

ساخت مدل پیشگوی ناپارامتری برای میدان فضایی با استفاده از قضیه تصویر

اسحاق الماسی^۱، مهدی امید^۲

^۱گروه آمار، دانشکده علوم پایه، دانشگاه رازی

^۲گروه ریاضی، دانشکده علوم پایه، دانشگاه ایلام

تاریخ دریافت: ۱۴۰۰/۰۱/۱۹ تاریخ پذیرش و انتشار: ۱۴۰۰/۰۶/۰۶

چکیده: تعیین بهترین پیشگوی فضایی برای مقادیر گمشده یکی از مسائل مهم در آمار فضایی به شمار می‌رود. در این راستا روش‌های مختلفی مطرح شده است که هر کدام از آن‌ها دارای مزیت و محدودیت‌هایی در کاربرد هستند. بر اساس روش کریگیدن بهترین پیشگوی خطی به دست می‌آید، اما این روش برای میدان تصادفی گاوسی مناسب است. نامشخص بودن توزیع میدان تصادفی، محققین را ملزم به استفاده از روش‌هایی می‌کند که بر اساس آن‌ها امکان پیشگویی ناگوسی میسر شود. در این مقاله با استفاده از قضیه تصویر یک روش ناپارامتری برای پیشگویی میدان تصادفی ارایه می‌شود و بر مبنای آن پیشگوی میدان ناگوسی بر اساس نزدیکترین همسایه‌ها معرفی می‌شود. در ادامه در یک مطالعه شبیه‌سازی میزان دقت این روش مورد ارزیابی قرار می‌گیرد. در پایان نیز نحوه کاربست روش معرفی شده در پیشگویی داده‌های بارندگی در استان خوزستان نشان داده می‌شود.

واژه‌های کلیدی: پیشگویی ناگوسی، میدان تصادفی، کریگیدن، قضیه تصویر.

۱ مقدمه

در آمار فضایی پیشگویی مقادیر گمشده معمولاً با فرض گاوسی بودن میدان تصادفی یا تبدیلی از آن انجام می‌شود. این فرض باعث می‌شود که استفاده از پیشگوهای خطی مانند کریکینگ (کرسی، ۱۹۹۳) امکان‌پذیر باشد. در برخی شرایط ممکن است این فرض واقع‌گرایانه نباشد، از این رو لازم است از روش‌های جایگزینی استفاده شود که بر اساس آن‌ها بتوان برای میدان‌های ناگاوسی پیشگویی انجام داد. یکی از روش‌های پیشگویی در میدان‌های ناگاوسی استفاده از تابع مفصل (نلسن، ۲۰۰۶) است، که در آن توزیع تحقق‌های میدان تصادفی با استفاده از محاسبات پیچیده و زمان‌بر به دست می‌آید. **گرالر و پبسا (۲۰۱۱)**، **امیدی و محمدزاده (۲۰۱۸)** و **جمشیدی و شمس (۱۳۹۸)** با استفاده از توابع مفصل پیشگویی ناگاوسی را برای میدان تصادفی مطرح کردند. وجود محدودیت‌هایی از قبیل حجم محاسبه بالا، وجود توزیع‌های چند متغیره شرطی و نسبی بودن این توابع از جمله مشکلات استفاده از این روش است.

براکول و دیویس (۱۹۹۱) پیشگویی در سری‌های زمانی را با استفاده از قضیه تصویر مطرح کردند، **الماسی و همکاران (۲۰۱۷)** با استفاده از قضیه تصویر روشی را برای پیش‌بینی آماره‌های ترتیبی سانسور شده معرفی نمودند. در این راستا **الماسی و همکاران (۲۰۲۰)** مدلی برای پیشگویی مقادیر گمشده در حالت چندمتغیره و با کاربرد آن در رگرسیون چندمتغیره با استفاده از قضیه تصویر ارائه دادند. در این مقاله بر اساس نزدیکترین همسایه‌ها به نقاط گمشده، مدلی ناپارامتری برای پیشگویی فضایی نقاط گمشده ارائه می‌شود. در بخش ۲، پیشگوی فضایی مورد بررسی قرار گرفته و ساختار همبستگی برای این داده‌ها مرور می‌شود. در بخش ۳، با استفاده از قضیه تصویر تابع پیشگو به صورت ناپارامتری برای میدان تصادفی مطرح می‌شود. در بخش ۴، با استفاده از شبیه‌سازی برای تابع کواریانس دلخواه دقت و صحت پیشگوی تعریف شده برای نزدیکترین همسایگی‌های دو، سه و چهار با استفاده از معیار $RMSE$ بررسی شده و در بخش ۵ کاربست آن‌ها در پیشگویی داده‌های بارش در استان خوزستان مورد بررسی قرار می‌گیرد.

۲ ساختار همبستگی میدان تصادفی

در آمار فضایی پیشگویی مقدار نامعلوم میدان تصادفی در مکان‌های مشخص معمولاً با فرض گاوسی بودن میدان تصادفی انجام می‌شود. تحت این فرض بهترین پیشگوی خطی ناریب به عنوان پیشگوی کریگیدن به دست می‌آید. توابع میانگین و کواریانس فضایی میدان تصادفی $Z = \{Z(s); s \in D \subset R^d\}$ ، که

در آن $d \geq 1$ ، به ترتیب به صورت

$$\begin{aligned}\mu(s) &= E(Z(s)); \quad s \in D \\ C(s, s') &= Cov(Z(s), Z(s')); \quad s, s' \in D\end{aligned}$$

تعریف می‌شوند. میدان تصادفی $Z(\cdot)$ ، مانای مرتبه دوم است هرگاه میانگین آن نسبت به s ثابت باشد، یعنی $\mu(s) = \mu$ و کواریانس فضایی تابعی از فاصله موقعیت‌ها به صورت $Cov(Z(s), Z(s')) = C(h)$ باشد، که در آن $h = s - s'$ تأخیر فضایی نامیده می‌شود. اگر تابع کواریانس فقط تابعی از اندازه تأخیرها بوده و به جهت آن‌ها بستگی نداشته باشد، همسانگرد نیز نامیده می‌شود. تغییرنگار فضایی به صورت $\gamma(s, s') = Var[Z(s) - Z(s')]$ تعریف می‌شود. چنانچه میانگین ثابت و تغییرنگار تابعی از تأخیر باشد، یعنی $\gamma(s, s') = \gamma(h)$ ، میدان مانای ذاتی نامیده می‌شود. تحت این فرض تابع تغییرنگار به صورت

$$2\gamma(h) = 2[C(\circ) - C(h)] \quad (1)$$

با تابع کواریانس فضایی در ارتباط است. توابع کواریانس فضایی ساختار همبستگی داده‌های فضایی را تفسیر نموده و خواص آن‌ها به گونه‌ای است که همواره اندازه همبستگی مثبت را به دست می‌دهند و مقدار آن برای مشاهداتی با فاصله نزدیک حداکثر و برای فواصل دورتر کم می‌شود. **امیدی و محمدزاده (۲۰۱۶)** روش‌هایی برای ساخت توابع کواریانس فضایی ارایه دادند و بر اساس آن توابع مختلف همیشه مثبت را به دست آوردند. به عنوان مثال تابع

$$C(h) = \sigma^2 \exp(-\alpha h^\beta) \quad (2)$$

برای σ و α مثبت و $\beta \in [0, 2]$ کواریانس فضایی معتبر است. هدف پیشگویی مقدار میدان تصادفی $Z(s)$ در یک موقعیت مشخص s_ℓ بر اساس تحقق‌های میدان تصادفی در موقعیت‌های s_1, \dots, s_n ، یعنی بردار مشاهدات $z = (z(s_1), \dots, z(s_n))$ است.

۳ قضیه تصویر و پیشگوهای ناپارامتری

برای میدان تصادفی $Z(s_\ell)$ ، فضای هیلبرت $L^2 = \{Z(s_\ell) : E(Z^2(s_\ell)) < \infty\}$ با ضرب داخلی $\langle Z(s_t), Z(s_k) \rangle = E(Z(s_t)Z(s_k))$ ، را در نظر بگیرید، که در آن $Z(s_t), Z(s_k) \in L^2$.

تعریف ۱. اگر \mathcal{M} یک زیر فضای بسته از L^2 شامل تمام توابع براساس $Z(s_1), \dots, Z(s_n)$ باشد، آن‌گاه تصویر $Z(s_\ell)$ روی \mathcal{M} به صورت $E_{\mathcal{M}}(z(s_\ell)) = E(Z(s_\ell)|z(s_1), \dots, z(s_n))$ تعریف می‌شود و بهترین پیشگو برای $Z(s_\ell)$ روی فضای \mathcal{M} تحت تابع زیان درجه دوم است. این پیشگو در حالت کلی به دلیل عدم اطلاع از توزیع توام تحقق‌های میدان تصادفی و پیچیدگی محاسبات سخت و زمان‌بر است. از این‌رو توجه خود را به زیر فضای توابع خطی از \mathcal{M} معطوف می‌کنیم.

تعریف ۲. فرض کنید $Z(s_1), \dots, Z(s_n)$ مشاهدات میدان تصادفی $Z(s_\ell) \in L^2$ باشند، در این صورت تصویر $Z(s_\ell)$ در زیر فضای بسته توابع خطی زیر مجموعه \mathcal{M} ، یعنی $\{1, Z(s_1), \dots, Z(s_n)\}$ ، به صورت $P_{s\bar{p}}(Z(s_\ell)) = \sum_{i=0}^n a_i Z(s_i)$ تعریف می‌شود، که در آن $Z(s_0) = 1$ و ضرایب a_i از حل همزمان معادلات زیر به دست می‌آیند.

$$\sum_{i=0}^n a_i E(Z(s_i)Z(s_j)) = E(Z(s_\ell)Z(s_j)), \quad j = 0, 1, \dots, n. \quad (۳)$$

امید شرطی، بر اساس قضیه تصویر، بهترین پیشگوی خطی برای $Z(s_\ell)$ تحت تابع زیان میانگین توان دوم خطا روی \mathcal{M} است (براکول و دیویس، ۱۹۹۱). الماسی و همکاران (۲۰۱۷) با استفاده از قضیه تصویر یک پیشگوی خطی برای آماره‌های ترتیبی سانسور شده به دست آوردند. این پیشگو با توجه به خاصیت مارکوفی آماره‌های مرتب و تنها بر اساس دو آماره ترتیبی نزدیک به آماره سانسور شده تعیین شد. هم‌چنین الماسی و همکاران (۲۰۲۰) با استفاده از قضیه تصویر مدلی برای بازسازی مقادیر گمشده در مدل‌های رگرسیون معرفی کردند. در این بخش این مدل به مدلی برای پیشگویی ناپارامتری فضایی تعمیم داده می‌شود. وجود همبستگی بالا میان داده‌های مرتبط با مکان مشاهده نشده با نزدیکترین همسایه‌های آن باعث افزایش دقت پیشگویی بر اساس قضیه تصویر شده که از جمله مزایای این تعمیم نسبت به مدل‌های قبلی است. با توجه به ساختار همبستگی فضایی داده‌ها، که در آن‌ها مکان‌های نزدیکتر دارای همبستگی بیشتر هستند، بهترین پیشگوی خطی با حل معادلات (۳) بر اساس دو، سه و چهار نزدیکترین همسایه به دست می‌آید.

۳.۱ بهترین پیشگو بر اساس نزدیکترین همسایگی‌ها

با توجه به ساختار همبستگی فضایی داده‌ها که در آن نقاط نزدیک دارای همبستگی بیشتری هستند، می‌توان بهترین پیشگوی خطی را بر اساس نزدیکترین نقاط همسایه تعریف کرد. قضیه تصویر این امکان را فراهم می‌کند که بهترین پیشگوی خطی بر اساس تعداد محدودی از نقاط همسایگی تعیین شود، به گونه‌ای که این محدودسازی امکان پیشگویی به صورت تحلیلی را برای تحلیل و تفسیر راحت‌تر فراهم می‌آورد.

فرض کنید $Z(s_\ell)$ داده در مکان s_ℓ فاقد مشاهده باشد. اگر $(s_{(i)}, z(s_{(i)}))$ - i امین مشاهده نزدیک به $Z(s_\ell)$ بر اساس فاصله اقلیدسی باشد، آنگاه طبق قضیه تصویر بهترین تابع پیشگوی خطی بر پایه $z(s_{(1)}), \dots, z(s_{(n)})$ به صورت $\hat{Z}(s_\ell) = \sum_{i=0}^n a_i z(s_{(i)})$ است، که در آن $z(s_{(0)}) = 1$ و ضرایب a_0, \dots, a_n از حل همزمان معادلات زیر به دست می‌آیند.

$$\sum_{i=0}^n a_i E(Z(s_{(i)})Z(s_{(j)})) = E(Z(s_\ell)Z(s_{(j)})), \quad j = 0, 1, \dots, n. \quad (4)$$

فرع ۰.۱ بهترین پیشگوی خطی $Z(s_\ell)$ بر اساس چهار نزدیکترین همسایه $(s_{(1)}, z(s_{(1)}))$ ، $(s_{(2)}, z(s_{(2)}))$ و $(s_{(3)}, z(s_{(3)}))$ و $(s_{(4)}, z(s_{(4)}))$ به صورت

$$\hat{Z}(s_\ell) = \sum_{i=0}^4 a_i z(s_{(i)}) \quad (5)$$

است، که ضرایب آن به صورت

$$\begin{aligned} a_0 &= \mu(s_\ell) - a_1 \mu(s_{(1)}) - a_2 \mu(s_{(2)}) - a_3 \mu(s_{(3)}) - a_4 \mu(s_{(4)}) \\ a_1 &= \frac{\varphi(\ell, 1; 2, 3, 4)}{\varphi(1, 2, 3, 4)} \sqrt{\frac{\text{Var}(z(s_\ell))}{\text{Var}(z(s_{(1)}))}} \\ a_2 &= \frac{\varphi(\ell, 2; 1, 3, 4)}{\varphi(1, 2, 3, 4)} \sqrt{\frac{\text{Var}(z(s_\ell))}{\text{Var}(z(s_{(2)}))}} \\ a_3 &= \frac{\varphi(\ell, 3; 1, 2, 4)}{\varphi(1, 2, 3, 4)} \sqrt{\frac{\text{Var}(z(s_\ell))}{\text{Var}(z(s_{(3)}))}} \\ a_4 &= \frac{\varphi(\ell, 4; 1, 2, 3)}{\varphi(1, 2, 3, 4)} \sqrt{\frac{\text{Var}(z(s_\ell))}{\text{Var}(z(s_{(4)}))}} \end{aligned}$$

به دست می‌آیند، که در آن‌ها $\varphi(\cdot)$ توابعی از همبستگی فضایی به صورت

$$\begin{aligned} \varphi(1, 2, 3, 4) = & 1 - \rho_{1,2}^{\check{}} - \rho_{1,3}^{\check{}} - \rho_{1,4}^{\check{}} - \rho_{2,3}^{\check{}} - \rho_{2,4}^{\check{}} - \rho_{3,4}^{\check{}} + \rho_{1,3}^{\check{}}\rho_{2,4}^{\check{}} \\ & + \rho_{1,4}^{\check{}}\rho_{2,3}^{\check{}} + \rho_{1,2}^{\check{}}\rho_{3,4}^{\check{}} + 2\rho_{1,2}^{\check{}}\rho_{1,3}^{\check{}}\rho_{2,3}^{\check{}} + 2\rho_{1,2}^{\check{}}\rho_{1,4}^{\check{}}\rho_{2,4}^{\check{}} + 2\rho_{1,3}^{\check{}}\rho_{1,4}^{\check{}}\rho_{3,4}^{\check{}} \\ & + 2\rho_{2,3}^{\check{}}\rho_{2,4}^{\check{}}\rho_{3,4}^{\check{}} - 2\rho_{1,2}^{\check{}}\rho_{2,3}^{\check{}}\rho_{3,4}^{\check{}}\rho_{1,4}^{\check{}} - 2\rho_{1,3}^{\check{}}\rho_{2,3}^{\check{}}\rho_{2,4}^{\check{}}\rho_{1,4}^{\check{}} - 2\rho_{1,2}^{\check{}}\rho_{1,3}^{\check{}}\rho_{2,4}^{\check{}}\rho_{3,4}^{\check{}} \end{aligned}$$

$$\begin{aligned} \varphi(\ell, r; k, s, m) = & \rho_{\ell,r}(\check{}}(1 - \rho_{k,m}^{\check{}} - \rho_{k,s}^{\check{}} - \rho_{s,m}^{\check{}} + 2\rho_{k,m}\rho_{k,s}\rho_{s,m}) \\ & - \rho_{\ell,k}(\rho_{r,k} - \rho_{r,k}\rho_{s,m}^{\check{}} - \rho_{r,s}\rho_{k,s} - \rho_{r,m}\rho_{k,m} + \rho_{r,m}\rho_{k,s}\rho_{s,m} + \rho_{r,s}\rho_{k,m}\rho_{s,m}) \\ & - \rho_{\ell,s}(\rho_{r,s} - \rho_{r,s}\rho_{k,m}^{\check{}} - \rho_{r,k}\rho_{k,s} - \rho_{r,m}\rho_{s,m} + \rho_{k,m}\rho_{r,m}\rho_{k,s} + \rho_{r,k}\rho_{k,m}\rho_{s,m}) \\ & - \rho_{\ell,m}(\rho_{r,m} - \rho_{r,m}\rho_{k,s}^{\check{}} - \rho_{r,s}\rho_{s,m} - \rho_{r,k}\rho_{k,m} + \rho_{r,k}\rho_{k,s}\rho_{s,m} + \rho_{r,s}\rho_{k,s}\rho_{k,m}) \end{aligned}$$

هستند. مقادیر $\rho_{\ell,i} = Corr(Z(s_{\ell}), Z(s_{(i)}))$ و $\rho_{i,j} = Corr(Z(s_{(i)}), Z(s_{(j)}))$ نیز بر اساس تابع همبستگی نگار برازش شده تعیین می‌شوند.

معادله پیشگو و ضرایب آن برای نزدیکترین دو همسایه و سه همسایه نیز به طریق مشابه و از حل همزمان معادلات (۴) به دست می‌آید. بهترین پیشگوی خطی $Z(s_{\ell})$ بر اساس دو نزدیکترین همسایه $z(s_{(1)}), z(s_{(2)})$ به صورت

$$\hat{Z}(s_{\ell}) = a_0 + a_1 z(s_{(1)}) + a_2 z(s_{(2)}) \quad (۶)$$

است، که در آن

$$\begin{aligned} a_0 &= \mu(s_{\ell}) - a_1 \mu(s_{(1)}) - a_2 \mu(s_{(2)}) \\ a_1 &= \frac{\rho_{\ell,1} - \rho_{1,2} \rho_{\ell,2}}{1 - \rho_{1,2}^{\check{}}} \sqrt{\frac{Var(z(s_{\ell}))}{Var(z(s_{(1)}))}} \\ a_2 &= \frac{\rho_{\ell,2} - \rho_{1,2} \rho_{\ell,1}}{1 - \rho_{1,2}^{\check{}}} \sqrt{\frac{Var(z(s_{\ell}))}{Var(z(s_{(2)}))}} \end{aligned}$$

بهترین پیشگوی خطی $Z(s_\ell)$ بر اساس سه نزدیکترین همسایه $z(s_{(1)}), z(s_{(2)}), z(s_{(3)})$ به صورت

$$\hat{Z}(s_\ell) = a_0 + a_1 z(s_{(1)}) + a_2 z(s_{(2)}) + a_3 z(s_{(3)}) \quad (7)$$

است، که در آن

$$\begin{aligned} a_0 &= \mu(s_\ell) - a_1 \mu(s_{(1)}) - a_2 \mu(s_{(2)}) - a_3 \mu(s_{(3)}) \\ a_1 &= \frac{\rho_{\ell,1} - \rho_{1,2} \rho_{\ell,2} - \rho_{1,3} \rho_{\ell,3} - \rho_{2,3}^2 \rho_{\ell,1} + \rho_{1,3} \rho_{2,3} \rho_{\ell,2} + \rho_{1,2} \rho_{2,3} \rho_{\ell,3}}{1 - \rho_{1,2}^2 - \rho_{1,3}^2 - \rho_{2,3}^2 + 2\rho_{1,2} \rho_{1,3} \rho_{2,3}} \\ &\times \sqrt{\frac{\text{Var}(z(s_\ell))}{\text{Var}(z(s_{(1)}))}} \\ a_2 &= \frac{\rho_{\ell,2} - \rho_{2,3} \rho_{\ell,3} - \rho_{1,2} \rho_{\ell,1} - \rho_{1,3}^2 \rho_{\ell,2} + \rho_{1,3} \rho_{2,3} \rho_{\ell,1} + \rho_{1,2} \rho_{1,3} \rho_{\ell,3}}{1 - \rho_{1,2}^2 - \rho_{1,3}^2 - \rho_{2,3}^2 + 2\rho_{1,2} \rho_{1,3} \rho_{2,3}} \\ &\times \sqrt{\frac{\text{Var}(z(s_\ell))}{\text{Var}(z(s_{(2)}))}} \\ a_3 &= \frac{\rho_{\ell,3} - \rho_{1,3} \rho_{\ell,1} - \rho_{2,3} \rho_{\ell,2} - \rho_{1,2}^2 \rho_{\ell,3} + \rho_{1,3} \rho_{1,2} \rho_{\ell,2} + \rho_{1,2} \rho_{2,3} \rho_{\ell,1}}{1 - \rho_{1,2}^2 - \rho_{1,3}^2 - \rho_{2,3}^2 + 2\rho_{1,2} \rho_{1,3} \rho_{2,3}} \\ &\times \sqrt{\frac{\text{Var}(z(s_\ell))}{\text{Var}(z(s_{(3)}))}} \end{aligned}$$

۴ مطالعه شبیه‌سازی

در این بخش با استفاده از شبیه‌سازی، دقت مدل‌های معرفی شده در روابط (۵)، (۶) و (۷) مورد بررسی قرار می‌گیرد. برای این منظور ابتدا میدانی تصادفی گاوسی به‌ازای تعداد مکان‌های ۵۰، ۱۰۰، ۵۰۰ و ۱۰۰۰ تولید شد. در ادامه برای بررسی ساختار وابستگی میدان تولید شده و همچنین تعیین مقادیر همبستگی مورد نیاز از تابع کواریانس (۲) استفاده شد. این مقادیر از طریق تابع همبستگی نگار و از تقسیم تابع کواریانس بر σ^2 به‌دست می‌آید. با در نظر گرفتن این عمل در محاسبه مقادیر همبستگی نیاز به مقدار σ^2 نیست. برای α مقادیر ۱، ۵، ۱۰ و ۲۰ و برای β مقادیر ۰/۵، ۱ و ۱/۵ در نظر گرفته شد.

به منظور کنترل اثر تصادفی داده‌های تولید شده، تولید میدان تصادفی در هر حالت هزار بار تکرار و با استفاده از جذر میانگین توان دوم خطا ($RMSE$) دقت توابع پیشگو ثبت شد. معیار ارزیابی نهایی

نیز میانگین $RMSE$ در این هزار بار اجرا در نظر گرفته شد. در جدول ۱ مقادیر پارامترها و میانگین $RMSE$ پیشگویی فضایی نقاط تولید شده در هزار بار تکرار آمده است. همانطور که ملاحظه می‌شود مقادیر میانگین $RMSE$ همگی نزدیک به ۱ است که به معنی زیاد بودن دقت پیشگویی است. همچنین نتایج نشان می‌دهد که افزایش تعداد نقاط همسایگی (n_b) تاثیر بسیار کمی در دقت پیشگویی بر اساس روش‌های معرفی شده دارد. این نتیجه برای مقادیر مختلف α و β یکسان است.

۵ مثال کاربردی

به منظور کاربری مدل‌های ناپارامتری به دست آمده در روابط (۵) تا (۷)، از داده‌های بارندگی در استان خوزستان که توسط **امیدی و محمدزاده (۲۰۱۸)** به‌کار برده شدند، استفاده شد. در این مطالعه تابع تغییرنگار کروی به عنوان بهترین تابع برای بررسی ساختار همبستگی فضایی داده‌ها انتخاب و با کاربری توابع مفصل توزیع تحقق میدان تصادفی برای پیشگویی ناگوسی داده‌های مشاهده نشده به دست آمد. تابع نیم‌تغییرنگار کروی به صورت $\gamma = b + C \cdot (1.5 \frac{h}{r} - 0.5 (\frac{h}{r})^3)$ است، که در آن C ، b و r به ترتیب اثر قطعه‌ای، ازاره و دامنه هستند. برای داده‌های بارش در استان خوزستان برآورد پارامترهای دامنه، اثر قطعه‌ای و ازاره به ترتیب با مقادیر $73/716$ ، $1/771$ و $31/650$ به دست آمد که برازش مناسب‌تری را به همبستگی فضایی داده‌ها نسبت به توابع تغییرنگار گاوسی و نمایی ارائه دادند. با استفاده از این تابع تغییرنگار و رابطه (۱)، تابع کواریانس فضایی و به تبع آن تابع همبستگی نگار برای داده‌ها به دست آمد.

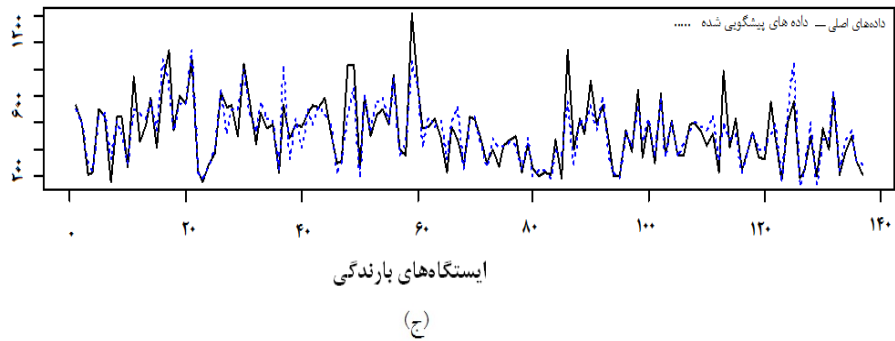
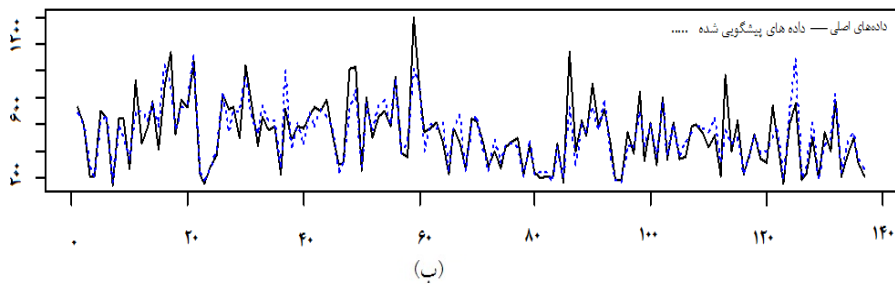
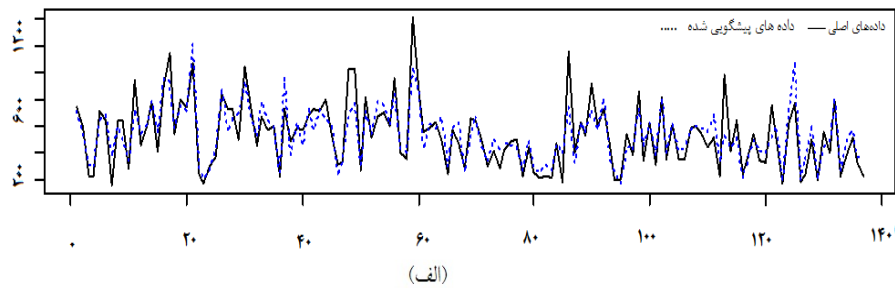
با در نظر گرفتن توابع همبستگی نگار و با به کارگیری روابط (۵) تا (۷) مقدار پیشگویی برای داده‌ها محاسبه که در جدول ۲ میانگین و انحراف معیار برای داده‌های اصلی و مقادیر پیشگویی شده خلاصه شده است. در شکل ۱ نیز نمودار خطی مربوط به داده‌های اصلی و نقاط پیشگویی شده به تفکیک دو، سه و چهار نزدیکترین همسایه آمده است. همانطور که ملاحظه می‌شود مقادیر به دست آمده به طور نسبی دقت بالایی مدل‌ها را نشان می‌دهند. **امیدی و محمدزاده (۲۰۱۸)** با استفاده از تابع مفصل توزیع میدان تصادفی داده‌های بارش باران در استان خوزستان را به‌طور نسبی تعیین و بر اساس آن 20° نقطه از میدان را انتخاب و دقت پیشگویی مدل مطرح شده را برای ۴ نزدیکترین همسایه بررسی نمودند. برای این نقاط با استفاده از رابطه (۵) دقت پیشگویی برابر $4/89$ به دست آمد که دقت بیشتری را نسبت به $RMSE$ به دست آمده توسط **امیدی و محمدزاده (۲۰۱۸)** با مقدار $5/39$ به دست داد.

جدول ۱. مقدار RMSE برای داده‌های شبیه‌سازی شده

α				n_b	β	n
γ_0	γ_1	δ	λ			
۱/۰۰۰۷۱۵	۰/۹۹۴۹۹	۰/۹۵۹۹۹	۱/۰۵۴۵۲	۲	۰/۵	
۱/۰۰۰۷۱۴	۰/۹۹۳۹۶	۰/۹۵۹۱۰	۱/۰۲۵۲۶	۳		
۱/۰۰۰۷۱۲	۰/۹۹۳۴۶	۰/۹۵۸۱۱	۱/۰۱۱۰۰	۴		
۰/۹۶۳۹۲	۰/۹۴۹۳۱	۱/۱۶۳۱۵	۱/۲۳۵۶۵	۲	۱	۵۰
۰/۹۶۲۱	۰/۹۴۲۸۳	۱/۱۳۵۴۵	۱/۲۰۲۶۷	۳		
۰/۹۶۱۵۶	۰/۹۴۰۱۰	۱/۱۲۲۱۳	۱/۱۸۷۲۸	۴		
۱/۰۲۶۲۴	۱/۳۳۷۷۸	۱/۲۴۵۱۲	۱/۳۶۷۶۵	۲	۱/۵	
۱/۰۲۴۸۷	۱/۳۳۵۹۲	۱/۲۴۳۶۵	۱/۳۵۱۰۹	۳		
۱/۰۱۷۲۴	۱/۳۳۱۵۸	۱/۲۴۰۷۱	۱/۳۳۷۲۴	۴		
۱/۰۰۰۰۲۸	۰/۹۸۸۶	۰/۹۵۶۷۸	۱/۰۴۸۱۳	۲	۰/۵	
۱/۰۰۰۰۲۱	۰/۹۸۷۹۳	۰/۹۵۲۴۶	۱/۰۱۷۰۳	۳		
۱/۰۰۰۰۱۷	۰/۹۸۷۵۷	۰/۹۴۹۴۱	۱/۰۱۰۷۷	۴		
۰/۹۴۲۹۷	۰/۹۴۹۲۵	۱/۰۱۵۷۸	۱/۲۲۵۶۶	۲	۱	۱۰۰
۰/۹۳۹۹۶	۰/۹۳۹۷۸	۱/۰۱۰۳۳	۱/۱۹۳۴۱	۳		
۰/۹۳۸۵۵	۰/۹۳۵۰۰	۱/۰۰۲۱۹	۱/۱۷۷۶۵	۴		
۱/۰۲۵۹۴	۱/۲۲۵۶۴	۱/۲۴۴۳۲	۱/۳۶۶۱۲	۲	۱/۵	
۱/۰۱۹۵۱	۱/۲۱۸۴۴	۱/۲۴۱۳۰	۱/۳۴۸۳۴	۳		
۱/۰۱۷۶۹	۱/۲۱۳۵۵	۱/۲۴۰۶۹	۱/۳۴۳۲۱	۴		
۰/۹۹۳۹۱	۰/۹۶۳۲	۰/۹۴۳۱۰	۱/۰۴۵۴۱	۲	۰/۵	
۰/۹۹۳۷۳	۰/۹۵۹۶۲	۰/۹۳۸۷۷	۱/۰۱۶۶	۳		
۰/۹۹۳۶۶	۰/۹۵۷۷۱	۰/۹۳۲۷۰	۱/۰۰۷۹۹	۴		
۰/۹۴۲۳۶	۰/۹۳۴۸۲	۱/۰۱۲۴۸	۱/۱۸۳۳۹	۲	۱	۵۰۰
۰/۹۳۶۲۵	۰/۹۳۴۱۹	۱/۰۰۹۷۷	۱/۱۵۷۷۲	۳		
۰/۹۳۰۷۰	۰/۹۳۱۹۵	۱/۰۰۱۹۷	۱/۱۴۵۷۳	۴		
۱/۰۱۸۲۷	۱/۲۱۴۳۹	۱/۱۶۴۸۸	۱/۳۴۴۹۳	۲	۱/۵	
۱/۰۰۰۷۷۴	۱/۲۱۲۰۲	۱/۱۶۲۴۱	۱/۳۴۱۸۹	۳		
۱/۰۰۰۶۹۷	۱/۲۱۰۴۶	۱/۱۵۸۱۵	۱/۳۴۱۲۱	۴		
۰/۹۸۹۳۳	۰/۹۵۱۱۶	۰/۹۴۰۹۳	۱/۰۰۱۵۹	۲	۰/۵	
۰/۹۸۸۸۵	۰/۹۴۵۹۴	۰/۹۳۷۱۶	۱/۰۰۱۱۲	۳		
۰/۹۸۸۶۳	۰/۹۴۳۰۱	۰/۹۲۹۴۳	۱/۰۰۰۲۹	۴		
۰/۹۴۱۵۵	۰/۹۳۴۰۶	۰/۹۹۵۳۳	۱/۱۶۵۸۴	۲	۱	۱۰۰۰
۰/۹۳۶۱۷	۰/۹۳۲۰۹	۰/۹۸۲۴۶	۱/۱۳۹۸۳	۳		
۰/۹۳۰۷۰	۰/۹۳۱۴۱	۰/۹۷۵۶۴	۱/۱۲۷۴۶	۴		
۱/۰۱۸۰۳	۱/۲۱۱۰۹	۱/۱۵۸۵۳	۱/۳۴۱۰۹	۲	۱/۵	
۱/۰۰۰۵۹۳	۱/۲۱۰۹۱	۱/۱۴۴۵۷	۱/۳۴۰۵۳	۳		
۱/۰۰۰۴۸۲	۱/۲۱۰۴۳	۱/۱۴۴۰۱	۱/۳۴۰۴۲	۴		

جدول ۲. میانگین و انحراف معیار داده های اصلی و پیشگویی شده

میانگین	انحراف معیار	داده های اصلی
۶۲/۵۳۱	۷۶/۲۴۶	پیشگویی بر اساس ۲ همسایه
۹۲/۵۳۴	۷۸/۲۰۳	پیشگویی بر اساس ۳ همسایه
۲۱/۵۳۰	۸۸/۲۱۷	پیشگویی بر اساس ۴ همسایه
۷۲/۵۳۰	۴۵/۲۲۱	



شکل ۱. نمودار داده های اصلی و نقاط پیشگویی شده، الف- برای دو نزدیکترین همسایه، ب- برای سه نزدیکترین همسایه و ج- برای چهار نزدیکترین همسایه

بحث و نتیجه‌گیری

در این مقاله با استفاده از قضیه تصویر یک روش ناپارامتری برای پیشگویی میدان تصادفی ارایه و بر اساس آن مدل‌هایی برای پیشگویی میدان ناگاوسی بر اساس نزدیکترین دو، سه و چهار همسایه مطرح شد. برای بررسی دقت این مدل‌ها مطالعه شبیه‌سازی انجام گرفت و با استفاده از معیار جذر میانگین توان دوم خطا ($RMSE$) دقت مدل‌ها مورد مقایسه قرار گرفت. نتایج شبیه‌سازی نشان داد تغییرات در مقادیر پارامترهای تابع کواریانس فضایی و همچنین افزایش تعداد مکان‌ها تاثیر پایینی بر دقت پیشگوها دارد. در ادامه مدل‌های به دست آمده برای پیشگویی داده‌های بارندگی در استان خوزستان مورد استفاده قرار گرفت که نتایج نشان از پیشگویی مناسب این مدل‌ها بود. در پایان نیز مدل پیشگوی ناپارامتری بر اساس نزدیکترین چهار همسایه با پیشگوی ساخته شده با استفاده از تابع مفصل توسط **امیدی و محمدزاده (۲۰۱۸)** مقایسه و نتیجه نشان از برتری نسبی روش ناپارامتری بر اساس قضیه تصویر نسبت به پیشگویی بر اساس تابع مفصل دارد. همچنین از دیگر مزایای این روش می‌توان به فرمول‌بندی ساده و شکل بسته روابط جهت تفسیرپذیری و تحلیل راحت تر مدل اشاره کرد. در این روش افزایش تعداد همسایگی‌ها باعث به وجود آمدن روابط پیچیده می‌شود اما با توجه به اینکه افزایش تعداد زیاد نقاط همسایگی، مثلا از سه به چهار همسایگی، افزایش محسوسی در دقت پیشگوها ایجاد نمی‌کند، بنابراین با توجه به اصل امساک در مدل‌سازی می‌توان از به کار گرفتن بیش از چهار همسایه اجتناب کرد. اگرچه در این مقاله استفاده از نزدیکترین همسایه‌ها مورد توجه قرار گرفت اما استفاده از معیارهای دیگر مانند شعاع همسایگی و کاربرد قضیه تصویر در این مباحث می‌تواند از موضوعات جدید در تحلیل داده‌های فضایی باشد.

تقدیر و تشکر

نویسندگان مقاله از سردبیر، هیئت تحریریه، داوران و ویراستار ارجمند که با پیشنهادات و نظرات ارزنده خود موجب ارائه بهتر مقاله شدند، تقدیر و تشکر می‌نمایند. همچنین از معاونت پژوهش و فناوری دانشگاه ایلام به خاطر حمایت در انجام مقاله کمال تشکر و قدردانی را دارند.

مراجع

جمشیدی، ز.، شمس، ص. (۱۳۹۸)، تحلیل داده‌های فضایی با خانواده مفصل خی‌دو، مجله علوم آماری، ۱۳، ۳۶۳-۳۸۴.

Almasi, I., Mohammadpour, A., and Mohammadi, M. (2017), Best Linear Unbiased Interpolation of Order Statistics. *Communications in Statistics-Simulation and Computation*, **46**, 4161–4171.

Almasi, I., Mohsen Salehi, M. and Moradi, M. (2020), A New Algorithm to Impute the Missing Values in the Multivariate Case, *JIRSS*, **19**, 133-143.

Arnold, B. C., Balakrishnan, N., and Nagaraja, H. N. (1992), *A First Course in Order Statistics (Vol. 54)*. Siam.

Brockwell, P. J., Davis, R. A. (1991). *Time Series: Theory and Methods*, second ed., Springer Series in Statistics, Springer-Verlag, New York.

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

Graler, B. and Pebesma, E., (2011), The Pair Copula Construction for Spatial Data, a New Approach to Model Spatial Dependency, *Procedia Environmental Sciences*, **7**, 206-211.

Nelsen, R. B., (2006), *An Introduction to Copulas*, New York, Springer.

Omidi, M. and Mohammadzadeh, M. (2016), A New Method to Build Spatio-Temporal Covariance Functions: Analysis of Ozone Data. *Statistical Papers*. **57**, 689-703.

Omidi, M. and Mohammadzadeh, M. (2018), Spatial Interpolation Using Copula for Non-Gaussian Modeling of Rainfall Data, *JIRSS*, **17** , 165-179.

Construction a Non-parametric Prediction Model for Spatial Random Field Using Projection Theorem

Almasi¹, **I. Omidi**, **M.**²

¹Department of Statistics, Razi University, Kermanshah , Iran.

²Department of Mathematics, Ilam University, Ilam , Iran.

Abstract: Identifying the best prediction of unobserved observation is one of the most critical issues in spatial statistics. In this line, various methods have been proposed, that each one has advantages and limitations in application. Although the best linear predictor is obtained according to the Kriging method, this model is applied for the Gaussian random field. The uncertainty in the distribution of random fields makes researchers use a method that makes the nongaussian prediction possible. In this paper, using the Projection theorem, a non-parametric method is presented to predict a random field. Then some models are proposed for predicting the nongaussian random field using the nearest neighbours. Then, the accuracy and precision of the predictor will be examined using a simulation study. Finally, the application of the introduced models is examined in the prediction of rainfall data in Khuzestan province.

Keywords: Nongaussian predictor, Random field, Kriging, Projection theorem.

Mathematics Subject Classification (2010): 62H11, 62M20.