مجله علوم دانشگاه تهران
جلد سی‌ام (۱۳۸۳) شماره ۱ (۱۴۴-۱۳۳)

پیشگویی فضایی بیزی برای یک میدان تصادفی تبدیل یافته

مجید جعفری خالدی و محسن محمدزاده
گروه آمار ـ دانشگاه تربیت‌مدرس
(دریافت: ۸۲/۱۰/۲ ؛ پذیرش: ۸۳/۳/۲۰)

چکیده

یکی از موضوعات مهم در تحلیل داده‌های فضایی، پیشگویی مقدار نامعلوم یک میدان تصادفی در موقعیتهای مشخص براساس بردار مشاهدات است. اگر میدان تصادفی مورد نظر گاوسی با ساختار میانگین و کوواریانس پارامتری باشد، پیشگوی بهینه و میانگین مجذور خطای آن قابل محاسبه است. اما در عمل با موارد زیادی مواجه می‌شویم که مشاهدات از مدل گاوسی تبعیت نمی‌کنند. در اینگونه موارد اگر تبدیلی غیر خطی از میدان تصادفی، گاوسی باشد پیشگویی فضایی امکان پذیر می‌گردد. اما اگر این تبدیل نامعلوم باشد، می‌توان آنرا متعلق به یک خانواده از تبدیلات پارامتری دانست. در اینصورت پیشگوی فضایی علاوه بر پارامترهای مدل به پارامتر تبدیل نیز وابسته خواهد شد و معمولا برآوردهای حداکثر درستنمایی آنها تعیین و در پیشگوی بهینه جایگذاری می‌شود. اما این کار موجب تردید در بهینگی پیشگوی حاصل خواهد شد. از طرف دیگر، معمولا تعیین میانگین مجذور خطای این پیشگو یا حتی ارائه تقریب مناسبی برای آن بسیار دشوار یا نشدنی است. لذا در این مقاله برای فائق آمدن بر این مسائل، رهیافت بیزی اتخاذ گردیده و با فرض آنکه پارامترهای مدل و تبدیل متغیرهایی تصادفی هستند، پیشگوی بهینه و میانگین مجذور خطای آن تعیین می‌شوند و نحوه اجرای این تکنیک در یک مثال کاربردی ارائه خواهد شد.

واژه‌های کلیدی: *داده‌های فضایی، میدان تصادفی، پیشگویی فضایی بیزی، خانواده تبدیلات.*

۱ـ مقدمه

داده‌هایی که نوعاً بر حسب موقعیت و مکان قرار گرفتن آنها در فضای مورد مطالعه وابسته هستند، داده‌های فضایی نامیده شده و در آمار فضایی برای مدل بندی انها از میدان تصادفی $\{Z(t); t \in D \subseteq R^d\}$، $d \geq 1$، که در آن D یک مجموعه اندیس گذار است، استفاده می‌شود. یکی از موضوعات مهم در ارتباط با تحلیل داده‌های فضایی پیشگویی مقدار نامعلوم میدان تصادفی در موقعیتهای مشخص براساس مشاهدات بدست آمده در موقعیتهای دیگر است. با فرض آنکه میدان تصادفی مورد نظر گاوسی و ساختار توابع میانگین و کوواریانس ان پارامتری باشد، پیشگوی بهینه و میانگین مجذور خطای آن قابل تعیین خواهد بود (Mohammadzadeh and Jafari, 2003). اما در عمل با موارد زیادی مواجه می‌شویم که مشاهدات از مدل گاوسی تبعیت نمی‌کنند و انتخاب این مدل ممکن است نتایج هر گونه پیشگویی را نامناسب سازد (Cressie 1993). لذا ارائه راهکاری مناسب برای پیشگویی فضایی از موضوعات مهم در این خصوص بشمار می‌رود. در چنین مواردی معمولا فرض می‌شود تبدیلی غیر خطی از میدان تصادفی مورد نظر، گاوسی با ساختار میانگین و کوواریانس پارامتری است. اگر این تبدیل معلوم باشد، پیشگویی فضایی امکان پذیر خواهد بود (Stien, 1999). در غیر اینصورت، می‌توان از یک خانواده تبدیلات پارامتری $G = \{g_\lambda(\cdot); \lambda \in \Lambda\}$ ، مانند خانواده باکس- کاکس (Box and Cox, 1964)، استفاده نمود. در این وضعیت پیشگوی فضایی علاوه بر پارامترهای مدل، یعنی پارامترهای توابع میانگین و کوواریانس فضایی، به پارامتر تبدیل λ نیز وابسته خواهد شد. اکنون می‌توان برآوردهای حداکثر درستنمایی پارامترهای نامعلوم مدل و تبدیل را تعیین و بعنوان مقادیر واقعی آنها در پیشگوی بهینه جایگزین نمود (Christensen and *et al*, 2001)، که در اینصورت پیشگوی با جایگذاری (Plug in Predictor) نامیده می‌شود. اما جایگزینی برآورد پارامترها بجای مقادیر واقعی آنها، بهینگی پیشگوی حاصل را با تردید مواجه می‌سازد. بعبارت دیگر، در صورتی که پارامترها بدرستی برآورد نشده باشند، پیشگوی فضایی از دقت لازم برخوردار نخواهد بود. بعلاوه تعیین میانگین مجذور خطای پیشگوی با جایگذاری یا حتی ارائه یک تقریب مناسب برای آن بسیار دشوار یا نشدنی خواهد بود. با توجه به مسائل مطروحه, اتخاذ یک رهیافت بیزی می‌تواند رهگشا باشد. در این مقاله، با فرض آنکه پارامترهای مدل و تبدیل متغیرهایی تصادفی هستند، یک توزیع پیشین که وابستگی پارامترهای مدل و تبدیل در آن لحاظ شده باشد، اختیار می‌گردد. بدین ترتیب پیشگوی بهینه و میانگین مجذور خطای آن با انتگرال‌گیری بر فضای پارامترهای مدل و تبدیل تعیین

می‌شوند. اما محاسبه پیشگوی بهینه و میانگین مجذور خطای آن معمولا دشوار است، لذا بکمک روشهای مونت کارلو, مقادیر تقریبی آنها ارائه می‌گردد.

در سالهای اخیر استفاده از روشهای بیزی برای تجزیه و تحلیل داده‌های فضایی رشد و توسعه چشمگیری یافته است. نخستین بار (Kitanidis, 1986) با اشاره به مشکلات روش کلاسیک, استفاده از رهیافت بیزی در تجزیه و تحلیل فضایی را پیشنهاد کرد. لی و زیک (Le and Zidek, 1992) روشهای بیزی را برای حالت ناپارامتری بطور تحلیلی بیان کردند.

هندکوک و اشتاین (Handcock and Stein, 1993) به مشکلات استفاده از روشهای بیزی از جمله تعیین توزیع پیشین و محاسبه پیچیده توزیع پیشگوی بیزی اشاره نمودند و برای سادگی محاسبات از توزیع‌های پیشین ناسره (Improper) استفاده کردند. هندکوک و والیس (Handcock and Wallis, 1994) برای مسائل مکانی- زمانی رهیافت بیزی را توسعه دادند. براون و همکاران (Brown and *et al*, 1995) به گسترش این موضوع کمک کردند. اکر و گلفند (Ecker and Gelfand, 1997) برآورد بیزی پارامترهای تغییرنگار کروی (Spherical Variogram)را مورد تجزیه و تحلیل قرار دادند.سان (Sun, 1998) به مقایسه پیشگوهای بیزی و کلاسیک با استفاده از مطالعات شبیه سازی پرداخت. تا قبل از سال ۱۹۹۹ بدلیل دشواری محاسبه توزیع پیشگوی بیزی, تجزیه و تحلیل بیزی داده‌های فضایی کمتر مورد توجه بود, اما بدنبال توسعه روشهای مونت کارلو در آمار بیزی, در این زمینه نیز شاهد تحول شگرفی بوده‌ایم. گادارد و همکاران (Gaudard and *et al*, 1999) برای یک میدان تصادفی گاوسی، با استفاده از روشهای مونت کارلو، پیشگویی فضایی بیزی را تعیین نمودند.

برگر و همکاران (Berger and *et al*, 2001) ,توزیع های پیشینی که توزیع های پسین ناسره را نتیجه می‌دهند مطالعه کردند.فونتس و اسمیت (Fuentes and Smith, 2002) روش بیزی را برای یک کلاس خاص از توابع ناهمسانگرد ارائه کردند. بانارجی و گلفند (Banerjee and Gelfand, 2002) با استفاده از روش دگردیسی فضایی(Spatial Deformation) به بررسی مسئله ناهمسانگردی برای مسائل پیشگویی فضایی پرداختند. یانگ و بانی (Hyoung and Bani, 2003) پیشگویی بیزی برای میدانهای تصادفی گاوسی چوله را مورد مطالعه قرار دادند.

محمدزاده و جعفری (Mohammadzadeh and Jafari, 2003) با بکارگیری توزیعهای پیشین سره، پیشگوی فضایی بیزی برای میدان تصادفی گاوسی را تعیین کردند که بدنبال آن در این مقاله این نوع پیشگویی برای میدانهای تصادفی که تبدیلی نامعلوم از آنها گاوسی باشد، ارائه می‌گردد.

۲ مدل آماری

فرض کنید $\{Z(t);t \in D\}$ یک میدان تصادفی مقدار حقیقی و بردار $Z = (Z(t_1),\dots Z(t_n))'$ میدان تصادفی $Z(\cdot)$ را در موقعیتهای $t_1,\dots,t_n$ نمایش دهد. بطور کلی پیشگویی مقدار نامعلوم میدان تصادفی در نقطه t_0، یعنی $Z(t_0)$، براساس بردار مشاهدات $z = (z(t_1),\dots z(t_n))$ مورد نظر است. فرض کنید $G = \{g_\lambda(\cdot);\lambda \in \Lambda\}$ یک خانوادهٔ پارامتری از تبدیلات غیر خطی یک به یک و مشتق پذیر با مشتقات پیوسته بر مجموعه $\Lambda \times R$ باشد و بازای بعضی مقادیر $\lambda \in \Lambda$ میدان تصادفی $\{Y(t) = g_\lambda(Z(t));t \in D\}$ گاوسی حقیقی مقدار با تابع میانگین $E(Y(t)) = f'(t)\beta$ و تابع کوواریانس $Cov(Y(s),Y(t)) = \frac{1}{\tau}\rho(s,t;\theta)$ باشد، که در آن $f(t) = (f_1(t),\dots,f_p(t))'$ یک بردار $p\times 1$ از مولفه‌های غیرتصادفی معلوم، $\beta = (\beta_1,\dots,\beta_p) \in R^p$ بردار پارامترهای رگرسیونی، $\tau = \frac{1}{Var(Y(t))}$ معکوس واریانس میدان و $\rho(s,t;\theta) = Corr(Y(s),Y(t))$ تابع همبستگی میدان با بردار پارامترهای $\theta = (\theta_1,\dots,\theta_q) \in \Theta \subseteq R^q$ است. با فرض آنکه برای هر $\theta \in \Theta$ تابع $\rho(.,.;\theta)$ پیوسته باشد، $Y = g_\lambda(Z) = (g_\lambda(Z(t_1)),\dots,g_\lambda(Z(t_n)))$ دارای توزیع نرمال n متغیره $N_n(X\beta,\frac{1}{\tau}\Sigma_\theta)$ است، که در آن ماتریس $X_{n\times q}$ و ماتریس همبستگی فضایی نمونه Σ_θ بصورت: $\Sigma_\theta = (\rho(t_i,t_j;\theta))$ و $X = \left(f'(t_1),\cdots,f'(t_n)\right)'$ تعریف می‌شوند. اگر X ماتریسی پر رتبه و Σ_θ ماتریسی معین مثبت باشد، لگاریتم تابع درستنمایی بعد از حذف ضرایب ثابت بصورت:

$$L(\eta;z) = (\frac{\tau}{2\pi})^{n/2} \mid \Sigma_\theta \mid^{-1/2} \exp\{-\frac{\tau}{2}(g_\lambda(z)-X\beta)'\Sigma_\theta^{-1}(g_\lambda(z)-X\beta)\}J_\lambda$$

خواهد بود، که در آن $J_\lambda = \prod_{i=1}^{n}|g'_\lambda(z_i)|$ ژاکوبین تبدیل و $\eta = (\phi,\lambda)$، با $\phi = (\beta,\tau,\theta)$، بردار پارامترهای مدل و تبدیل است. باتوجه به گاوسی بودن میدان تصادفی $Y(.)$، توزیع توام $(Y(t_0),Y)$ نرمال $n+1$ متغیره است و می‌توان نشان داد توزیع $Y(t_0)$ بشرط $Z = z$ و ϕ بصورت:

$$Y(t_0) \mid z,\phi \sim N(f'(t_0)\beta + r'_\theta\,\Sigma_\theta^{-1}(g_\lambda(z)-X\beta),\frac{1}{\tau}(1-r'_\theta\,\Sigma_\theta^{-1}r_\theta)) \qquad (۱)$$

خواهد بود، که در آن بردار $r'_\theta = (\rho(t_0, t_i, \theta))_{n\times 1}$ همبستگی فضایی متغیر تصادفی $Y(t_0)$ با بردار تصادفی Y است. با استفاده از توزیع شرطی (۱) توزیع پیشگوی بهینه (Optimal Predictor Distribution) برای $Z(t_0) = g_\lambda^{-1}(Y(t_0))$ بصورت:

$$f(z(t_0) \mid z) = (\frac{\tau}{2\pi})^{1/2}(1 - r'_\theta \Sigma_\theta^{-1} r_\theta) \mid g'_\lambda(z(t_0)) \mid \exp\{\frac{-\tau}{2(1 - r'_\theta \Sigma_\theta^{-1} r_\theta)}[g_\lambda(z(t_0)) - m_{\beta,\theta,\lambda}]^2\} \quad (۲)$$

خواهد شد، که در آن $m_{\beta,\theta,\lambda} = f'(t_0)\beta + r'_\theta \Sigma_\theta^{-1}(g_\lambda(z) - X\beta)$ است. اکنون براساس (۲) و یک تابع زیان مشخص می‌توان پیشگوی بهینه را تعیین نمود. برای تابع زیان درجه دوم و با فرض معلوم بودن پارامتر η، پیشگوی بهینه $E(Z(t_0) \mid z, \eta)$ و میانگین مجذور خطای آن $\sigma^2(t_0) = Var(Z(t_0) \mid z, \eta)$ خواهد بود. چون پیشگوی بهینه به پارامتر $p + q + 2$ بعدی η بستگی دارد، هنگامی که η نامعلوم است، معمولا برآورد حداکثر درستنمایی η بعنوان مقدار واقعی آن در پیشگوی بهینه جایگزین می‌شود. اما این جایگذاری، بهینگی پیشگوی حاصل را نامطمئن می‌سازد. همچنین، اغلب مواقع تعیین میانگین مجذور خطای پیشگوی با جایگذاری، یعنی $MSE(\hat{Z}_{\hat{\eta}}(t_0)) = E[\hat{Z}_{\hat{\eta}}(t_0) - Z(t_0)]^2$, یا حتی ارائه تقریب مناسبی برای آن بسیار دشوار یا نشدنی است. در بخش بعد بر اساس رهیافت بیزی به رفع این مسائل پرداخته می‌شود.

۳- پیشگویی فضایی بیزی

برای تعیین پیشگوی فضایی بیزی لازم است یک توزیع پیشین برای پارامترهای مدل و تبدیل اختیار شود. چون بردار پارامترهای مدل ϕ به پارامتر تبدیل λ وابسته است، در نظر گرفتن استقلال آنها ممکن است نتایج هر گونه تجزیه و تحلیلی را نامعتبر سازد. لذا بر اساس استدلالهایی که (Pericchi, 1981) و (Hinkley and Runger, 1984) در رابطه با رگرسیون بیزی کلاسیک مطرح نمودند، برای پارامترهای مدل و تبدیل توزیع پیشین توام:

$$\pi(\eta) = \pi(\beta, \tau, \theta, \lambda) = \frac{\pi(\lambda)\pi(\theta)}{\tau J_\lambda^{p/n}} \qquad (۳)$$

اختیار می‌شود، که در آن $\pi(\theta)$ و $\pi(\lambda)$ بترتیب توابع چگالی حاشیه‌ای θ و λ و همچنین، J_λ ژاکوبین تبدیل است. از آنجا که $\int_0^\infty \int_{-\infty}^\infty \frac{1}{\tau} d\beta d\tau = \infty$ است، توزیع پیشین (۳) ناسره است، لذا لازم است سره بودن توزیع پسین پارامترها، یعنی $\pi(\eta \mid z)$، مورد بررسی قرار

گیــرد. بــرای ایــن منظــور فــرض کنیــد $\Omega = R^p \times (0,\infty) \times \Theta \times \Lambda$ فــضای پــارامتری، $\hat{\beta}_{\theta,\lambda} = (X'\Sigma_\theta^{-1}X)^{-1}X'\Sigma_\theta^{-1}g_\lambda(z)$ و $q_{\theta,\lambda} = (g_\lambda(z) - X\hat{\beta}_{\theta,\lambda})'\Sigma_\theta^{-1}(g_\lambda(z) - X\hat{\beta}_{\theta,\lambda})$ باشد. اگر $\Theta \times \Lambda$ کراندار باشد تابع چگالی:

$$f(z) = \int_\Omega f(z \mid \eta)\pi(\eta)d\eta$$

$$= \int_\Lambda\int_\Theta |\Sigma_\theta|^{-1/2}|X'\Sigma_\theta^{-1}X|^{-1/2}\, q_{\theta,\lambda}^{-[(n-p)/2]}J_\lambda^{1-(p/n)}\pi(\lambda)\pi(\theta)d\theta \qquad (۴)$$

متناهی و مخالف صفر خواهد بود زیرا با توجه به مفروضات بیان شده عبارت داخل انتگـرال (۴) بـرای تقریبـا تمـام $(\theta,\lambda) \in \Theta \times \Lambda$ مثبـت و پیوسـته اسـت. بنـابراین سـره بـودن توزیـع پسین $\pi(\eta \mid z)$ نتیجه خواهد شد. لذا لازم است شـرط کرانـداری $\Theta \times \Lambda$ در انتخـاب توزیـع پیشین (۳) لحاظ گردد. از آنجا که تعیین توزیع پسین پارامترها از رابطه

$$\pi(\eta / z) \propto f(z / \eta)\frac{\pi(\lambda)\pi(\theta)}{\tau J_\lambda^{p/n}}$$

مشکل است، ابتدا توزیع پسین حاشیه‌ای β با شرطی کـردن پارامترهـای τ، θ و λ تعیـین می‌شود, که بصورت:

$$\beta \mid z,\tau,\theta,\lambda \sim N(\hat{\beta}_{\theta,\lambda}, \frac{1}{\tau}(X'\Sigma_\theta^{-1}X)^{-1}) \qquad (۵)$$

خواهد شد. سپس توزیع پسین حاشیه‌ای τ با شرطی کردن پارامتر θ و λ بصورت:

$$\tau \mid z,\theta,\lambda \sim Gamma(\frac{n-p}{2}, \frac{2}{q_{\theta,\lambda}}) \qquad (۶)$$

بدست می‌آید. در نهایت توزیع پسین حاشیه‌ای (θ,λ) بصورت:

$$\pi(\theta,\lambda \mid z) \propto \frac{f(z \mid \eta)\pi(\lambda)\pi(\theta)}{\pi(\beta \mid z,\tau,\theta,\lambda)\pi(\tau \mid z,\theta,\lambda)J_\lambda^{p/n}} \qquad (۷)$$

$$= |\Sigma_\theta|^{-1/2}|X'\Sigma_\theta^{-1}X|\, q_{\theta,\lambda}^{-(n-p)/2}J_\lambda^{1-(p/n)}\pi(\lambda)\pi(\theta)$$

تعیین می‌شود. اکنون براساس قاعده بیز و با جایگذاری توزیعهای پـسین حاشـیه‌ای (۵)، (۶) و (۷) در رابطه

$$\pi(\eta \mid z) = \pi(\beta,\tau,\theta,\lambda \mid z) = \pi(\beta \mid z,\tau,\theta.\lambda)\pi(\tau \mid z,\theta,\lambda)\pi(\theta,\lambda \mid z) \qquad (۸)$$

توزیع پسین پارامترهای مدل قابل تعیین می‌باشد. از آنجا که عموما تعیین $\pi(\theta,\lambda \mid z)$ و بـه تبع آن توزیع پسین (۸) بسیار دشوار است، با استفاده از روشهای مونت کـارلوی زنجیرمـارکف

(MCMC: Monte Carlo Markov Chain)، یک نمونه تصادفی از توزیعهای (۷) و (۸) بدست آورده می‌شود و استنباط بیزی پارامترهای مدل براساس این نمونه صورت می‌پذیرد (Cowles,2003). نظربه اینکه تعیین توزیع پیشگوی بهینه $Z(t_0)$ از انتگرال $p+q+2$ گانه:

$$f(z(t_0)\mid z)=\int f(z(t_0)\mid z,\eta)\pi(\eta\mid z)d\eta \qquad (۹)$$

دشوار است، با انتگرال‌گیری از $\pi(\beta,\tau\mid z,\theta,\lambda)f(z\mid\eta)$ نسبت به β و τ

$$g_\lambda(Z(t_0))\mid z,\theta,\lambda\sim t_{n-p}(a_{\theta,\lambda},b_{\theta,\lambda}^{-1})$$

است (Mohammadzadeh and Jafari, 2003)، که در آن:

$$a_{\theta,\lambda}=r_\theta'\Sigma_\theta^{-1}g_\lambda(z)+(f'(t_0)-r_\theta'\Sigma_\theta^{-1}X)'\hat{\beta}_{\theta,\lambda}$$

$$b_{\theta,\lambda}=q_{\theta,\lambda}(1-r_\theta'\Sigma_\theta^{-1}r_\theta+(f'(t_0)-r_\theta'\Sigma_\theta^{-1}X)'(X\Sigma_\theta^{-1}X)^{-1}(f'(t_0)-r_\theta'\Sigma_\theta^{-1}X))$$

بنابراین:

$$f(z(t_0)\mid z,\theta,\lambda)=\frac{\Gamma(\frac{n-p+k}{2})\mid g_\lambda'(z(t_0))\mid}{\Gamma(\frac{n-p}{2})\pi^{1/2}\mid b_{\theta,\lambda}\mid^{1/2}}[1+\frac{(g_\lambda(z(t_0))-a_{\theta,\lambda})^2}{b_{\theta,\lambda}}]^{(n-p+k)/2} \qquad (۱۰)$$

می‌باشد. از طرف دیگر بر اساس رابطه (۷) داریم:

$$f(z\mid\theta,\lambda)\propto\mid\Sigma_\theta\mid^{-1/2}\mid X\Sigma_\theta^{-1}X\mid^{-1/2}q_{\theta,\lambda}^{-(n-p)/2}J_\lambda^{1-(p/n)} \qquad (۱۱)$$

اکنون با استفاده از روابط (۱۰) و (۱۱) می‌توان توزیع پیشگوی بهینه (۹) را بصورت:

$$f(z(t_0)\mid z)=\int_\Lambda\int_\Theta\pi(\theta,\lambda\mid z)f(z(t_0)\mid z,\theta,\lambda)d\theta d\lambda$$
$$=\frac{\int_\Lambda\int_\Theta f(z(t_0)\mid z,\theta,\lambda)f(z\mid\theta,\lambda)\pi(\theta)\pi(\lambda)d\theta d\lambda}{\int_\Lambda\int_\Theta f(z\mid\theta,\lambda)\pi(\theta)\pi(\lambda)d\theta d\lambda}$$

نوشت، که یک انتگرال $q+1$ گانه و تعیین آن ساده‌تر از رابطه (۹) می‌باشد. با این وجود چون محاسبه تحلیلی آن هنوز امکان پذیر نیست، روش مونت کارلو برای تقریب آن مورد استفاده قرار می‌گیرد. بدین منظور، از توزیعهای پیشین $\pi(\theta)$ و $\pi(\lambda)$ یک نمونه تصادفی i.i.d به حجم M بصورت $\{\theta_i,\lambda_i\}_{i=1}^M$ تولید می‌شود و برای مقادیر به اندازه کافی بزرگ M می‌توان تقریب

$$f(z(t_0)\,|\,z) \approx \frac{\sum_{i=1}^{M} f(z(t_0)\,|\,z,\theta_i,\lambda_i)\,f(z\,|\,\theta_i,\lambda_i)}{\sum_{i=1}^{M} f(z\,|\,\theta_i,\lambda_i)} \qquad (۱۲)$$

را برای توزیع پیشگوی بهینه بدست آورد. هر اندازه مقدار M بزرگتر باشد، تقریب دقیق‌تر خواهد بود (Chen and *et al.*, 2002). اکنون براساس (۱۲) و یک تابع زیان داده شده، پیشگوی بهینه تقریبی و واریانس آن تعیین می‌شود. بعنوان مثال، برای زیان درجه دوم پیشگوی بهینه تقریبی برای $Z(t_0)$ و میانگین مجذور خطای تقریبی آن بترتیب بصورت:

$$\hat{Z}(t_0) = E(Z(t_0)\,|\,z) \approx \frac{\sum_{i=1}^{M} E(Z(t_0)\,|\,z,\theta_i,\lambda_i)\,f(z\,|\,\theta,\lambda_{ii})}{\sum_{i=1}^{M} f(z\,|\,\theta_i,\lambda_i)} \qquad (۱۳)$$

$$\sigma^2(t_0) = Var(Z(t_0)\,|\,z)$$
$$\approx \frac{\sum_{i=1}^{M}\{Var(Z(t_0)\,|\,z,\theta_i,\lambda_i) + [E(Z(t_0)\,|\,z,\theta_i,\lambda_i) - E(Z(t_0)\,|\,z)]^2\}\,f(z\,|\,\theta_i,\lambda_i)}{\sum_{i=1}^{M} f(z\,|\,\theta_i,\lambda_i)} \qquad (۱۴)$$

هستند، که با استفاده از آنها انجام استنباط آماری برای $Z(t_0)$ نیز میسر می‌شود. بعنوان مثال یک فاصله پیشگویی ۹۵% تقریبی برای $Z(t_0)$ بصورت $\hat{Z}(t_0) \pm 1.96\sigma(t_0)$ می‌باشد. از آنجا که ممکن است برای بعضی تبدیلات $E(Z(t_0)\,|\,z,\theta,\lambda)$ و بدنبال آن $E(Z(t_0)\,|\,z)$ موجود نباشد، معمولا تابع زیان قدرمطلق خطا مورد استفاده قرار می‌گیرد، که در اینصورت پیشگوی بهینه برابر $\hat{Z}(t_0) = Median[Z(t_0)\,|\,z]$ خواهد بود.

۴ مثال کاربردی

در این بخش، نحوه تعیین توزیع پیشگو و پیشگوی بهینه و ارزیابی آنها در یک مثال عددی بکار گرفته می‌شود. جدول ۱ میزان بارندگی دوازدهمین هفته سال ۱۹۹۱، در ۲۴ موقعیت معین از منطقه‌ای به ابعاد 12×12 کیلومتر مربع واقع در ناحیه داروین استرالیا (Keenan and Manton, 1996) را نشان می‌دهد. همانطوریکه در شکل ۱ ملاحظه می‌شود هیستوگرام مشاهدات نامتقارن و چوله به راست می‌باشد، بنابراین توزیع آنها نمی‌تواند گاوسی باشد. برای پیشگویی فضایی دریک موقعیت دلخواه، خانواده تبدیلات باکس- کاکس، یعنی:

$$g_\lambda(z) = \begin{cases} \dfrac{z^\lambda - 1}{\lambda} & \lambda \neq 0 \\ Ln(z) & \lambda = 0 \end{cases}$$

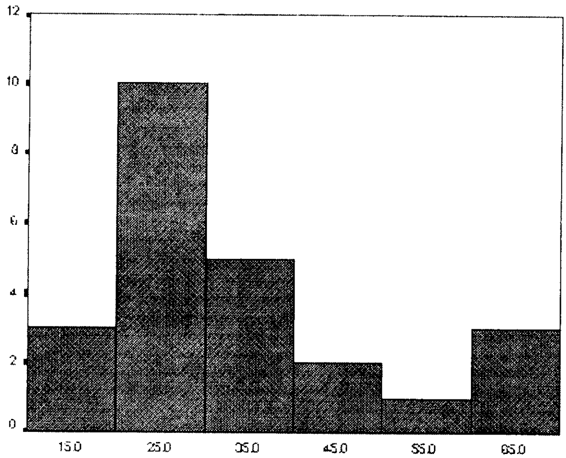
را با میانگین ثابت $E(Y(t)) = \beta_0$ و تابع همبستگی همسانگرد(Isotropic) نمایی:

$$\rho(s,t;\theta) = \exp\{-a \| t-s \|^{\theta_2})\} = \theta_1^{\|t-s\|^{\theta_2}}$$

که در آن $a > 0$، $\theta_1 = e^{-a} \in (0,1)$ و $\theta_2 \in (0,2]$ پارامترهای نامعلوم هستند، در نظر می‌گیریم. چون اطلاعاتی از قبل در مورد پارامترهای مدل و تبدیل در اختیار نیست، از توزیعهای پیشین $\theta_1 \sim U(0,1)$، $\theta_2 \sim U(0,2]$ و $\lambda \sim U(-3,3)$ استفاده می‌شود، که در این صورت توزیع پسین سره خواهد بود. با انجام روش مونت کارلو با $M = 500$ بار تکرار، توزیع پیشگوی بیزی (۱۳) و بدنبال آن پیشگوی فضایی بیزی در موقعیت $t_0 = (6,5)$، بصورت $\hat{Z}(t_0) = Median[Z(t_0) \mid z] = 26.5$ محاسبه می‌شود. با تعیین مقادیر پیشگوی فضایی بیزی برای هر کدام از موقعیتهای نمونه بر اساس۲۳ مشاهده دیگر، معیار میانگین مجذور خطای پیشبینی بروش اعتبار متقابل (Cross-Validation)، یعنی $MSEP_1 = \frac{\sum_{i=1}^{24} (z(t_i) - \hat{z}(t_i))^2}{24} = 49.10$، خواهد شد. چنانچه از $\bar{Z}$ برای پیشگویی استفاده شود، آنگاه این معیار $MSEP_2 = 217.238$ می‌شود که در مقایسه با $MSPE_1$، مقدار بسیار بزرگی است. بنابراین در صورت استفاده از روشهای ارائه شده در این مقاله، شاهد بهبود دقت پیشگویی خواهیم بود.

جدول۱ : موقعیتها و میزان بارندگی

(x , y)	Z(x,y)	(x , y)	Z(x,y)	(x , y)	Z(x,y)	(x , y)	Z(x,y)
(4.5 , 10.5)	29.92	(2 , 8)	68.25	(5 , 5)	33.68	(11 , 2)	31.27
(7 , 10.5)	23.60	(1 , 6)	63.10	(6 , 1)	29.55	(11 , 6)	23.69
(6 , 10)	21.70	(3 , 3)	54.35	(6 , 7)	24.47	(9.5 , 3)	22.82
(3 , 9)	31.14	(3.5 , 6)	46.06	(8 , 3)	33.98	(7 , 5)	19.31
(9.5 , 9.5)	17.72	(4 , 8)	26.83	(9 , 1)	41.85	(7 , 8)	18.24
(2 , 5)	66.76	(5 , 2.5)	26.76	(12 , 8)	28.98	(6.5 , 4)	30.57



شکل ۱: نمودار مستطیلی داده‌ها

۵ بحث و نتیجه‌گیری

در حالتیکه تبدیلی غیر خطی و پارامتری از میدان تصادفی مورد نظر گاوسی باشد، با توجه به مسائل مبتلابه روشهای کلاسیک پیشگویی فضایی، رهیافت بیزی برای تعیین پیشگوی فضایی اتخاذ گردید، تا با در نظر گرفتن کلیه مقادیر ممکن پارامتر تبدیل در پیشگویی، امکان محاسبه پیشگوی بهینه و میانگین مجذور خطای آن فراهم شود. براین اساس با تجزیه و تحلیل داده‌های مربوط به میزان بارندگی، پیشگوی فضایی بیزی تعیین و مناسب بودن آن مورد بررسی قرار گرفت. دراین مثال خانواده تبدیلات باکس- کاکس انتخاب شد. اما در حالت کلی چگونگی انتخاب یک خانواده تبدیلات و میزان تاثیر این انتخاب بر دقت پیشگو، از مسائل مهم در این خصوص بشمار می‌رود.

References

Banerjee, S., and Gelfand, A.E. (2002) *Prediction, Interpolation and Regression for Spatially Misaligned Data*. Sankaya, **64(A),** 227-245.

Berger, J.O., Oliveira, V.D., and Sanso, B. (2001) *Objective Bayesian Analysis of Spatially Correlated Data*. JASA **96**, 1351-1370.

Box, G.E.P., and Cox, D.R. (1964) *An Analysis of Transformation*. Journal of the Royal Statistical Society **26(B),** 211-252.

Brown, P.J., Le, N.D., and Zidek, J.V. (1995) *Multivariate Spatial Interpolation and Exposure to Air Pollutants.* The Canadian Journal of Statistics. **22**, 489-509.

Chen, M., Shao, Q., and Ibrahim, J.G. (2002) *Monte Carlo Methods in Bayesian Computation.* Springer Series in Statistics.

Chritensen, O.F., Diggle, P.J., and Ribeiro, P.J. (2001) *Analyzing Positive-Valued Spatial Data: The Transformed Gaussian Model.* Geostatistics for Environmental Applications. **28**, 287-298.

Cowles, M.K. (2003) *Efficient Model Fitting and Model Comparison for High-Dimensional Bayesian Geostatistical Models.* Journal Of Statistical Planning and Infrrence. (Article in Press).

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

Ecker, M.D. and Gelfand, A.E. (1997) *Bayesian Variogram Modeling for an Isotropic Spatial Process.* J. Agric. Biol. Environ. Stat. **2**, 347-369.

Fuentes, M., and Smith, R.L. (2002) *A New Class of Nonstationary Spatial Models.* Technical Report, North Carolina State University.

Gaudard, M., Karson, M., Linder, E., Sinha, D. (1999) *Bayesian Spatial Prediction.* Environmental and Ecological Statistics **6**, 147-171.

Handcock, M., and Stein, M. (1993) *A Bayesian Analysis of Kriging.* Technometrics **35**, 403–410.

Handcock, M., and Wallis, J. (1994) *An Approach to Statistical Spatial-Temporal Modelling of Metereological Fields (with Discussion).* J. Amer. Statist. Assoc., **89**, 368-390.

Hyoung, M. K., and Bani, K.M. (2003) *A Bayesian Prediction Using the Skew Gaussian Distribution. Statistical Planning and Inference* (Arrticle in Press).

Hinkley, D.V., and Runger, G. (1984) *The Analysis of Transformed Data.* JASA. **79**, 302- 320.

Keenan, T.D., and Manton, M.J. (1996) *Darwin Climate Monitoring and Research Station, Bureau of Meteorology Research Center.* Research Report **53**, Victoria, Australia.

Kitanidis, P.K. (1986) *Parameter Uncertainty in Estimation of Spatial Functions: Bayesian Analysis.* Water Resources Research, **22**, 499-507.

Le, N.D., and Zidek, J.V. (1992) *Interpolation with Uncertain Spatial Covariance: a Bayesian Alternative to Kriging*. J. Multivariate Anal., **43**, 351-374.

Mohammadzadeh, M., and Jafari, K.M. (2003) *Bayesian Spatial Prediction for Gaussian Random Field.* To appear in Proceedings of the Fourth Seminar of Probability and Stochastic Process, Gorgan, Iran.

Pericchi, L.R. (1981) *A Bayesian Approach to Transformation to Normality*, Biometrika, **68**, 35-43.

Stein, M.L. (1999) *Interpolation of Spatial Data.* New York. Environmetrics, **9**, 445-457.

Sun, W. (1998) Comparison of a Cokriging Method with a Bayesian Background. Water Resources Research. **36**, 218-235.