



## مدل‌های کم‌رتبه پویا برای تحلیل بیزی داده‌های فضایی-زمانی حجیم

بهمن حمیدیان<sup>۱</sup>، حسین باغیشنی<sup>۱</sup>

<sup>۱</sup> گروه آمار، دانشگاه صنعتی شاهرود

چکیده: هدف این مقاله، تحلیل داده‌های فضایی-زمانی با استفاده از رده‌ای از مدل‌های بیزی پویا است. این رده به مدل‌های فرآیند پیش‌گو معروف شده است. ساختار این مدل‌ها قادر به شناسایی الگوی فضایی-زمانی موجود در داده‌های زمین‌آماري حجیم است. برای نمایش عملکرد این رده از مدل‌ها، داده‌های فضایی-زمانی کیفیت آب منطقه وسیعی از استان گلستان را در بازه زمانی سال‌های ۱۳۸۲ تا ۱۳۹۲ مورد تحلیل قرار داده‌ایم.

واژه‌های کلیدی: استنباط بیزی، داده‌های فضایی-زمانی، مدل‌های پویا، مدل‌های فرآیند پیش‌گو.

کد موضوع‌بندی ریاضی (۲۰۱۰): 62F15، 62H11.

### ۱ مقدمه

پیشرفت‌های اخیر در پیدایش دستگاه‌های قدرتمند ثبت و نگهداری داده‌ها، حضور داده‌های فضایی-زمانی را در تحقیقات علمی به شدت پررنگ کرده است. مدل‌بندی و تحلیل داده‌های فضایی-زمانی مورد توجه محققین مختلفی در زمینه‌های علمی گوناگون قرار گرفته است (کرسی و ویکل، ۲۰۱۱). ساختار پیچیده این نوع از داده‌ها، حجم بالای آن‌ها و نقص مدل‌های موجود، کاربران آن‌ها را به سمت توسعه مدل‌های جدید و پویا روان کرده‌اند. به عنوان چند نمونه می‌توان به **بنرجی و همکاران (۲۰۰۸)**، **فینلی و همکاران (۲۰۰۹)** و **فینلی و همکاران (۲۰۱۲)** اشاره کرد. مدل‌های سلسله‌مراتبی کاربرد وسیعی در تحلیل داده‌های زمین‌آماري دارند که معمولاً از چارچوب استنباط بیزی پیروی می‌کنند و تحلیل آن‌ها بر اساس روش‌های مبتنی بر نمونه‌گیری، مانند الگوریتم‌های مونت کارلوی زنجیر

<sup>۱</sup> بهمن حمیدیان: bhamidian@shahroodut.ac.ir

مارکوفی<sup>۱</sup> (MCMC) صورت می‌پذیرد (رابرت و کسلا، ۲۰۰۴). استفاده از مدل‌های سلسله‌مراتبی منجر به افزایش کارایی محاسباتی و انعطاف‌پذیری مدل‌های زمین‌آماري شده است. اما مشکل اصلی در تحلیل داده‌های فضایی-زمانی با مدل‌های سلسله‌مراتبی بیزی، سرعت کند الگوریتم‌های MCMC و تشخیص همگرایی زنجیر در حضور حجم بالای این نوع داده‌ها است. با توجه به این ویژگی داده‌های فضایی-زمانی، بنزجی و همکاران (۲۰۰۸) رده‌ای از مدل‌های سلسله‌مراتبی معروف به مدل‌های فرآیند پیش‌گو<sup>۲</sup> را معرفی کردند که بر اساس دستورالعملی، در مقابل از دست دادن سهمی از کارایی، مدل را با حجم کمتری از داده‌ها تقریب می‌زنند. این مدل‌ها به نوعی یک مدل کم‌تبه<sup>۳</sup> محسوب می‌شوند. در این مقاله به معرفی و استفاده از مدل‌های فرآیند پیش‌گو، برای تحلیل داده‌های زمین‌آماري گاوسی، می‌پردازیم که با کاهش فضای پارامتر موجب می‌شود تا نرخ همگرایی الگوریتم‌های MCMC و در نتیجه سرعت محاسبات بهبود یابد. کارایی مدل پیشنهادی را نیز با تحلیل داده‌های زمین‌آماري کیفیت آب بخش وسیعی از استان گلستان مورد ارزیابی قرار می‌دهیم.

## ۲ مدل خطی فضایی-زمانی پویا

فرض کنید  $y_t(s)$  متغیر پاسخ در مکان  $s$  و زمان  $t$  باشد، به طوری که  $t = 1, 2, \dots, N_t$  و  $s \in D \subseteq R^2$ . مدل مورد نظر ما در قالب یک مدل فضای حالت، ترکیبی از دو دسته معادله مشاهده و معادله‌های حالت است که برای وارد کردن مولفه‌های رگرسیونی و فضایی-زمانی به مدل استفاده و به صورت زیر نوشته می‌شوند:

$$\begin{aligned} y_t(s) &= x_t(s)^T \beta_t + u_t(s) + \varepsilon_t(s), & \varepsilon_t(s) &\stackrel{ind.}{\sim} N(0, \tau^2); \\ \beta_t &= \beta_{t-1} + \eta_t, & \eta_t &\stackrel{i.i.d.}{\sim} N(0, \Sigma_\eta); \\ u_t(s) &= u_{t-1}(s) + w_t(s), & w_t(s) &\stackrel{ind.}{\sim} GP(0, C_t(\cdot, \theta_t)), \quad t = 1, 2, \dots, N_t \end{aligned} \quad (1.2)$$

که اختصارات  $ind.$  و  $i.i.d.$  به ترتیب نشان‌دهنده مستقل و هم‌توزیع هستند. در این جا  $x_t(s)$  بردار  $p$ -بعدی متغیرهای تبیینی و  $\beta_t$  بردار  $p$ -بعدی ضرایب رگرسیونی وابسته به زمان متناظر با متغیرهای تبیینی می‌باشد. نماد  $GP(0, C_t(\cdot, \theta_t))$  نیز نمایانگر فرآیند گاوسی با تابع کوواریانس پارامتری  $C_t(\cdot; \theta_t)$  است. معمولاً شکل تابع کوواریانس به صورت  $C_t(s_1, s_2; \theta_t) = \sigma_t^2 \rho(s_1, s_2; \phi_t)$  است، که در آن  $\theta_t = \{\sigma_t^2, \phi_t\}$ ، تابع همبستگی با پارامتر  $\phi_t$ ، و  $\sigma_t^2$  مولفه واریانس فضایی می‌باشد. یک انتخاب برای تعریف تابع همبستگی فضایی، تابع نمایی است که به صورت  $C_t(s_1, s_2; \theta_t) = \sigma_t^2 \exp\{-\phi_t \|s_1 - s_2\|\}$  نمایش داده می‌شود و در آن  $\|s_1 - s_2\|$ ، فاصله اقلیدسی بین مکان‌های  $s_1$  و  $s_2$  است. با این وجود، هر تابع همبستگی فضایی معتبر را نیز می‌توان به کار برد (کرسی، ۱۹۹۳). با پذیرفتن  $\beta_0 \sim N(m_0, \Sigma_0)$  و  $u_0(s) = 0$ ، برای  $t = 0$ ، مدل سلسله‌مراتبی بیزی (۱.۲) کامل می‌شود. برای هر نقطه زمانی، تعداد مکان‌های یکسانی را در نظر می‌گیریم. بنابراین در ماتریس مکانی-زمانی داده‌ها، سطرها نمایانگر مکان و ستون‌ها نمایشگر زمان هستند؛ به عبارتی  $(i, j)$  امین عنصر،  $y_j(s_i)$  می‌باشد.

در مواردی که تعداد مکان‌ها به اندازه کافی بزرگ (یعنی سطرها زیاد برای ماتریس فضایی-زمانی داده‌ها) و تعداد نقطه‌های زمانی تحت کنترل یا محدود باشند، انجام استنباط بیزی کامل برای مدل (۱.۲)، از نظر محاسباتی، طاقت‌فرسا است. برای تشریح این سختی محاسباتی، فرض کنید از روش نمونه‌گیری گیبز برای تولید نمونه از توزیع پسین پارامترها (شامل اثرات تصادفی فضایی-زمانی) استفاده

<sup>۱</sup>Markov chain Monte Carlo

<sup>۲</sup>Predictive Process Models

<sup>۳</sup>Low-rank Model

کنیم. اجرای این الگوریتم نیازمند تجزیه ماتریس کوواریانس فضایی-زمانی است که برای هر نقطه زمانی، تعداد عملیات مورد نیاز از مرتبه سوم تعداد مکان‌ها است. یعنی برای مثلا  $n$  مکان، تعداد عملیات لازم برابر  $O(n^3)$  است. این تعداد عملیات برای تمام مراحل الگوریتم باید در تعداد تکرارهای نمونه‌گیری نیز ضرب شود.

## ۱.۲ فرآیند پیش‌گو

برای دوری از تنگنای محاسباتی مذکور، فرآیند فضایی-زمانی در (۱.۲) را با یک فرآیند کم‌رتبه معروف به فرآیند پیش‌گو جایگزین می‌کنیم (بنرجی و همکاران، ۲۰۰۸). برای این منظور، مجموعه کوچک‌تری از ناحیه مورد مطالعه  $D$  را که نقاط گره<sup>۴</sup> نام دارد، در نظر می‌گیریم و آن را با  $S^* = \{s_1^*, s_2^*, \dots, s_m^*\}$ ، به طوری که  $m \ll n$ ، نمایش می‌دهیم. تعریف می‌کنیم  $\tilde{w}_t(s) = E[w_t(s)|w_t^*]$  که در آن  $w_t^* = (w_t(s_1^*), w_t(s_2^*), \dots, w_t(s_m^*))^T$  صورت فرآیند پیش‌گو برای مدل (۱.۲) با جایگزینی  $u_t(s)$  در مدل (۱.۲) با  $\tilde{u}_t(s) = \sum_{k=1}^m [\tilde{w}_k(s) + \tilde{\varepsilon}_k(s)]$  حاصل می‌شود، که در آن  $\tilde{\varepsilon}_k(s)$  مؤلفه‌ای برای کاهش کم‌برآوردی تغییرات فضایی حاصل از فرآیند پیش‌گو است. اضافه کردن  $\tilde{\varepsilon}_k(s)$  به فرآیند، فرآیند پیش‌گوی اصلاح‌شده **فینلی و همکاران (۲۰۰۹)** را در پی خواهد داشت. با اجرای فرآیند بالا پیچیدگی محاسباتی از  $O(n^3)$  به  $O(m^3)$  (و در نتیجه زمان اجرا) کاهش می‌یابد. نکته کلیدی در این مدل‌ها، انتخاب گره است. روش‌های مختلفی برای انتخاب مکان گره‌ها پیشنهاد شده‌اند؛ برای مثال **رول و نیکا (۱۹۹۸)** و **فینلی و همکاران (۲۰۰۹)** را ببینید. یکی از روش‌ها، طراحی شبکه منظم است که در آن گره‌ها به صورت یک شبکه با فواصل یکسان پخش شده‌اند. در عمل اگر داده‌ها به طور همگن در ناحیه تحت مطالعه پراکنده باشند، تفاوت ناچیزی در روش‌های انتخاب گره وجود دارد. اما تعداد گره‌ها تاثیر زیادی در برآورد پارامترها و پیش‌گویی‌های بعدی می‌گذارد. بنابراین این تعداد باید، بسته به حساسیت استنباط و هزینه محاسباتی، تعیین شود.

## ۲.۲ برآزش مدل

فرض کنید مجموعه  $S = \{s_1, s_2, \dots, s_n\}$  شامل مکان‌هایی است که متغیرهای پاسخ و تبیینی مدل در آن‌ها مشاهده شده‌اند و برای هر مکان در  $N_t$  زمان مختلف، مشاهده داشته باشیم. همچنین فرض کنید  $(y_t(s_1), y_t(s_2), \dots, y_t(s_n))'$  بردار  $n$ -بعدی پاسخ در زمان  $t$  و  $X_t$  ماتریس  $n \times p$  است که  $i$ امین سطر آن بردار  $x_t(s_i)$  می‌باشد. مدل فرآیند پیش‌گو را می‌توان به صورت زیر نوشت:

$$y_t(s) = X_t(s)' \beta_t + \tilde{u}_t(s) + \varepsilon_t(s); \quad \varepsilon_t(s) \stackrel{ind.}{\sim} N(0, \tau_t^2 I_n), \quad t = 1, 2, \dots, N_t, \quad (2.2)$$

که در آن  $\tilde{u}_t = (\tilde{u}_t(s_1), \tilde{u}_t(s_2), \dots, \tilde{u}_t(s_n))'$  تحقیق‌های فرآیند پیش‌گو است. همچنین می‌توان نوشت:

$$\tilde{u}_t = \sum_{k=1}^m \left[ C_k(\theta_k)' C_k^*(\theta_k)^{-1} w_k^* + \tilde{\varepsilon}_k \right], \quad t = 1, 2, \dots, N_t,$$

که در آن  $C_k(\theta_k)'$  یک ماتریس  $n \times n^*$  است که  $(i, j)$ امین مولفه آن کوواریانس بین  $w_k(s_i)$  و  $w_k(s_j^*)$ ،  $C_k(s_i, s_j^*; \theta_k)$ ، است و  $C_k^*(\theta_k)$  ماتریسی  $n^* \times n^*$  با مولفه  $(i, j)$ ام  $C_k(s_i^*, s_j^*; \theta_k)$  است. همچنین  $\tilde{\varepsilon}_k \stackrel{ind.}{\sim} N(0, \tilde{D}_k)$  به طوری که  $\tilde{D}_k$  یک ماتریس قطری  $n \times n$  با  $i$ امین عنصر روی قطر  $C_k(s_i, s_i) - c_k(s_i)' C_k^*(\theta_k)^{-1} c_k(s_i)$  است، که در آن  $c_k(s_i)'$   $i$ امین سطر ماتریس  $C_k(\theta_k)'$  می‌باشد. برای انجام استنباط بیزی، باید توزیع پسین پارامترهای  $\gamma = \{\beta_0, \Sigma_\eta, \{\theta_t\}, \{\beta_t\}, \{\tilde{u}_t\}, \{w_t^*\}, \{\tau_t^2\}\}$

<sup>۴</sup>Knots

را، تحت توزیع پیشین  $\pi(\gamma)$ ، به دست آوریم. با فرض استقلال بین مولفه‌های  $\gamma$  و در نظر گرفتن توزیع‌های پیشین نرمال برای  $\beta$ ،  $\tilde{u}$  و  $w^*$ ، و یشارت معکوس برای  $\Sigma_\eta$ ، و گامای معکوس برای پارامترهای وابستگی  $\tau^2$  و  $\theta = (\phi, \sigma^2)$ ، توزیع پسین کامل مدل سلسله‌مراتبی متناسب است با

$$\prod_{t=1}^{N_t} IG(\phi_t | a_\phi, b_\phi) \prod_{t=1}^{N_t} IG(\sigma_t^2 | a_\sigma, b_\sigma) \prod_{t=1}^{N_t} IG(\tau_t^2 | a_\tau, b_\tau) N(\beta_o | m_o, \Sigma_o) IW(\Sigma_\eta | r_\eta, \gamma_\eta) \times \\ \prod_{t=1}^{N_t} N(\beta_t | \beta_{t-1}, \Sigma_\eta) \prod_{t=1}^{N_t} N(w_t^* | o, C_t^*(\theta_t)) \prod_{t=1}^{N_t} N(\tilde{u}_t | \tilde{u}_{t-1} + C_t(\theta_t)' C_t^*(\theta_t)^{-1} w_t^*, \bar{D}_t) \times \\ \prod_{t=1}^{N_t} N(y_t | X_t \beta_t + \tilde{u}_t, \tau_t^2 I_n).$$

تولید نمونه از این توزیع پسین با استفاده از الگوریتم نمونه‌گیری گیبز، نیازمند محاسبه توزیع‌های شرطی کامل است که جزئیات محاسبه آن‌ها در **فینلی و همکاران (۲۰۱۲)** آمده است. برای ارزیابی عملکرد مدل می‌توان از روش شبیه‌سازی مشاهدات برحسب مدل برازش داده‌شده و مقایسه با مقادیر واقعی استفاده کرد. برای این منظور، باید توزیع پسین پیش‌گو، یعنی

$$f(y_{rep,t}(s_i) | y) = \int N(y_{rep,t}(s_i); x_t(s_i)' \beta_t + \tilde{u}_t(s_i), \tau_t^2) \pi(\beta_t, \tilde{u}_t(s_i), \tau_t^2 | y),$$

را برای هر نقطه زمانی  $t$  محاسبه کنیم که در آن  $\pi(\beta_t, \tilde{u}_t(s_i), \tau_t^2 | y)$  توزیع پسین کناری پارامترهای مدل است. شبیه‌سازی از توزیع پسین پیش‌گو با وجود نمونه‌های MCMC تولیدشده کار سراسری است. کافی است به ازای نمونه‌های تولیدشده از توزیع پسین کناری  $\{\beta_t, \tilde{u}_t(s_i), \tau_t^2\}$ ، مشاهدات از توزیع  $N(x_t(s_i)' \beta_t + \tilde{u}_t(s_i), \tau_t^2)$  تولید شوند.

### ۳ عوامل موثر بر کیفیت آب‌های زیر زمینی استان گلستان

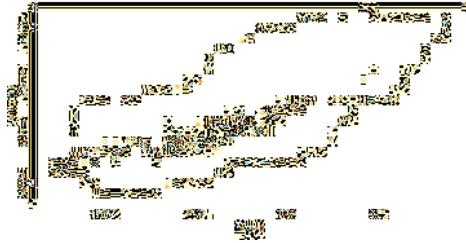
توانایی اندازه‌گیری آب برای عبور جریان الکتریکی، هدایت الکتریکی<sup>۵</sup> (EC) نامیده می‌شود. هدایت الکتریکی تابعی از غلظت آنیون‌ها، کاتیون‌ها و دمای محلول بوده، با مقدار کل جامدات محلول<sup>۶</sup> (TDS) همبستگی دارد و یکی از سریع‌ترین روش‌های ارزیابی کلی کیفیت آب‌های زیرزمینی می‌باشد. قابلیت هدایت الکتریکی یا عکس آن، یعنی مقاومت الکتریکی، یکی از عوامل مهم در تحقیقات هیدروژئولوژی می‌باشد و در اندازه‌گیری آن احتیاجی به دستگاه‌های پیچیده نیست. بنابراین مشکلات نمونه‌برداری وجود ندارد و اندازه‌گیری آن سریع و دقیق است. مقدار استاندارد تعیین‌شده برای EC توسط سازمان بهداشت جهانی ۵۰۰ میکروموس است. با توجه به این ویژگی‌های EC، بر آن شدید تا کیفیت آب‌های زیرزمینی بخشی از استان گلستان را، بر اساس آن، مورد مطالعه قرار دهیم. داده‌های موجود شامل اندازه‌های ثبت‌شده در ۱۵۰ ایستگاه از آب‌های زیرزمینی استان گلستان است که از آبان‌ماه سال ۸۲ تا آبان‌ماه سال ۹۲ گردآوری شده‌اند. منظور از ایستگاه‌های آب زیرزمینی، چشمه‌ها و چاه‌های عمیق و نیمه‌عمیق است و زمان‌های نمونه‌گیری نیز در دو ماه سال یعنی اردیبهشت و آبان انجام گرفته‌اند. بنابراین مجموعه داده‌ها شامل ۱۵۰ نقطه مکانی (سطر) و ۲۱ نقطه زمانی (ستون) و در مجموع ۳۱۵۰ مشاهده فضایی- زمانی برای هر متغیر است. متغیر پاسخ EC و متغیرهای تبیینی شامل سختی کل<sup>۷</sup> (th)، پتاسیم (k)، سدیم (na)، سولفات (so4)، کلر (cl)، بیکربنات (hco3)، ph و کل جامدات محلول هستند. این داده‌ها را با هر دو مدل پرتبه (۱.۲) و فرآیند پیش‌گو (۲.۲) با

<sup>۵</sup>Electric Conductivity

<sup>۶</sup>Total Dissolved Solution

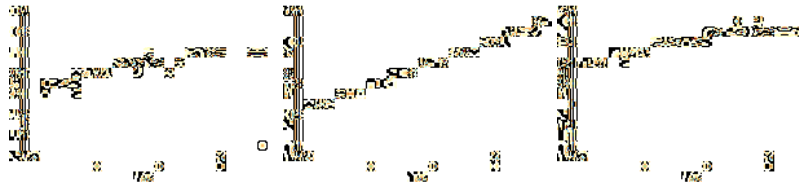
<sup>۷</sup>Total Hardness

انتخاب ۵۰ گره، مورد تحلیل قرار دادیم. برای این‌که عملکرد مدل پرتبه را ارزیابی کنیم در ۶ ایستگاه (علامت‌های ستاره در شکل ۱)، ۱۰ نقطه زمانی را کنار گذاشته و مدل را با بقیه داده‌ها برازش دادیم. به طور مشابه برای مدل فرآیند پیش‌گو، ۳ ایستگاه (شکل ۶، سمت چپ) حذف شدند. مکان‌های جغرافیایی ایستگاه‌های نمونه‌گیری در شکل ۱ نمایش داده شده‌اند. بر اساس تحلیل اکتشافی داده‌ها (شکل



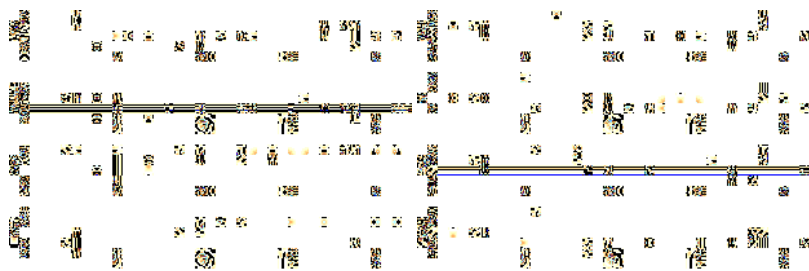
شکل ۱: پراکنش جغرافیایی ایستگاه‌های آب زیرزمینی

۲ برای سه زمان ۱، ۲ و ۲۰ (به عنوان نمونه)، تابع تغییرنگار نمایی به خوبی به تغییرنگار تجربی برازش دارد. در نتیجه تابع کوواریانس نمایی را برای ساختار وابستگی فضایی انتخاب کردیم. برای ضرایب رگرسیونی، توزیع پیشین نرمال  $\beta \sim N(0, I_8)$  را انتخاب کردیم

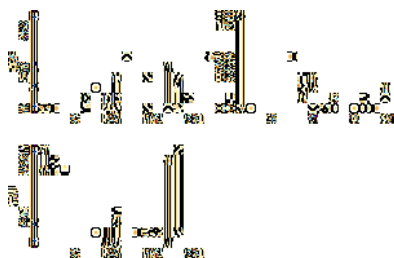


شکل ۲: تغییرنگار برازش‌شده به داده‌ها در سه زمان ۱، ۲ و ۲۰

که در آن  $I_8$  ماتریس همانی با بعد ۸ است. همچنین برای ضرایب وابستگی فضایی توزیع گامای معکوس  $\phi_t, \sigma_t^2, \tau_t^2 \sim IG(2, 25)$  انتخاب شد. علاوه بر این فرض کردیم  $(\Sigma_\eta \sim IW(2, \text{diag}(0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1)))$  که در آن  $\text{diag}(0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1)$  یک ماتریس قطری با بعد ۸ با درآیه‌های روی قطر ۰,۰۰۱ است. برآوردهای بیزی ضرایب رگرسیونی وابسته به زمان به همراه فواصل اعتبار ۹۵ درصد آن‌ها در شکل ۳ در طول ۲۱ زمان نمایش داده شده‌اند. این شکل می‌تواند در توضیح تغییرپذیری زمانی ضرایب مفید باشد. از روی شکل می‌توان تاثیر زمانی هر کدام از متغیرهای تبیینی را بر روی EC تعبیر کنیم. به عنوان یک نمونه، همان‌طور که دیده می‌شود متغیر سختی کل، پتاسیم و سدیم در اکثر زمان‌ها تاثیر مثبتی بر EC دارند. همچنین هر سه متغیر رفتار مشابهی در زمان‌های مختلف دارند. پس هر چه متغیرهای سختی کل، پتاسیم و سدیم مقادیر بزرگتری اختیار کنند، EC نیز افزایش می‌یابد و در نتیجه کیفیت آب کاهش می‌یابد. بنابراین برای جلوگیری از کاستن کیفیت آب باید جلوی افزایش این مقادیر را در آب‌های زیرزمینی گرفت. شکل ۴ مشابه شکل ۳ است ولی برای پارامترهای وابستگی مدل و به طور مشابه قابل تعبیر. در مجموع، با توجه به وجود ۱۵۰ ایستگاه در هر زمان مدل تقریباً توانسته است پارامترهای فرآیند و تغییرپذیری آن‌ها را به خوبی برآورد کند. شکل ۵ مقادیر مشاهده‌شده و برآوردشده ۶ ایستگاه حذف‌شده برای اعتبارسنجی مدل را نمایش می‌دهد. در این شکل دایره‌ها نشان‌دهنده مشاهدات مورد استفاده در برازش مدل و ستاره‌ها نشان‌دهنده مشاهدات حذف‌شده هستند. همان‌طور که مشهود است برآورد فواصل اعتبار در این ایستگاه‌ها کوتاه است که نشان از عملکرد پیش‌گویی خوب مدل دارد. همچنین داده‌های حذف‌شده همگی در داخل فاصله اعتبار قرار گرفته‌اند که حاکی از سازگاری مدل برازش‌شده در تشریح

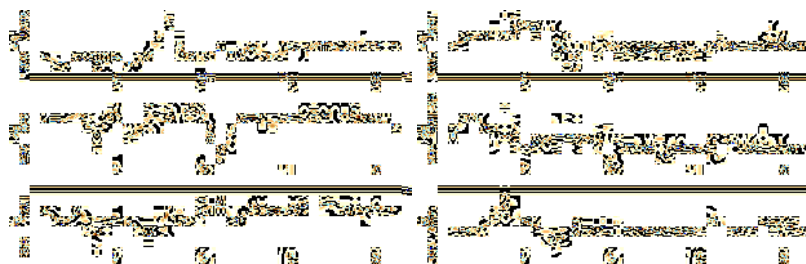


شکل ۳: برآورد (میانه) و فواصل اعتبار بیزی ۹۵٪ برای اثرات رگرسیونی وابسته به زمان در مدل پرتبه (۱.۲)



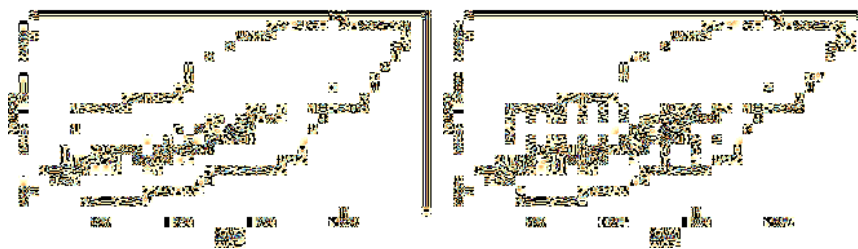
شکل ۴: میانه و فواصل اعتبار بیزی ۹۵٪ برای پارامترهای ساختار وابستگی مدل پرتبه (۱.۲)

روند فضایی-زمانی مدل است.

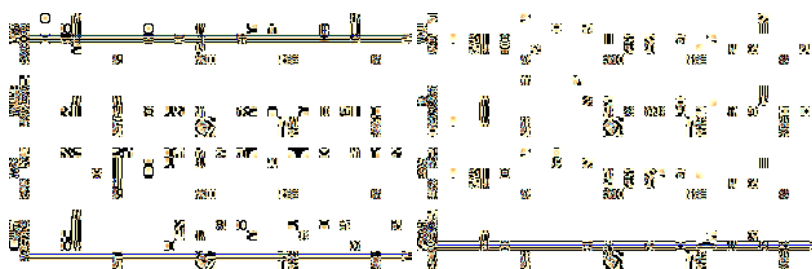


شکل ۵: میانه توزیع پیش‌گو و فواصل اعتبار بیزی ۹۵٪ که به ترتیب با خطوط غیرمنقطع و منقطع برای ۶ ایستگاه حذف‌شده در مدل پرتبه (۱.۲) نشان داده شده‌اند.

برای ارزیابی عملکرد مدل فرآیند پیش‌گو (۲.۲)، تعداد ۵۰ گره و مطابق با روش شبکه منظم انتخاب شدند که در شکل ۶ نشان داده شده‌اند. لازم به ذکر است که در این مثال، انتخاب تعداد گره‌های کمتر از ۵۰، دقت تحلیل را کاهش و بیشتر از آن دقتی مشابه با ۵۰ گره و در عین حال هزینه بیشتر را در پی داشته است. همانند مدل پرتبه، برآوردهای بیزی ضرایب رگرسیونی وابسته به زمان و پارامترهای وابستگی فضایی، به همراه فواصل اعتبار بیزی ۹۵ درصد آن‌ها، در شکل‌های ۷ و ۸ برای ۲۱ زمان نمایش داده شده‌اند. این شکل‌ها مشابه مدل پرتبه قابل تعبیر و نتایجی مشابه قابل استخراج هستند. شکل ۹ مقادیر مشاهده‌شده و پیش‌گویی برای ۳ ایستگاهی که برای اعتبارسنجی مدل حذف شده بودند را نمایش می‌دهد. همانند مدل پرتبه، فواصل اعتبار کوتاه نشان از قدرت پیش‌گویی خوب



شکل ۶: مکان گره‌ها و ایستگاه‌های کنار گذاشته شده روی نقشه استان گلستان که به ترتیب با علامت ستاره در سمت راست و چپ شکل مشخص شده‌اند.



شکل ۷: میانه و فواصل اعتبار بیزی ۹۵٪ برای اثرات متغیرهای تبیینی مدل فرآیند پیش‌گو (۲۰۲)

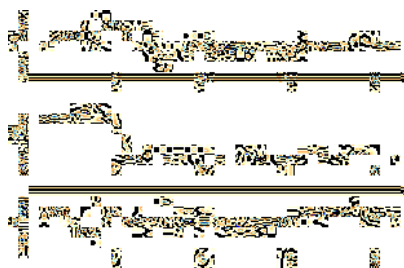


شکل ۸: میانه و فواصل اعتبار بیزی ۹۵٪ برای پارامترهای وابستگی فضایی در مدل فرآیند پیش‌گو (۲۰۲)

مدل دارد. تمام داده‌های حذف شده نیز در داخل این فاصله‌ها قرار دارند.

## بحث و نتیجه‌گیری

با مدل‌بندی داده‌های واقعی کیفیت آب‌های زیرزمینی استان گلستان، توانستیم عملکرد مدل کم‌رتبه فرآیند پیش‌گو که یک تقریب بسیار خوب از مدل کلی می‌باشد را به نمایش درآوریم. در واقع نتایج به‌دست آمده از مدل فرآیند پیش‌گو تفاوت چشمگیری با مدل پرتبه ندارد. این تقریب قابل قبول با کاهش هزینه محاسباتی نیز همراه است؛ در واقع هزینه محاسباتی از ۱۵۷۷/۹۰ ثانیه برای مدل پرتبه به ۳۳۹/۰۲ ثانیه برای مدل کم‌رتبه تقلیل یافته است و این زمان قابل توجهی را برای کاربر ذخیره می‌کند.



شکل ۹: میانه توزیع پیش‌گو و فواصل اعتبار ۹۵٪ که به ترتیب با خطوط غیرمنقطع و منقطع برای سه ایستگاه حذف‌شده مدل فرآیند پیش‌گو (۲.۲) نشان داده شده‌اند. دایره‌های خالی نشان‌گر مشاهدات مورد استفاده برای اعتبارسنجی مدل هستند. ستاره‌ها نشان‌گر مشاهدات مورد استفاده برای اعتبارسنجی مدل هستند.

## سپاسگزاری

از سازمان آب منطقه‌ای استان گلستان بابت در اختیار گذاشتن داده‌های آب زیرزمینی استان گلستان قدردانی می‌کنیم.

## مراجع

- Banerjee, S., Gelfand, A.E., Finely, A.O. and Sang, H. (2008), Gaussian predictive process models for large spatial data sets, *Journal of the Royal Statistical Society, Series B*, **70(4)**, 825-848.
- Cressie, N. (1993), *Statistics for Spatial Data*, John Wiley, New York.
- Cressie, N. and Wikle, C. (2011), *Statistics for Spatio-Temporal Data*, John Wiley, London.
- Finely, A.O., Banerjee, S. and Gelfand, A.E. (2012), Bayesian dynamic modeling for large space-time datasets using Gaussian predictive processes, *Journal of Geographical Systems*, **14(1)**, 29-47.
- Finely, A.O., Sang, H., Banerjee, S. and Gelfand, A.E. (2009), Improving the performance of predictive process modeling for large datasets, *Computational Statistics and Data Analysis*, **53(8)**, 2873-2884.
- Robert, C. and Casella, G. (2004), *Monte Carlo Statistical Methods*, Second edition, Springer, New York.
- Royle, J. and Nychka, D. (1998), An algorithm for the construction of spatial coverage designs with implementation in S-PLUS, *Computers and Geosciences*, **24(5)**, 479-488.