پیش‌گویی بر اساس مشاهدات در نزدیکترین همسایگی با استفاده از مدل‌های اتورگرسیو فضایی نامانا

آزاده مجیری[۱] ، یداله واقعی[۲]
[۱] گروه آمار، دانشگاه زابل
[۲] گروه آمار، دانشگاه بیرجند

چکیده: یک روش برای مدل‌بندی و تحلیل داده‌های فضایی، استفاده از مدل‌های اتورگرسیو فضایی است که بین متغیر وابسته در موقعیت‌های مختلف رابطه اتورگرسیو برقرار می‌کند. اغلب مدل‌های اتورگرسیو فضایی برای داده‌های فضایی مانا به کار می‌روند و بر اساس آن‌ها پیش‌گویی فضایی در موقعیت مشخص به‌دست می‌آید. در این مقاله، دو پیش‌گویی خطی بر اساس مشاهدات در نزدیکترین همسایگی با استفاده از مدل اتورگرسیو فضایی نامانا ارائه می‌شود سپس کاربرد روش‌های ارائه شده با استفاده از یک مجموعه داده فضایی مشبکه‌ای مورد تحلیل قرار خواهد گرفت.
واژه‌های کلیدی: پیش‌گویی، داده فضایی مشبکه‌ای، مدل اتورگرسیو فضایی نامانا.
کد موضوع‌بندی ریاضی (۲۰۱۰): 62M20، 62M30، 62M10.

۱ مقدمه

به داده‌هایی که علاوه بر مقدار متغیر مورد نظر، مختصات فضایی آن ثبت و در تحلیل‌های آماری مورد استفاده قرار می‌گیرد، داده‌های فضایی گفته می‌شود. داده‌های مشبکه‌ای نوع خاصی از داده‌های فضایی هستند که موقعیت فضایی داده‌ها به صورت ناحیه‌ای است، این مکان‌ها می‌تواند منظم یا نامنظم باشند. اگر مکان‌ها هم‌شکل و هم‌اندازه باشند ناحیه‌ها منظم و در غیر این‌صورت نامنظم‌اند (محمدزاده ، ۱۳۹۴).

اولین بار ویتل (۱۹۵۴) با فرض این که متغیرهای وابسته و مستقل بر روی یک شبکه منظم مشاهده شده باشند، وابستگی فضایی را در یک مدل اتورگرسیو فضایی (SAR) لحاظ کرد. باسو و راینسل (۱۹۹۳) خواص مدل‌های اتورگرسیو فضایی یک‌طرفه مرتبه اول را به‌دست آوردند و پارامترهای این مدل توسط ارد (۱۹۷۹) و انسلین (۱۹۸۸) به روش ماکسیمم درستنمایی (ML) و کمترین توان‌های دوم خطا (LS) برآورد گردید. همچنین باسو و راینسل (۱۹۹۲) برآورد پارامترهای

[۱] آزاده مجیری، amojirii@uoz.ac.ir

مدل SAR مرتبه اول را به روش یول واکر (YW) محاسبه کرد. پارامترهای مدل SAR مرتبه دوم ($SAR(۲, ۱)$)، به سه روش ML، LS و YW برای داده‌هایی با میانگین صفر توسط آوانگ و شیتان (۲۰۰۶) برآورد گردید.

به طور معمول با فرض مانایی میدان تصادفی پیش‌گویی بهینه از می‌نیمم کردن میانگین توان دوم خطای پیش‌گویی به‌دست آورده می‌شود که همان میانگین شرطی $Z_{k,s}$ به شرط سایر مشاهدات است (کرسی، ۱۹۹۳). باسو و راینسل (۱۹۹۳) با استفاده از مشاهدات در نزدیکترین همسایگی روشی برای پیش‌گویی درون قلمرو داده‌ها (درونیابی) پیشنهاد دادند. روش پیش‌گویی باسو و راینسل برای مدل SAR مرتبه دوم ($SAR(۲, ۱)$) توسط مجیری و همکاران (۲۰۱۸a) تعمیم داده شد و در کنار آن ایده جدیدی مبتنی بر پیش‌گویی به کمک مشاهدات ربع سوم ارائه گردید سپس مجیری و همکاران (۲۰۱۸b) روش پیش‌گویی درون قلمرو داده‌ها را با روش پیش‌گویی کریگیدن مقایسه کردند. صابر و نعمت الهی (۲۰۱۷) سه پیشگو درون قلمرو داده‌ها بر اساس مشاهدات در نزدیکترین همسایگی برای مدل SAR مرتبه اول تفکیک پذیر مانا، ارائه دادند.

یکی از مسائل مهم در آنالیز سری‌های زمانی، بررسی وجود روند و رسیدن به یک سری زمانی ماناست که با استفاده از مدلسازی روند و تفاضل‌گیری روند داده‌ها حذف می‌شود. در این مقاله دو تعمیم از مدل‌های SAR نامانا ارائه می‌شود. در روش اول، یک مدل خطی پارامتری بر حسب شماره موقعیت های طول و عرض داده‌ها برای میانگین داده‌ها در نظر گرفته شده، سپس با استفاده از روش ماکسیمم درستنمایی پارامترهای آن را برآورد کرده و یک الگوریتم برای برآورد پارامترها ارائه می‌دهیم. روش دوم بر اساس ایده مدل‌های میانگین متحرک خود انباشته (ARIMA) در سری زمانی است که با استفاده از تفاضل‌گیری نامانایی در میانگین داده‌ها از بین می‌بریم. در بخش چهارم، پیش‌گویی فضایی بر اساس مشاهدات در نزدیکترین همسایگی درون قلمرو داده‌ها برای این دو روش ارائه می‌دهیم و در پایان کاربرد روش‌های پیش‌گویی ارائه شده را با یک سری داده مشبکه‌ای منظم نشان داده شده و خطاهای پیشگویی محاسبه می‌گردد.

۲ مدل اتورگرسیو فضایی

مدل اتورگرسیو فضایی یک‌طرفه ربع سوم ($SAR(p_۱, p_۲)$) را می‌توان به‌صورت

$$\Phi(B_۱, B_۲)Z_{i,j} = \varepsilon_{i,j};\quad i = ۱, \dots, m,\ j = ۱, \dots, n, \tag{۱.۲}$$

نوشت که در آن $\Phi(B_۱, B_۲) = \sum_{k=۰}^{p_۱}\sum_{\ell=۰}^{p_۲} \alpha_{k,\ell}B_۱^kB_۲^\ell,\ \alpha_{۰,۰} = ۰$ است. با قرار دادن $p_۱ = p_۲ = ۱$ و $p_۱ = ۱$، $p_۲ = ۲$ در مدل (۱.۲) به ترتیب مدل $SAR(۱, ۱)$ و $SAR(۲, ۱)$ به‌دست می‌آید. برای ساده‌تر شدن نمادها، فرض کنید $\alpha_{۱,۰} = \alpha_۱$، $\alpha_{۰,۱} = \alpha_۲$، $\alpha_{۱,۱} = \alpha_۳$، $\alpha_{۲,۰} = \alpha_۴$ و $\alpha_{۲,۱} = \alpha_۵$، در این‌صورت برای مدل $SAR(۲, ۱)$ داریم:

$$Z_{i,j} = \alpha_۱Z_{i-۱,j} + \alpha_۲Z_{i,j-۱} + \alpha_۳Z_{i-۱,j-۱} + \alpha_۴Z_{i-۲,j} + \alpha_۵Z_{i-۲,j-۱} + \varepsilon_{i,j}. \tag{۲.۲}$$

فرض کنید داده‌های فضایی بر روی مشبکه منظم قرار بگیرند و $\{Z_{i,j};\ i = ۱, ۲, \dots, m,\ \ j = ۱, ۲, \dots n\}$ متغیرهای فضایی باشند. همچنین $\mathbf{Z} = (Z_{۱,۱}, Z_{۱,۲}, ..., Z_{۱,n}, ..., Z_{m,n})'$ و $\varepsilon = (\varepsilon_{۱,۱}, \varepsilon_{۱,۲}, ..., \varepsilon_{۱,n}, ..., \varepsilon_{m,n})'$ بردارهایی $N = m \times n$تایی باشد. آوانگ و شیتان (۲۰۰۶) مدل (۲.۲) را برای $i = ۱, \dots, m$ و $j = ۱, \dots, n$ به فرم ماتریسی

$$\mathbf{Z} = \mathbf{AZ} + \varepsilon \rightarrow (\mathbf{I}_N - \mathbf{A})\,\mathbf{Z} = \varepsilon, \tag{۳.۲}$$

نوشتند، که در آن $\mathbf{I}_N$ ماتریس همانی $N \times N$ و ماتریس پارامتری

$$\mathbf{A} = (\alpha_۱\mathbf{W}_۱ + \alpha_۲\mathbf{W}_۲ + \alpha_۳\mathbf{W}_۳ + \alpha_۴\mathbf{W}_۴ + \alpha_۵\mathbf{W}_۵)$$

است. با فرض $\mathrm{E}(\varepsilon_{i,j}) = ۰$ تحت پذیرش مانایی در مدل (۲.۲)، $\mathrm{E}(Z_{i,j}) = ۰$ نتیجه می‌شود. در صورتی‌که $\mathrm{E}\,(Z_{i,j}) = \mu$، مشابه الگوی ماتریسی (۳.۲)، $(\mathbf{I}_N - \mathbf{A})\,(\mathbf{Z} - \mu\mathbf{1}) = \varepsilon$ نوشته می‌شود.

۳ مدل اتورگرسیو فضایی نامانا

میدان‌های تصادفی فاقد شرایط مانایی از جمله عدم ثبات میانگین یا واریانس میدان‌های فضایی نامانا خواهد بود. در این بخش برای مدلسازی میدان‌های تصادفی با نامانایی در میانگین دو مدل اتورگرسیو فضایی نامانا پیشنهاد می‌دهیم.

۱.۳ مدل اتورگرسیو فضایی با روند چندجمله‌ای

فرض کنید $\boldsymbol{\mu} = (\mu_{۱,۱}, \ldots, \mu_{۱,n}, \ldots, \mu_{m,۱}, \ldots, \mu_{m,n})'$ میانگین بردار $\mathbf{Z}$ باشد، در صورتیکه $\mu_{i,j}$ را به‌صورت یک چند جمله ای پارامتری بر حسب شماره سطر و ستون‌های مشاهدات بنویسیم، با نماد ماتریسی خواهیم داشت $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ و به عنوان تعمیم دیگری از (۳.۲) مدل ماتریسی $(\mathbf{I}_N - \mathbf{A})(\mathbf{Z} - \mathbf{X}\boldsymbol{\beta}) = \boldsymbol{\varepsilon}$ حاصل می‌شود. اگر در این مدل ، ε_{ij}ها متغیرهای تصادفی مستقل و هم توزیع با توزیع $\mathrm{N}(۰, \sigma^۲)$ باشد، آنگاه $\boldsymbol{\varepsilon} \sim \mathrm{N}_{m\times n}(۰, \sigma^۲\mathbf{I}_N)$ می‌گردد. بردار تصادفی $\boldsymbol{Z} = \mathbf{B}^{-۱}\boldsymbol{\varepsilon} + \mathbf{X}\boldsymbol{\beta} \sim \mathrm{N}_{m\times n}(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Gamma})$ دارای ماتریس کواریانس $\boldsymbol{\Gamma} = \mathrm{Cov}(\mathbf{Z}) = \mathbf{B}^{-۱}\sigma^۲\mathbf{I}_N\mathbf{B}$ است، که $\mathbf{B} = (\mathbf{I_N} - \mathbf{A})$ ماتریس پایین مثلثی با عناصر یک روی قطر اصلی، $|\mathbf{B}| = ۱$ و $|\boldsymbol{\Gamma}|^{-۱} = |\sigma^۲\mathbf{I}_N|^{-۱}$ هستند.

لگاریتم تابع درستنمایی $\boldsymbol{\theta} = (\boldsymbol{\alpha}', \boldsymbol{\beta}', \sigma^۲)'$، $\boldsymbol{\alpha}' = (\alpha_۱, \alpha_۲, \alpha_۳, \alpha_۴, \alpha_۵)$ به‌صورت

$$\ell(\boldsymbol{\theta}|\mathbf{Z}) = Ln(L(\boldsymbol{\theta}|\mathbf{Z})) = -\frac{N}{۲}\ln\left(۲\pi\sigma^۲\right)\left\{(\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})'\mathbf{B}'\mathbf{B}\left(\mathbf{Z} - \boldsymbol{X}\boldsymbol{\beta}\right)\right\}. \qquad (۱.۳)$$

است. برای برآورد پارامترها به روش ماکسیمم درستنمایی مشتق تابع (۱.۳) نسبت به $\boldsymbol{\alpha}$، $\boldsymbol{\beta}$ و $\sigma^۲$ برابر صفر قرار می دهیم. فرض کنید $\mathbf{Y} = \mathbf{Z} - \mathbf{X}\boldsymbol{\beta}$ و $\mathbf{S_i} = \mathbf{W_i}\mathbf{Y}$ باشد، با صفر قرار دادن مشتق لگاریتم تابع درستنمایی نسبت به α_i و افراز $\mathbf{X}^* = [\mathbf{S}_۱|\mathbf{S}_۲|\mathbf{S}_۳|\mathbf{S}_۴|\mathbf{S}_۵]$، داریم:

$$\hat{\boldsymbol{\alpha}} = (\mathbf{X}^{*\prime}\mathbf{X}^*)^{-۱}\mathbf{X}^{*\prime}(\mathbf{Z} - \mathbf{X}\hat{\boldsymbol{\beta}}). \qquad (۲.۳)$$

مشتق رابطه (۱.۳)، نسبت به $\boldsymbol{\beta}$ و $\sigma^۲$ عبارتست از:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\hat{\mathbf{B}}'\hat{\mathbf{B}}\mathbf{X})^{-۱}\mathbf{X}'\hat{\mathbf{B}}'\hat{\mathbf{B}}\mathbf{Z}, \qquad (۳.۳)$$

$$\hat{\sigma^۲} = \frac{\left\{(\mathbf{Z} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'\hat{\mathbf{B}}'\hat{\mathbf{B}}(\mathbf{Z} - \boldsymbol{X}\hat{\boldsymbol{\beta}})\right\}}{N}. \qquad (۴.۳)$$

توجه کنید که بنا به روابط (۲.۳) و (۳.۳)، $\hat{\boldsymbol{\alpha}}$ و $\hat{\boldsymbol{\beta}}$ به یکدیگر وابسته‌اند که از طریق برازش یک مدل رگرسیونی و برآورد پارامترهای آن، برآورد اولیه $\hat{\boldsymbol{\beta}}_\bullet$ را به‌دست آورده و بردار $\mathbf{Y} = \mathbf{Z} - \mathbf{X}\hat{\boldsymbol{\beta}}_\bullet$ حساب می‌شود سپس الگوریتم ۱.۳ به‌کار گرفته می‌شود.

الگوریتم ۱.۳ . *الگوریتم تکراری برآورد پارامترهای مدل* $\mathrm{SAR}(۲,۱)$ *با روند چند جمله‌ای:*

گام ۱ ــ ابتدا $\hat{\boldsymbol{\beta}}_\bullet$ *توسط برازش یک مدل رگرسیونی و برآورد پارامترهای آن، برآورد شده و بردار* $\mathbf{Y} = \mathbf{Z} - \mathbf{X}\hat{\boldsymbol{\beta}}_\bullet$ *محاسبه شود.*

گام ۲ ــ $\hat{\boldsymbol{\alpha}}$ *با استفاده از رابطه (۲.۳) برآورد شود.*

گام ۳ــ با استفاده از $\hat{\boldsymbol{\alpha}}$ *در گام ۲ و رابطه (۳.۳)،* $\hat{\boldsymbol{\beta}}$ *محاسبه شود.*

گام ۴ ــ $\hat{\sigma^۲}$ *با استفاده از رابطه (۴.۳) برآورد شود.*

۲.۳ مدل اتورگرسیو فضایی تفاضلی شده

در این بخش همانند مدل‌های سری‌زمانی ARIMA برای حذف روند از عملگرهای تفاضلی استفاده می‌شود. فرض کنید ∇_1 و ∇_2 به ترتیب عملگرهای تفاضلی افقی و عمودی به‌صورت $\nabla_1 Z_{i,j} = (1-B_1)Z_{i,j} = Z_{i,j} - Z_{i-1,j}$ و $\nabla_2 Z_{i,j} = (1-B_2)Z_{i,j} = Z_{i,j} - Z_{i,j-1}$ باشد. به طریق مشابه $\nabla_1^{d_1} Z_{i,j} = \nabla_1(\nabla_1^{d_1-1} Z_{i,j});\ d_1 \geq 2$ و $\nabla_2^{d_2} Z_{i,j} = \nabla_2(\nabla_2^{d_2-1} Z_{i,j});\ d_2 \geq 2$ تعریف می‌شود. با استفاده از عملگرهای $\nabla_1^{d_1}$ و $\nabla_2^{d_2}$ می‌توان روندهای خطی و چندجمله‌ای‌های غیرخطی در میانگین یک میدان فضایی را از بین برد.

تعریف ۱.۳. *اگر $Y_{i,j} = \nabla_1^{d_1}\nabla_2^{d_2} Z_{i,j}$ دارای مدل $\mathrm{SAR}(p_1,p_2)$ تعریف شده در (۱.۲) باشد آن‌گاه میدان تصادفی $\{Z_{i,j}\}$ دارای مدل اتورگرسیو فضایی تفاضلی شده است که با $\{Z_{i,j}\} \sim \mathrm{SARI}(p_1,p_2,d_1,d_2)$ نمایش داده می‌شود. p_1 و p_2 مرتبه مدل اتورگرسیو، اعداد صحیح نامنفی d_1 و d_2 به ترتیب مرتبه تفاضلی کردن در جهت محور افقی و عمودی است.*

۴ پیش‌گویی فضایی

با فرض $\mathrm{E}(Z_{i,j}) = 0$، برای پیش‌گویی متغیر درون قلمرو داده‌ها باسو و راینسل (۱۹۹۳) با استفاده از مدل $\mathrm{SAR}(1,1)$ و بر اساس مشاهدات در نزدیکترین همسایگی، پیش‌گویی $Z_{k,s}$ را برای $k=2,\ldots,m-1$ و $s=2,\ldots,n-1$ به‌صورت

$$\begin{aligned}\hat{Z}_{k,s} &= \frac{1}{1+\alpha_1^2+\alpha_2^2+\alpha_3^2}\Big((\alpha_1-\alpha_2\alpha_3)\left(Z_{k-1,s}+Z_{k+1,s}\right) + (\alpha_2-\alpha_1\alpha_3)\left(Z_{k,s-1}+Z_{k,s+1}\right) \\ &\quad +\alpha_3\left(Z_{k-1,s-1}+Z_{k+1,s+1}\right) - \alpha_1\alpha_2\left(Z_{k-1,s+1}+Z_{k+1,s-1}\right)\Big), \qquad (1.4)\end{aligned}$$

معرفی کردند. مجیری و همکاران (۲۰۱۸a) با استفاده از این روش، پیشگویی مدل $\mathrm{SAR}(2,1)$ که ترکیب خطی از ۱۴ مشاهده مجاور است برای $k=3,\ldots,m-2$ و $s=2,\ldots,n-1$ به دست آوردند:

$$\begin{aligned}\hat{Z}_{k,s} &= \frac{1}{1+\alpha_1^2+\alpha_2^2+\alpha_3^2+\alpha_4^2+\alpha_5^2}\Big((\alpha_1-\alpha_1\alpha_4-\alpha_2\alpha_3-\alpha_3\alpha_5)\left(Z_{k-1,s}+Z_{k+1,s}\right) \\ &\quad +(\alpha_2-\alpha_1\alpha_3-\alpha_4\alpha_5)\left(Z_{k,s-1}+Z_{k,s+1}\right) - (\alpha_1\alpha_2+\alpha_3\alpha_4)\left(Z_{k-1,s+1}+Z_{k+1,s-1}\right) \\ &\quad +(\alpha_3-\alpha_1\alpha_5)\left(Z_{k-1,s-1}+Z_{k+1,s+1}\right) + (\alpha_4-\alpha_2\alpha_5)\left(Z_{k-2,s}+Z_{k+2,s}\right) \\ &\quad -\alpha_2\alpha_4\left(Z_{k-2,s+1}+Z_{k+2,s-1}\right) + \alpha_5\left(Z_{k-2,s-1}+Z_{k+2,s+1}\right)\Big). \qquad (2.4)\end{aligned}$$

۱.۴ پیش‌گویی فضایی برای مدل اتورگرسیو فضایی با روند چندجمله‌ای

وقتی $\mathrm{E}(Z_{i,j}) = \mu_{i,j}$ باشد، برای پیش‌گویی $Y_{i,j}$ با استفاده مدل‌های SAR کافی است، $Y_{i,j} = Z_{i,j} - \mu_{i,j}$ جایگزین $Z_{i,j}$ شود در این‌صورت $\mathrm{E}(Y_{i,j}) = 0$ و از روابط (۱.۴) و (۲.۴) می‌توان استفاده کرد. بنابراین $\hat{Y}_{i,j}$ در مدل $\mathrm{SAR}(2,1)$

برای $i = 3, \dots, m-2$ و $j = 2, \dots, n-1$ عبارتست از:

$$\begin{aligned}
\hat{Y}_{i,j} &= \mathrm{E}\big[Y_{i,j} \mid \mathrm{Y}^*\big] = \mathrm{E}\big[Y_{i,j} \mid Y_{k,\ell} = y_{k,\ell};\ k = 1, 2, \dots, m,\ \ell = 1, 2, \dots, n,\ (i,j) \neq (k,\ell)\big] \\
&= \frac{1}{1 + \hat{\alpha}_1^2 + \hat{\alpha}_2^2 + \hat{\alpha}_3^2 + \hat{\alpha}_4^2 + \hat{\alpha}_5^2}\Big((\hat{\alpha}_1 - \hat{\alpha}_1\hat{\alpha}_4 - \hat{\alpha}_2\hat{\alpha}_3 - \hat{\alpha}_3\hat{\alpha}_5)(y_{i-1,j} + y_{i+1,j}) \\
&\quad + (\hat{\alpha}_2 - \hat{\alpha}_1\hat{\alpha}_3 - \hat{\alpha}_4\hat{\alpha}_5)(y_{i,j-1} + y_{i,j+1}) + (\hat{\alpha}_3 - \hat{\alpha}_1\hat{\alpha}_5)(y_{i-1,j-1} + y_{i+1,j+1}) \\
&\quad - (\hat{\alpha}_1\hat{\alpha}_2 + \hat{\alpha}_3\hat{\alpha}_4)(y_{i-1,j+1} + y_{i+1,j-1}) + (\hat{\alpha}_4 - \hat{\alpha}_2\hat{\alpha}_5)(y_{i-2,j} + y_{i+2,j}) \\
&\quad - \hat{\alpha}_2\hat{\alpha}_4(y_{i-2,j+1} + y_{i+2,j-1}) + \hat{\alpha}_5(y_{i-2,j-1} + y_{i+2,j+1})\Big). \qquad (3.4)
\end{aligned}$$

۲.۴ پیش‌گویی فضایی برای مدل اتورگرسیو فضایی تفاضلی شده

در این روش پیش‌گویی به‌کمک $Y_{i,j} = \nabla_1^{d_1}\nabla_2^{d_2}Z_{i,j}$ روند داده‌ها به‌طور غیرمستقیم حذف می‌شود. بدین‌صورت که به‌وسیله تعریف ۱.۳ و با استفاده از مدل‌های SAR مانا، $\hat{Y}_{i,j}$ محاسبه شده سپس از روی آن $\hat{Z}_{i,j}$ به‌دست می‌آید. در مدل $\mathrm{SARI}(1,1,1,1)$ و $\mathrm{SARI}(2,1,1,1)$ با یکبار تفاضلی کردن در جهت محور افقی و عمودی روند داده‌ها حذف می‌شود. بنابراین با جای‌گذاری $\nabla_1\nabla_2 Z_{i,j} = Z_{i,j} - Z_{i-1,j} - Z_{i,j-1} + Z_{i-1,j-1}$ به جای $Z_{i,j}$ در رابطه (۲.۴) و ساده‌سازی پیش‌گویی مدل $\mathrm{SARI}(2,1,1,1)$ برای $i = 4, \dots, m-2$ و $j = 3, \dots, n-1$ حاصل می‌شود:

$$\begin{aligned}
\hat{Z}_{i,j} &= \frac{1}{\big(1 + \sum_{i=1}^{5} + A + B - C\big)}\Big(\big(1 + \sum_{i=1}^{5}\hat{\alpha}_i^2 + A + B + D\big)\big(z_{i-1,j} + z_{i,j-1}\big) - \big(A + D - E - \hat{\alpha}_2\hat{\alpha}_4\big) \\
&\quad \times \big(z_{i-2,j} + z_{i+1,j-1}\big) - \big(1 + \sum_{i=1}^{5}\hat{\alpha}_i^2 + A + B - C\big)z_{i-1,j-1} + (B - C)\big(z_{i-1,j-2} + z_{i,j+1}\big) \\
&\quad + \big(A - C - E + \hat{\alpha}_5\big)\big(z_{i+1,j} + z_{i-2,j-1}\big) - (B + D)\big(z_{i,j-2} + z_{i-1,j+1}\big) + (C - \hat{\alpha}_5) \qquad (4.4) \\
&\quad \times \big(z_{i+1,j+1} + z_{i-2,j-2}\big) + (D - \hat{\alpha}_2\hat{\alpha}_4)\big(z_{i-2,j+1} + z_{i+1,j-2}\big) + (E - \hat{\alpha}_5)\big(z_{i-3,j-1} + z_{i+2,j}\big) \\
&\quad - (E + \hat{\alpha}_2\hat{\alpha}_4)\big(z_{i-3,j} + z_{i+2,j-1}\big) + \hat{\alpha}_2\hat{\alpha}_4\big(z_{i-3,j+1} + z_{i+2,j-2}\big) + \hat{\alpha}_5\big(z_{i-3,j-2} + z_{i+2,j+1}\big)\Big),
\end{aligned}$$

که در آن $A = (\hat{\alpha}_1 - \hat{\alpha}_1\hat{\alpha}_4 - \hat{\alpha}_2\hat{\alpha}_3 - \hat{\alpha}_3\hat{\alpha}_5)$، $B = (\hat{\alpha}_2 - \hat{\alpha}_1\hat{\alpha}_3 - \hat{\alpha}_4\hat{\alpha}_5)$، $C = (\hat{\alpha}_3 - \hat{\alpha}_1\hat{\alpha}_5)$، $D = (\hat{\alpha}_1\hat{\alpha}_2 + \hat{\alpha}_3\hat{\alpha}_4)$ و $E = (\hat{\alpha}_4 - \hat{\alpha}_2\hat{\alpha}_5)$ است. توجه شود که اگر به عنوان مثال $\mathrm{E}(Z_{i,j}) = \beta_0 + \beta_1 i + \beta_2 j + \beta_3 ij + \beta_4 i^2 + \beta_5 j^2$ باشد $\mathrm{E}(\nabla_1\nabla_2 Z_{i,j}) = \hat{\beta}_3$ می‌شود لذا کافی است برای پیش‌گویی در مدل $\mathrm{SARI}(2,1,1,1)$ از $Y_{i,j} = \nabla_1\nabla_2 Z_{i,j} - \hat{\beta}_3$ استفاده شود.

۵ مثال کاربردی

در این مثال مدل‌های $\mathrm{SAR}(1,1)$ و $\mathrm{SAR}(2,1)$ با میانگین ثابت و سه مدل مختلف برای روند، $(1):\ \mu_{i,j} = \beta_0 + \beta_1 i + \beta_2 j$، $(2):\ \mu_{i,j} = \beta_0 + \beta_1 i + \beta_2 j + \beta_3 ij$ و $(3):\ \mu_{i,j} = \beta_0 + \beta_1 i + \beta_2 j + \beta_3 ij + \beta_4 i^2 + \beta_5 j^2$ بر روی مجموعه داده‌های پستی و بلندی آتشفشان مونگا در مجموعه داده‌ای با نام volcano با ابعاد 61×87 در نرم افزار R موجود است، برازش داده می‌شود.

جدول ۱: برآورد پارامترهای مدل‌های SAR(۱, ۱)، SAR(۲, ۱) و ملاک اطلاعاتی آکائیکه

$\mu_{i,j}$	مدل	$\hat{\boldsymbol{\alpha}}$	$\hat{\boldsymbol{\beta}}$	AIC
μ	SAR(۱, ۱)	(۰/۹۴, ۰/۹۳, −۰/۸۷)	۱۰۲/۸۴	۱۴۳۲۵/۹۱
	SAR(۲, ۱)	(۰/۸۵, ۰/۹۶, −۰/۸, −۰/۱۲, −۰/۱۳)	۱۳۱/۰۸	۲۳۵۸۴/۲۳
(۱)	SAR(۱, ۱)	(۰/۹۴, ۰/۹۳, −۰/۸۷)	(۱۰۱/۷, ۰/۰۵, ۰/۰۶)	۱۴۳۲۵/۳۲
	SAR(۲, ۱)	(۰/۹۶, ۰/۷۱, −۰/۴۳, −۰/۰۰۹, −۰/۲۴)	(۱۰۰/۵۳, ۰/۰۶, ۰/۱۲)	۱۳۶۵۲/۲۹
(۲)	SAR(۱, ۱)	(۰/۹۴, ۰/۹۲, −۰/۸۶)	(۱۰۱/۱۴, ۰/۰۹, ۰/۱۱, −۰/۰۱)	۱۴۳۰۴/۸۳
	SAR(۲, ۱)	(۰/۹۶, ۰/۷۲, −۰/۴۳, −۰/۰۰۹, −۰/۲۴)	(۱۰۰/۴۷, ۰/۰۷, ۰/۱۲, −۰/۰۰۳)	۱۳۶۵۲
(۳)	SAR(۱, ۱)	(۰/۹۲, ۰/۹۱, −۰/۸۳)	(۹۷/۳۴, ۰/۹۳, ۰/۷۶, −۰/۰۰۶, −۰/۰۱, −۰/۰۱)	۱۴۱۹۳/۳۲
	SAR(۲, ۱)	(۰/۹۶, ۰/۷۱, −۰/۴۱, −۰/۰۱, −۰/۲۴)	(۹۶/۸۴, ۰/۹۵, ۰/۴۵, −۰/۰۰۲, −۰/۰۱, −۰/۰۰۵)	۱۳۵۵۰/۱

پارامترهای مدل SAR(۲, ۱) به‌وسیله الگوریتم ۱.۳ و برای مدل SAR(۱, ۱) به طریق مشابه برآورد شده است. برای مقایسه مدل‌ها از ملاک اطلاعاتی آکائیکه (AIC) با رابطه $\mathrm{AIC} = ۲k - ۲\ln(\mathrm{L}(\hat{\boldsymbol{\theta}}))$ استفاده می‌شود، که $\mathrm{L}(\hat{\boldsymbol{\theta}})$ مقدار ماکسیمم تابع درستنمایی و k تعداد پارامترهای برآورد شده است. برآورد پارامترهای مدل‌های SAR(۱, ۱)، SAR(۲, ۱) و AIC در جدول ۵ آمده است. با توجه به این‌که AIC مدل SAR(۲, ۱) با میانگین (۳) کمتر از بقیه مدل‌هاست، لذا برای این داده‌ها مدل SAR(۲, ۱) با پارامترهای برآورد شده $\mu_{i,j} = ۹۶/۸۴ + ۰/۹۵i + ۰/۴۵j - ۰/۰۰۲ij - ۰/۰۱i^۲ - ۰/۰۰۵j^۲$ و $\hat{\boldsymbol{\alpha}} = (۰/۹۶, ۰/۷۱, -۰/۴۱, -۰/۰۱, -۰/۲۴)'$ بهترین مدل است.

برای پیش‌گویی فضایی مدل اتورگرسیو فضایی با روند چندجمله‌ای، ابتدا از داده‌ها $\hat{\mu}_{i,j}$ کسر می‌گردد و $\hat{Y}_{i,j}$ از رابطه (۳.۴) محاسبه شده سپس $\hat{Z}_{i,j} = \hat{Y}_{i,j} + \hat{\mu}_{i,j}$ به دست می‌آید. برای پیش‌گویی فضایی مدل اتورگرسیو فضایی تفاضلی شده، بر طبق نکته ذکر شده در بخش ۲.۴ بایستی به رابطه (۴.۴)، عبارت

$$\frac{-۲\hat{\beta}_۳}{۱ + \sum_{i=۱}^{۵} \hat{\alpha}_i^۲}\Big(A + B - C + D + E - \hat{\alpha}_۲\hat{\alpha}_۴ + \hat{\alpha}_۵\Big) + \hat{\beta}_۳$$

اضافه گردد. دقت پیش‌گویی مدل‌های اتورگرسیو فضایی با روند چندجمله‌ای و تفاضلی شده توسط میانگین قدر مطلق خطای پیش‌گویی (MAPE) بر طبق رابطه $\mathrm{MAPE} = \frac{۱}{N}\sum_i \sum_j |Z_{i,j} - \hat{Z}_{i,j}|$ ارزیابی گردید. نتایج بیان‌گر آن است که میانگین قدرمطلق خطای پیش‌گویی مدل‌های اتورگرسیو فضایی با روند چندجمله‌ای و تفاضلی شده به ترتیب برابر ۰/۴۳ و ۳۴/۷۵ است.

بحث و نتیجه‌گیری

در مدل‌های متداول اتورگرسیو فضایی، برآورد پارامترها فقط در شرایط مانایی قابل به دست آوردن هستند، این در حالی است که اغلب به دلایلی از جمله وجود روند فضایی در داده‌ها استفاده از مدل‌های SAR مانا منجر به خطای زیادی می‌شود. لذا در این مقاله دو نوع مدل SAR نامانا برای حل مشکل نامانایی در میانگین داده‌ها ارائه شد. بر اساس نتایج حاصل از داده‌های آتشفشان مونگا، میانگین قدرمطلق خطای پیش‌گویی برای مدل SAR با روند چندجمله‌ای کمتر از مدل SAR تفاضلی شده است. از آنجایی‌که یکی از اهداف پیش‌گویی درون داده‌ها ارزیابی دقت مدل‌ها است، مدل با روند چندجمله‌ای پیشنهاد می‌شود زیرا تعمیم آن به مدل‌های SAR مراتب بالا بسیار ساده‌تر است.

مراجع

محمدزاده، م.، (۱۳۹۸)، *آمار فضایی و کاربردهای آن*، چاپ سوم، مرکز نشر آثار علمی دانشگاه تربیت مدرس، تهران،

Anselin, L. (1988), *Spatial econometrics: methods and models*, Springer Science and Business Media.

Awang, N., Shitan, M. (2006), Estimating the parameters of the second order spatial unilateral autoregressive model, *International Journal of Statistical Sciences*, **5**, 37-58.

Basu, S., Reinsel, G. C. (1992), A note on properties of spatial Yule-Walker estimators, *Journal of statistical computation and simulation*, **41**, 243-255.

Basu, S., Reinsel, G. C. (1993), Properties of the spatial unilateral first-order ARMA model, *Advances in Applied Probability*, **25(3)**, 631-648.

Cressie, N. (1993), *Statistics for spatial data*, John Wiley, New York.

Mojiri, A., Waghei, Y., NiliSani, H. R., Mohtashami Borzadaran, G., R. (2018a), Spatial prediction by using unilateral autoregressive models in two dimensional space, *Journal of Statistical Sciences*, **12(1)**, 189-208.

Mojiri, A., Waghei, Y., NiliSani, H. R., Mohtashami Borzadaran, G., R. (2018b), Comparison of predictions by kriging and spatial autoregressive models, *Communications in Statistics-Simulation and Computation*, **47(6)**, 1785-1795.

Saber, M. M., Nematollahi, A. R. (2017). Comparison of spatial interpolation methods in the first order stationary multiplicative spatial autoregressive models, *Communications in Statistics-Theory and Methods*, **46(18)**, 9230-9246.

Ord, K. (1975), Estimation methods for models of spatial interaction, *Journal of the American Statistical Association*, **70**, 120-126.

Whittle, P. (1954), On sationary processes in the plane, *Biometrika*, **41**, 434-449.