (۱.۲) به‌صورت

$$2\ell(\Psi,\eta|Y) = -N\log(2\pi) - \log|\tilde{\Sigma}(\Psi)| - \left(Y - \tilde{X}\eta^T\right)^T \tilde{\Sigma}^{-1}(\Psi)\left(Y - \tilde{X}\eta^T\right) \quad (۱.۴)$$

است و برای ماکسیمم‌کردن آن از الگوریتم بهینه‌سازی بیرد و همکاران (۱۹۹۵) استفاده می‌شود. با استفاده از ساختار قطری بلوکی Σ_ν و Σ_B لگاریتم درستنمائی با اتحادهای ماتریسی هارویل (۱۹۹۷) ساده می‌شود (لیندستروم و همکاران، ۲۰۱۳). اگر η با تعمیم برآورد کمترین توان‌های دوم جایگذاری شود آنگاه باتوجه به $\hat{\eta}^T = \left(\tilde{X}^T\tilde{\Sigma}^{-1}\tilde{X}\right)^{-1}\left(\tilde{X}^T\tilde{\Sigma}^{-1}\tilde{X}\right)$ ، لگاریتم درستنمائی (۱.۴) می‌تواند توسط

$$\begin{aligned} 2\ell_{PROF}(\Psi|Y) &= -\log|\Sigma_\nu(\theta_\nu)| - \log|\Sigma_B(\theta_B)| - \log|\Sigma^{-1}_{B|Y}(\Psi)| - Y^T\hat{\Sigma}(\Psi)Y \\ &+ Y^T\hat{\Sigma}(\Psi)\mathcal{M}\left(\mathcal{M}^T\hat{\Sigma}(\Psi)\mathcal{M}\right)^{-1}\mathcal{M}^T\hat{\Sigma}(\Psi)Y + const. \end{aligned} \quad (۲.۴)$$

جایگزین و برآورد پارامتر Ψ با ماکسیمم سازی (۲.۴) به‌صورت

$$\hat{\Psi}_{PROF} = \arg\max_{\Psi} \ell_{PROF}(\Psi|Y) \quad (۳.۴)$$

حاصل می‌شود.

باتوجه به ساختار مدل معرفی شده، پیشگویی مقدار Y را می‌توان به بخش‌های پیشگویی بر اساس ضرایب رگرسیونی معلوم $(\mathcal{M}\hat{\gamma} + FX\hat{\alpha})$، پیشگویی بر اساس میدان‌های برآوردشده $\beta_i(s)$ $(\mathcal{M}\hat{\gamma} + FE(B_u|Y))$ و پیشگویی‌های کامل (همراه با اثرتصادفی) تجزیه کرد. با توجه به مشاهدات Y و تخمین پارامترهای کوواریانس Ψ، ضرایب رگرسیونی با واریانس $Var\left(\hat{\eta}^T|Y,\Psi\right) = \left(\tilde{X}^T\tilde{\Sigma}^{-1}\tilde{X}\right)^{-1}$ برآورد می‌شود. فرض کنید X_u، $\mathcal{M}_u$ و F_u به‌ترتیب نشان‌دهنده متغیرهای کمکی فضایی-زمانی، متغیرهای کمکی جغرافیایی و توابع پایه زمانی باشند که در زمان یا موقعیت فضایی آن‌ها مشاهده‌ای وجود ندارد. بعلاوه B_u نیز نشان‌دهنده مجموعه‌ای از میدان‌های β در مکان‌های فاقد مشاهده است، $\Sigma_{B,uo}$ و $\Sigma_{\nu,uo}$ ماتریس کوورایانس متقابل میان موقعیت‌های دارای مشاهده و فاقد مشاهده است و $\Sigma_{B,uu}$ و $\Sigma_{\nu,uu}$ ماتریس کوواریانس برای نقاط فاقد مشاهده است. با استفاده از این نمادها، می‌توان ماتریس‌های

$$\tilde{X} = [\ \mathcal{M}_u \quad F_uX_u\ ] \quad , \quad \tilde{\Sigma}_{uo} = \Sigma_{\nu,uo} + F_u\Sigma_{B,uo}F^T \quad (۴.۴)$$

را معرفی کرد. پیشگوها تحت میدان هموار v^* با Y^* نشان داده می‌شود و با پیشگوهای Y تنها در مکان‌های مشاهده شده فرق دارند. با توجه به (۴.۴) پیشگوی Y_u^* به‌صورت $E(Y_u^*|Y,\Psi) = \tilde{X}_u\hat{\eta}^T + \tilde{\Sigma}^*_{uo}\tilde{\Sigma}^{-1}\left(Y - \tilde{X}\hat{\eta}^T\right)$ با واریانس

$$Var\left(Y_u^*|Y,\Psi,\eta\right) = \tilde{\Sigma}^*_{uu} - \tilde{\Sigma}^*_{uo}\tilde{\Sigma}^{-1}\tilde{\Sigma}^*_{ou}$$

معرفی می‌شود، که در آن‌ها $\tilde{\Sigma}^*_{uo}$ ماتریس کوواریانس متقابل بدون اثرقطعه‌ای در ν است. در نهایت برای بررسی میزان دقت روش جانهی از روش MSE استفاده می‌شود.

۵ تحلیل داده‌های آلودگی هوا

در این بخش برای مدل‌بندی و تحلیل داده‌های آلودگی هوای شهر لس‌آنجلس، دو روش جانهی الگوریتم فوئنتس و همکاران و کریگیدن عادی به‌کار گرفته شده و پیشگویی داده‌ها مورد ارزیابی قرار می‌گیرد. یکي از گازهاي بسيار خطرناك كه از ديرباز افراد زيادي را به مرگ خاموش مي‌كشاند گازهای بی‌رنگ و بی‌بو NO_x و CO است كه به دليل ميل تركيبي بيشتر گاز CO با هموگلوبين خون نسبت به اكسيژن، مزيد بر علت مرگ تعداد زيادي از انسان‌ها شده است. در این مطالعه با ۷۰۰۰ داده فضایی-زمانی سروکار داریم که دارای ۲۴۲۳، معادل تقریباً ۳۴ درصد مشاهدات گمشده‌اند. تقریباً ۱۰ سال اندازه‌گیری در سایت‌های ثابت AQS از تاریخ ۱۹۹۹/۰۱/۱۳ تا ۲۰۰۹/۰۹/۲۳ انجام شده و در نیمه دوم دوره اندازه‌گیری، از تاریخ ۲۰۰۵/۱۲/۰۷ تا ۲۰۰۹/۰۷/۰۱ اندازه‌گیری در سایت‌های $MESA$ اضافه شده است. از آنجا که توزیع مشاهدات اصلی چوله هستند از مشاهدات تبدیل‌یافته لگاریتمی برای مدل‌بندی استفاده خواهیم کرد. داده‌ها لگاریتم متوسط‌های دو هفته‌ای غلظت $NO_x(ppb)$ در ۲۰ سایت AQS و ۵ مانیتور $MESA$ هستند و برای هریک از ۲۵ مکان، ۲۸۰ اندازه‌گیری صورت گرفته است.

برای تعیین تعداد توابع پایه زمانی مورد نیاز در مدل، نباید داده گمشده داشته باشیم. تا به حال برای جانهی از روش فوئنتس استفاده شده است و روشی که برای جانهی پیشنهاد می‌کنیم کریگیدن عادی است. در این روش اگر سطر و ستون‌ها به ترتیب مبین زمان و مکان مشاهدات باشد، ابتدا در هر سطر، مشاهدات را روندزدود کرده و برای هر زمان (سطر) یک تابع کوواریانس نمایی با روش کمترین توان‌های موزون[۱۰] (WLS) برازش داده و با کریگیدن عادی برای مقادیر گمشده پیشگویی نموده و پس از افزودن روند به آن، جانهی صورت می‌گیرد. برای تعیین تعداد توابع پایه زمانی از برخی ملاک‌های ارزیابی استفاده می‌شود که برای هر دو روش در جدول ۱ ارائه شده است. همان‌طور که ملاحظه می‌شود نتایج حاصل از هردو روش، نشان می‌دهند که دو تابع پایه زمانی هموار به خوبی تغییرات زمانی در داده‌ها را تبیین می‌کند.

جدول ۱: مقادیر ملاک‌های ارزیابی دو روش جانهی برای تعداد توابع پایه زمانی هموار مختلف

تعداد	روش فوئنتس MSE	-BIC	-AIC	$R^۲$	روش کریگیدن عادی MSE	-BIC	-AIC	$R^۲$
۰	۰/۳۸۸	۲۹۲/۸۵۸	۲۹۶/۴۹۳	۰/۰۰۰	۰/۴۰۶	۲۷۶/۰۱۹	۲۷۹/۶۵۴	۰/۰۰۰
۱	۰/۰۷۴	۷۶۲/۷۷۲	۷۷۰/۰۴۱	۰/۷۷۵	۰/۰۶۹	۷۵۴/۹۴۸	۷۶۲/۲۱۸	۰/۷۸۹
۲	۰/۰۶۳	۷۹۴/۲۵۸	۸۰۵/۱۶۲	۰/۸۱۴	۰/۰۵۹	۷۸۵/۸۵۹	۷۹۶/۷۶۳	۰/۸۲۹
۳	۰/۰۵۱	۸۳۲/۰۵۰	۸۴۶/۵۸۹	۰/۸۴۵	۰/۰۵۶	۷۹۲/۴۹۰	۸۰۷/۰۲۹	۰/۸۳۹
۴	۰/۰۴۷	۸۴۴/۱۸۰	۸۶۲/۳۵۴	۰/۸۵۵	۰/۰۵۵	۷۹۳/۵۴۶	۸۱۱/۷۲۰	۰/۸۳۹

برای مقایسه میزان دقت روش‌های جانهی، از ملاک میانگین توان‌های دوم خطاها استفاده می‌شود. با استفاده از این روش ابتدا برای مقادیر گمشده، توسط مدل برازش داده شده، پیشگویی انجام می‌دهیم و سپس با مقادیر جانهی شده مقایسه می‌کنیم. اگر مقدار جانهی شده توسط الگوریتم فوئنتس را با F، مقدار جانهی شده توسط کریگیدن عادی را با K و مقدار پیشگویی شده در آن نقطه را با P نشان

[۱۰] Weigthed List Square

دهیم آنگاه ملاک‌های ارزیابی آن‌ها به صورت

$$MSE_{Fuentes} = \frac{\sum_{i=1}^{n}(P_i - F_i)^2}{n} \qquad (۱.۵)$$
$$MSE_{Ordinal\ Kriging} = \frac{\sum_{i=1}^{n}(P_i - K_i)^2}{n}$$

محاسبه می‌شوند. بر اساس روابط (۱.۵) مقادیر $MSE_{Fuentes}$ و $MSE_{Ordinal\ Kriging}$ به ترتیب برابر با ۲۶۶/۳۸ و ۲۴۱/۶۰ است و نتایج حاکی از آن است که جانهی با روش کریگیدن عادی از دقت بالاتری برخوردار است. برای برآورد پارامترها و پیشگویی تنها از مشاهدات در دسترس استفاده شد و کاربست این جانهی تنها برای تعیین تعداد توابع پایه زمانی هموار است و برای برآورد پارامترها و پیشگویی کاربردی ندارد. حال اگر مقادیر گمشده را توسط هر دو روش جانهی کنیم و سپس برای ۷۰۰۰ مشاهده، پارامترها را تخمین بزنیم و مدل برازش دهیم، دو مدل با پارامترهای مختلف در دست داریم. برای هریک از مدل‌ها مقادیر $MSE_{Fuentes}$، $MSE_{Ordinal\ Kriging}$ و $MSE_{withoutfill}$ به ترتیب برابر با ۴۰/۱۷۱، ۷/۷۰۲ و ۱۸/۰۵۲ حاصل شده‌اند. همان‌طور که ملاحظه می‌شود مدلی که مقادیر گمشده آن توسط روش کریگیدن عادی جانهی می‌شود از دقت بالاتری برخوردار است. اگر مدل‌بندی بدون جانهی مقادیر گمشده انجام شود و برای مشاهدات در دسترس پیشگویی انجام دهیم آنگاه مقدار MSE برابر با ۱۸/۰۵۲ است. یعنی جانهی مقادیر گمشده با روش کریگیدن عادی به دقت پیشگویی‌های حاصل از مدل می‌افزاید.

بحث و نتیجه‌گیری

در این مقاله مدل فضایی-زمانی انعطاف‌پذیر با توابع پایه زمانی هموار معرفی شد و نشان دادیم برای تحلیل داده‌های آلودگی هوا دو تابع پایه زمانی هموار به خوبی تغییرات زمانی در داده‌ها را تبیین می‌کند و درصورتی که مشاهدات گمشده با روش کریگیدن عادی جانهی شود و مدل‌بندی با حضور مقادیر جانهی شده انجام شود، میزان دقت پیشگویی‌های حاصل از مدل افزایش پیدا می‌کند.

مراجع

محمدزاده، م. (۱۳۹۴). آمار فضایی و کاربردها، چاپ دوم، انتشارات دانشگاه تربیت‌مدرس، تهران.

Cressie, N. (1993), *Statistics for Spatial Data Revised edition*, London, Wiley.

Fuentes, M., Guttorp, P., and Sampson, P. D. (2006), *Using Transforms to Analyze Space-Time Processes*, In: Finkenstädt B, Held L, Isham V (eds) Statistical Methods For Spatio-Temporal Systems, London, Chapman and Hall/CRC.

Gaetan, C. and Grigoletto, M. (2004), Smoothing Sample Extremes eith Dynamic Models, *Extremes*, **7** ,221–236.

Gneiting, T. (2002), Nonseparable Stationary Covariance Functions for Space-Time Data, *Journal of American Statistical Association*, **97**, 221–236.

Gneiting, T., and Guttorp, P. (2010), *Continuous Parameter Spatio-Temporal Processes*, Handbook of Spatial Statistics, Chapman and Hall/CRC,427–436.

Harville, D. A. (1997), Matrix Algebra From a Statistician's Perspective, *first edition, Springer*.

Huerta, G., Sanso, B. and Stroud, J. R. (2004), A Spatio-Temporal Models for Mexico City Ozone Levels, *Applied Statistics*, **53**, 231-248.

Kent, J. T., Mohammadzadeh, M. and Mohammadian, M. A. (2011), The Dimple in Gneiting's Spatial-Temporal Covariance Model, *Biometrika, 98, No. 2,* 489-494.

Lindström, J., Szpiro, A. A., Sampson, P. D., Oron, A., Richards, M., Larson, T., and Sheppard, L. (2013), A Flexible Spatio-Temporal Model for Air Pollution with Spatial and Spatio-Temporal Covariates, *Under revision for Environmental and Ecological Statistics, TBD.*

McMillan, N., Bortnick, S. M., Irwin, M. E. and Berliner, M. (2005), A Hierachical Bayesian Model to Estimate and Forecast Ozone Through Space and Time, *Atmospheric Environment*, **39**, 1373–1382.

Omidi, M. and Mohammadzadeh, M. (2015), A New Method to Build Spatio-Temporal Covariance Functions: Analysis of Ozone Data, *Statistical Papers, DOI 10.1007/s00362-015-0674-2*.

Stroud, R. j,. Muller, P. and Sanso, B. (2001), Dynamic Models Spatio-Temporal Data, *Journal of the Royal Statistical Society, Series B*, **63**, 637–689.

Szpiro, A. A., Sampson, P. D., Sheppard, L., Lumley, T., Adar, S., and Kaufman, J. (2010), Predicting intra-urban Variation in Air Pollution Concentrations with Complex Spatio-temporal Dependencies, *Environmetrics*, **21**, 606–631.

Thampson, M. L., Reynolds, j., Cox, L. H., Guttorp, P. and Sampson, P. D. (2001), A Review of Statistical Methods for the Meteorologial Adjustment of Tropospheric Ozone, *Atmospheric Environment*, **35**, 617–630.