

## تحلیل داده‌های فضایی-زمانی شمارشی با مدل منعطف گامای شمارشی

مهسا نادى فر<sup>۱</sup>، حسین باغیشنى<sup>۱</sup>، افشین فلاح<sup>۲</sup>

<sup>۱</sup>گروه آمار، دانشکده علوم ریاضی، دانشگاه صنعتی شاهرود

<sup>۲</sup>گروه آمار، دانشگاه بین‌المللی امام خمینی

تاریخ دریافت: ۱۳۹۹/۰۱/۲۱ تاریخ پذیرش و انتشار: ۱۳۹۹/۰۸/۰۳

**چکیده:** بسیاری از داده‌های فضایی-زمانی، به‌ویژه در پزشکی و تصویرسازی بیماری‌ها، شمارشی هستند. معمولاً این نوع داده‌های شمارشی دارای تغییرپذیری مضاعفی هستند که کارایی مدل پواسون را برای تحلیل آن‌ها خدشه‌دار می‌کند. لحاظ کردن این تغییرپذیری در فرآیند مدل‌بندی، نقش مهمی در افزایش کارایی تحلیل داده‌های فضایی-زمانی شمارشی دارد. با این هدف، در این مقاله، یک مدل فضایی-زمانی بیزی جدید با انعطاف بالا برای مدل‌بندی پراکنش داده‌ها، به نام گامای شمارشی، معرفی شده است. برای برازش و استنباط این مدل، روش تقریب لاپلاس آشیانی جمع‌بسته به کار گرفته شده است. برای ارزیابی عملکرد مدل پیشنهادی در مقایسه با مدل‌های سنتی پواسون و دوجمله‌ای منفی، از یک مطالعه شبیه‌سازی استفاده شده است. افزون بر این، کاربست مدل در تحلیل داده‌های سرطان خون در استان خراسان رضوی، نشان داده شده است.

**واژه‌های کلیدی:** اثر متقابل فضایی-زمانی، بیش‌پراکنش، توزیع پیشین با پیچیدگی توانانیده، داده‌های شمارشی، مدل گامای شمارشی.

## ۱ مقدمه

در شاخه‌های مختلفی از علوم مانند همه‌گیرشناسی، پزشکی، محیط زیست، و اقتصاد، پدیده‌هایی وجود دارند که مشاهدات آن‌ها در موقعیت‌های مختلف جغرافیایی و زمانی ثبت می‌شوند. به این نوع داده‌ها، معمولاً یک ساختار وابستگی القا می‌شود که ناشی از هر دو منبع ناحیه جغرافیایی تحت مطالعه و بازه زمانی گردآوری آن‌ها است. در متون آماری به این نوع داده‌ها، فضایی-زمانی و به ساختار وابستگی القاشده، ساختار وابستگی فضایی-زمانی گفته می‌شود. داده‌های فضایی-زمانی، مقیاس‌های جغرافیایی وسیعی را در طول زمان نشان می‌دهند و معمولاً تحت تاثیر اطلاعات فردی قرار نمی‌گیرند. این نوع داده‌ها، به‌ویژه، در تدوین سیاست‌های بهداشتی، اقتصادی، و جمعیتی کشورها، نقش بسزایی دارند. تحلیل داده‌های فضایی-زمانی بر شناسایی تغییرات محلی و جهانی الگوها و روندهای فضایی در طول زمان، و تعیین کمیت‌های ارتباطی متمرکز است (اوتازی و همکاران، ۲۰۱۸). در سال‌های اخیر مدل‌بندی داده‌های مربوط به بهداشت و سلامت از جمله بیماری‌های سرطان، بر حسب مکان و زمان، بسیار مورد توجه قرار گرفته است. این نوع داده‌ها اغلب شمارشی هستند و استفاده از توزیع‌های شمارشی برای تحلیل آن‌ها ضروری است.

رگرسیون پواسون (مک‌کالاک و نلدر، ۱۹۸۹) رایج‌ترین مدل برای تحلیل پاسخ‌های شمارشی است. یکی از پذیره‌های اصلی این مدل، برابری میانگین و واریانس متغیر پاسخ است. اما این پذیره در بسیاری از کاربردها برقرار نیست. حالتی را که واریانس مشاهدات از میانگین آن‌ها بزرگتر است، بیش‌پراکندگی و وارون آن را کم‌پراکندگی می‌نامند. برای برخورد با بیش یا کم‌پراکندگی پاسخ‌های شمارشی، تاکنون پژوهش‌های گوناگونی انجام و راه‌حل‌هایی پیشنهاد شده‌اند. از جمله مدل‌های جایگزین می‌توان به مدل‌های پواسون تعمیم‌یافته و پواسن آمیخته اشاره کرد (کمرون و تریودی، ۲۰۱۳؛ فلاح و همکاران، ۱۳۹۲). با این وجود، اغلب مطالعات موجود بر چگونگی برخورد با پدیده بیش‌پراکندگی تمرکز دارند و برای مساله کم‌پراکندگی پژوهش‌های اندکی انجام شده‌اند. به‌عنوان نمونه، می‌توان به مدل رگرسیون کانوی-مکسول-پواسون<sup>۱</sup> (سلرز و شیمولی، ۲۰۱۰) اشاره کرد که در آن، به دلیل وجود یک سری نامتناهی در عبارت میانگین پاسخ، مدل روی تقریبی از میانگین این متغیر تعریف می‌شود. وینکلن (۱۹۹۵) نشان داد که با استفاده از نظریه تجدید<sup>۲</sup> و با در نظر گرفتن توزیع‌های احتمالی مختلف برای زمان‌های بین رخداد پیشامدها، می‌توان به توزیع‌های شمارشی جدیدی برای تعداد رخدادها دست یافت که انعطاف‌پذیری بیشتری

<sup>۱</sup>Conway-Maxwell-Poisson<sup>۲</sup>Renewal Theory

برای تحلیل داده‌ها دارند. برای این منظور، می‌توان از توزیع‌های پیوسته با تکیه‌گاه مثبت مانند نمایی، گاما، وایبل و لگ‌نرمال استفاده کرد. اگر توزیع زمان‌های بین رخداد پیشامدها گاما در نظر گرفته شود، توزیعی به نام گامای شمارشی<sup>۱</sup> (GC) به دست می‌آید. در این توزیع برای مدل‌بندی پراکنش داده‌ها، پارامتری وجود دارد که با در نظر گرفتن مقادیر مختلف برای آن می‌توان بیش پراکنش، کم پراکنش یا پراکنش متعادل را در مدل لحاظ کرد. اولین بار **وینکلمن** (۱۹۹۵) استفاده از این توزیع را در تحلیل داده‌های شمارشی پیشنهاد کرد. این پیشنهاد تا مدت‌ها بی‌توجه ماند تا این‌که **زیوانی و همکاران** (۲۰۱۴) یک مدل رگرسیونی مبتنی بر توزیع گامای شمارشی را برای تحلیل داده‌های دارای کم پراکنش با یک رویکرد مبتنی بر درست‌نمایی معرفی کردند. اخیراً بسته نرم‌افزاری countR در محیط نرم‌افزار R برای تحلیل مبتنی بر درست‌نمایی داده‌های شمارشی معرفی شده است که توزیع گامای شمارشی را نیز شامل می‌شود (**خراط و همکاران**، ۲۰۱۹).

با پیشرفت روش‌های استنباط بیزی، در سال‌های اخیر، مدل‌های رگرسیونی بیزی شامل مولفه‌های فضایی و فضایی-زمانی در تحلیل داده‌های پزشکی به‌طور گسترده مورد بررسی قرار گرفته‌اند، که از آن میان می‌توان به **گرینلند** (۲۰۰۶)، **بری و همکاران** (۲۰۱۱) و **بایو** (۲۰۱۲) اشاره کرد. استفاده از رویکرد بیزی این امکان را فراهم می‌آورد تا تحلیل‌گر بتواند باورهای پیشین خود را درباره جامعه یا پارامترهای آن با در نظر گرفتن توزیع‌های پیشین مناسب، به فرآیند مدل‌بندی وارد کند. افزون بر این، رهیافت بیزی بستری مناسب برای در نظر گرفتن ساختارهای سلسله‌مراتبی فراهم می‌کند که به کمک آن‌ها می‌توان یک مدل پیچیده را در قالب مدل‌های ساده‌تر مطالعه کرد. در کنار این مزیت‌ها، چالش اصلی در رویکرد بیزی، بار محاسباتی سنگین آن است. به این معنی که در بسیاری از مدل‌های بیزی، توزیع پسین و کمیت‌های پسینی مورد علاقه دارای نمایش بسته نیستند. روش‌های پرترفدار مبتنی بر نمونه‌گیری مانند الگوریتم‌های مونت کارلو زنجیر مارکوفی<sup>۲</sup> (MCMC)، با تولید نمونه از توزیع پسین، تقریبی برای کمیت‌های پسینی مورد علاقه ارائه می‌کنند. سادگی استفاده از نرم‌افزارهای بیزی مانند BayesX (**لانگ و همکاران**، ۲۰۰۵)، OpenBugs (**لون و همکاران**، ۲۰۰۹)، Stan (**گلن و همکاران**، ۲۰۱۵) و بسته‌های نرم‌افزاری موجود در نرم‌افزار R، یکی از دلایل استقبال از این الگوریتم‌ها است. این روش‌ها برخلاف سودمندی‌ها و نقشی که در ساخت استنباط‌های بیزی دارند، برای تقریب توزیع پسین در مدل‌های پیچیده، مانند مدل‌های فضایی-زمانی، با مشکلات محاسباتی جدی مواجه می‌شوند. رسیدن به نرخ همگرایی مطلوب برای الگوریتم‌های MCMC در مدل‌های پیچیده فضایی-زمانی، اغلب زمان‌بر است و این الگوریتم‌ها از کارایی محاسباتی قابل قبولی برخوردار نیستند. برای پرهیز از مشکلات محاسباتی اشاره‌شده، **رو و همکاران** (۲۰۰۹) یک روش بیزی

<sup>1</sup>Gamma-Count

<sup>2</sup>Markov Chain Monte Carlo

تقریبی با نام تقریب لاپلاس آشیانه‌ای جمع‌بسته<sup>۱</sup> (INLA) معرفی کردند که از نظر محاسباتی بسیار سریع‌تر از الگوریتم‌های MCMC است و دقت نتایج حاصل از آن نیز با نتایج روش‌های MCMC رقابت می‌کند. روش INLA برای مدل‌های گاوسی پنهان که یک رده بزرگ از مدل‌های پرکاربرد شامل مدل‌های فضایی و فضایی-زمانی است، طراحی شده است. به همین دلیل، می‌توان این روش را در کاربردهای متنوعی مورد استفاده قرار داد (ریبلر و همکاران، ۲۰۱۲؛ لی و همکاران، ۲۰۱۲؛ بلانگیاردو و همکاران، ۲۰۱۳).

در این مقاله، یک مدل فضایی-زمانی شمارشی جدید بر پایه توزیع GC برای تحلیل داده‌های شمارشی با همبستگی فضایی-زمانی توسعه داده شده است. مدل پیشنهادی، تعمیمی از کار نادفر و همکاران (۲۰۱۹) است که تاکنون تنها پژوهش ارائه‌شده برای تحلیل داده‌های شمارشی فضایی بر اساس توزیع GC است. چگونگی تحلیل مدل پیشنهادی با استفاده از رویکرد بیزی و روش INLA مورد بحث قرار گرفته است. در بخش ۲، توزیع گامای شمارشی و مدل رگرسیونی مبتنی بر آن معرفی شده است. ویژگی‌های مدل و چگونگی برازش آن در بخش ۳ بررسی شده است. در بخش ۴، عملکرد مدل پیشنهادی با یک مطالعه شبیه‌سازی مورد ارزیابی قرار گرفته و کارایی آن با مدل‌های رقیب پواسون و دوجمله‌ای منفی (NB)، مقایسه شده است. در بخش ۵، داده‌های مربوط به سرطان خون در استان خراسان رضوی، با کمک مدل فضایی-زمانی GC تحلیل شده است. در پایان، بحث و نتیجه‌گیری ارائه شده است.

## ۲ مدل رگرسیونی گامای شمارشی

دنباله‌ای از پیشامدها را در نظر بگیرید که زمان‌های رخداد آن‌ها مشاهده نشده و تنها تعداد پیشامدها قبل از یک نقطه زمانی مشخص، مشاهده شده باشند. نظریه تجدید (کاکس، ۱۹۶۲)، با کمک استدلال‌های احتمالی، نشان می‌دهد که با آگاهی از توزیع زمان‌های رخداد پیشامدها می‌توان توزیع تعداد رخدادها را شناسایی کرد. فرض کنید  $\{T_k, k \in \mathbb{N}\}$  دنباله زمان‌های مورد انتظار بین پیشامدهای  $(k-1)$ ام و  $k$ ام باشد. بنابراین زمان رخداد  $\ell$ امین پیشامد، برای  $\ell = 1, 2, \dots$  به صورت  $\nu_\ell = \sum_{k=1}^{\ell} T_k$  به دست می‌آید. اکنون فرض کنید فرآیند شمارشی  $\{Y_t, t \geq 0\}$  تعداد کل رخدادها را در بازه زمانی  $[0, t]$  نشان می‌دهد. با کمک تعاریف  $Y_t$  و  $\nu_\ell$  به راحتی می‌توان گفت که  $Y_t < \ell$  اگر و فقط اگر  $\nu_\ell \geq t$ . بنابراین  $P(Y_t < \ell) = P(\nu_\ell \geq t) = 1 - F_\ell(t)$  که در آن  $F_\ell(t)$  تابع توزیع تجمعی  $\nu_\ell$  است. بنابراین،

<sup>1</sup>Integrated Nested Laplace Approximation

تابع جرم احتمال  $Y_t$  به صورت

$$p_{Y_t}(y) = P(Y_t = y) = P(Y_t < y + 1) - P(Y_t < y) = F_y(t) - F_{y+1}(t) \quad (۱)$$

حاصل می‌شود (کاکس، ۱۹۶۲؛ کمرون و تریودی، ۲۰۱۳). تساوی (۱) رابطه پایه‌ای بین توزیع زمان‌های رخداد پیشامدها و توزیع تعداد رخدادها را بیان می‌کند. اگر دنباله  $\{\tau_k\}$  مستقل با تابع چگالی احتمال مشترک  $f(\tau)$  باشند، آن را فرآیند تجدید می‌نامند.

توزیع گامای شمارشی یک توزیع گسسته است که با پذیرفتن توزیع گاما برای  $\tau_k$ ها به دست می‌آید. به طور مشخص اگر فرض کنیم  $\tau_k$ ها متغیرهایی مستقل و هم‌توزیع با توزیع گاما،  $\text{Gamma}(\alpha, \gamma)$ ، با میانگین  $E(\tau) = \alpha/\gamma$  و واریانس  $\text{Var}(\tau) = \alpha/\gamma^2$  هستند، آن‌گاه  $Y_t$  دارای توزیع گامای شمارشی با پارامترهای  $\alpha$  و  $\gamma$  است که با نماد  $Y_t \sim \text{GC}(\alpha, \gamma)$  نشان می‌دهیم (وینکلن، ۱۹۹۵). تابع جرم احتمال این توزیع به صورت

$$p_{Y_t}(y) = G(y\alpha, \gamma t) - G((y+1)\alpha, \gamma t), \quad y = 0, 1, 2, \dots$$

است، که در آن  $G(y\alpha, \gamma t) = \frac{1}{\Gamma(y\alpha)} \int_0^{\gamma t} x^{y\alpha-1} e^{-x} dx$ . میانگین این توزیع نیز به صورت

$$E(Y_t) = \sum_{i=1}^{\infty} G(i\alpha, \gamma t) \quad (۲)$$

محاسبه می‌شود. پارامتر  $\alpha$  در توزیع GC وضعیت پراکنش داده‌ها را کنترل می‌کند. به ازای  $\alpha = 1$ ، توزیع زمان‌های بین رخدادها به نمایی و توزیع  $Y_t$  به پواسون تبدیل می‌شود. برای مقادیر  $\alpha < 1$  توزیع GC یک توزیع بیش‌پراکنده و در حالت  $\alpha > 1$ ، توزیعی کم‌پراکنده است.

برای تعریف یک مدل رگرسیونی مبتنی بر توزیع GC، بر اساس نظریه مدل‌های خطی تعمیم‌یافته، باید میانگین متغیر پاسخ و متغیرهای تبیینی به کمک یک تابع پیوند مناسب، مرتبط شوند. بدون کاستن از کلیت مسئله، فرض کنید طول بازه زمانی مورد انتظار برای رخداد پیشامدها، برای همه مشاهدات یکسان و برابر واحد است. در مدل رگرسیون GC، به دلیل پیچیدگی رابطه (۲) نمی‌توان مدل رگرسیونی را به طور مستقیم برحسب میانگین پاسخ  $Y_i$ ،  $i = 1, \dots, n$ ، نوشت. به همین دلیل، زیوانی و همکاران (۲۰۱۴)

مدل رگرسیونی را برحسب میانگین زمان‌های مورد انتظار،  $\tau_i$ ، تعریف کردند. برای این منظور، رابطه

$$\log(E(\tau_i|x_i)) = \log\left(\frac{\alpha}{\gamma(x_i)}\right) = -x_i'\beta = -\eta_i$$

را در نظر بگیرید، که در آن  $\beta = (\beta_1, \dots, \beta_p)'$  و  $x_i' = (x_{i1}, \dots, x_{ip})$  به ترتیب بردارهای ضرایب رگرسیونی و متغیرهای تبیینی را نشان می‌دهند. با فرض  $\gamma(x_i) = \alpha \exp(\eta_i)$  مدل رگرسیونی GC، برای  $i = 1, \dots, n$ ، به صورت  $Y_i|x_i, \eta_i \sim GC(\alpha, \alpha \exp(\eta_i))$  معرفی می‌شود. اگر  $\mathcal{D}$  مجموعه همه مشاهدات باشد، تابع درستنمایی مدل را می‌توان به صورت

$$L(\eta_i|\mathcal{D}) = \prod_{i=1}^n \{G(y_i\alpha, \alpha \exp(\eta_i)) - G((y_i + 1)\alpha, \alpha \exp(\eta_i))\} \quad (3)$$

نوشت. از آن‌جا که تابع درستنمایی (۳) بر اساس  $\eta_i$  غیرخطی و پیچیده است، برآوردهای ماکسیم درستنمایی پارامترهای مدل دارای نمایش بسته نیستند و باید با استفاده از روش‌های عددی محاسبه شوند (وینکلن، ۱۹۹۵؛ زیوانی و همکاران، ۲۰۱۴).

### ۳ وابستگی فضایی-زمانی در مدل گامای شمارشی

فرض کنید  $Y_{st}$  متغیر پاسخ شمارشی در مکان  $s$  و زمان  $t$  باشد، به‌طوری که تعداد کل نقاط مکانی  $S$  و تعداد کل نقاط زمانی  $T$  هستند. نقاط مکانی می‌توانند نواحی مختلف جغرافیایی در یک ناحیه مورد علاقه، مانند شهرهای یک استان باشند. با توجه به مثال کاربردی این مقاله، تمرکز تنها بر فرآیندهای تصادفی فضایی-زمانی شبکه‌ای است و تحلیل داده‌های فضایی-زمانی از نوع زمین‌آماري مورد نظر نیست. مدل رگرسیونی گامای شمارشی فضایی-زمانی به صورت

$$Y_{st}|\eta_{st} \sim GC(\alpha, \alpha \exp(\eta_{st})), \quad s = 1, \dots, S, \quad t = 1, \dots, T$$

تعریف می‌شود، به‌طوری که پیشگوی ساختاری  $\eta_{st}$  را می‌توان به صورت

$$\eta_{st} = \beta_0 + x_{st}'\beta + f_S(z_s) + f_T(t) + f_{ST}(z_s, t) \quad (4)$$

نوشت، که در آن  $\beta_0$  عرض از مبدا،  $x_{st}$  بردار متغیرهای تبیینی مربوط به مکان  $s$  و زمان  $t$ ، و  $\beta$  بردار ضرایب رگرسیونی متناظر با آن‌هاست. همچنین  $f_S(z_s)$  اثر فضایی،  $f_T(t)$  اثر زمانی و  $f_{ST}(z_s, t)$  اثر متقابل فضایی-زمانی است. در متون آمار فضایی-زمانی برای وارد کردن اثرات پنهان فضایی، زمانی و فضایی-زمانی به مدل، از فرآیندهای تصادفی مشخصی استفاده می‌شود (کرس و ویکل، ۲۰۱۱). در چارچوب بیزی، این فرآیندها در قالب توزیع‌های پیشین اثرات پنهان تعریف می‌شوند. در ادامه، به جزئیات هر کدام به‌طور جدا پرداخته می‌شود.

### ۳.۱ مدل‌های تصادفی برای اثرات پنهان فضایی و زمانی

اثر فضایی: از آن جا که در داده‌های شبکه‌ای، نواحی نزدیک‌تر از یکدیگر تاثیر بیشتری می‌پذیرند، یکی از ابزارهای مورد توجه در زمینه مدل‌بندی این نوع از داده‌ها، مدل‌بندی ضرایب رگرسیونی و اثرات تصادفی بر اساس نظریه میدان‌های تصادفی مارکوفی گاوسی ذاتی<sup>۱</sup> است (رو و همکاران، ۲۰۰۹). اثر تصادفی فضایی را می‌توان به صورت

$$f_S(z_s) = \nu_s + u_s \quad (5)$$

به دو مولفه ساختاری و غیرساختاری تجزیه و از هر دو مولفه در مدل استفاده کرد. مولفه  $\nu_s$  یک اثر فضایی غیرساختاری است که برای آن توزیع پیشین  $\mathcal{N}(0, \tau_\nu^{-1})$  در نظر گرفته می‌شود (بلانگیارد و کاملتی، ۲۰۱۵). این اثر در واقع وظیفه مدل‌بندی خطای اندازه‌گیری در نقاط مکانی را دارد و  $\tau_\nu$  پارامتر دقت آن است. برای اثر فضایی ساختاری  $u_s$  نیز معمولاً از یک فرآیند اتورگرسیو شرطی ذاتی<sup>۲</sup> (ICAR) به صورت

$$u_s | \mathbf{u}_{-s} \sim \mathcal{N}_S \left( \sum_{\ell: \ell \neq s} \frac{w_{s\ell}}{w_{s+}} u_\ell, \frac{\sigma_u^2}{w_{s+}} \right), \quad s = 1, \dots, S, \quad \ell = 1, \dots, S$$

<sup>۱</sup>Intrinsic Gaussian Markov Random Fields

<sup>۲</sup>Intrinsic Conditional Auto Regressive

استفاده می‌شود، که در آن  $w_{s+} = \sum_{s \neq \ell} w_{s\ell}$  و  $u_{-s} = \{u_\ell : \ell \neq s\}$  به طوری که

$$w_{s\ell} = \begin{cases} 1 & \text{اگر ناحیه } s \text{ با ناحیه } \ell \text{ همسایه باشد} \\ 0 & \text{اگر } s = \ell \\ 0 & \text{اگر ناحیه } s \text{ با ناحیه } \ell \text{ همسایه نباشد} \end{cases}$$

با این تعریف فرآیند شرطی (بی‌سگ و همکاران، ۱۹۹۱)، توزیع توام اثر ساختاری به صورت

$$u_s \sim \mathcal{N}_S(\cdot, \tau_u^{-1} R_u) \quad (۶)$$

خواهد بود، که در آن  $R_u = (D - W)^{-1}$ ، ماتریسی با درایه‌های  $w_{s\ell}$  و  $D$  ماتریس قطری با درایه‌های روی قطر  $w_{s+}$ ،  $s = 1, \dots, S$ ، است. همچنین  $\tau_u = 1/\sigma_u^2$ . اثر زمانی: اثر زمانی را نیز می‌توان مشابه اثر فضایی به دو مولفه ساختاری و غیرساختاری به صورت

$$f_T(t) = \varepsilon_t + \gamma_t \quad (۷)$$

تجزیه کرد. البته در هر دو رابطه (۵) و (۷)، فرض بر آن است که تغییرات بزرگ مقیاس زمانی و فضایی توسط بردار متغیرهای تبیینی  $x_{st}$  مدل‌بندی می‌شوند. همچنین لازم به ذکر است که در (۴) به جای خطی وارد کردن اثرات  $x_{st}$ ، می‌توان از صورت‌های ناپارامتری برای وارد کردن اثرات غیرخطی متغیرهای تبیینی استفاده کرد (بایور و همکاران، ۲۰۱۵). برای اثر زمانی غیرساختاری  $\varepsilon_t$ ، توزیع پیشین  $\mathcal{N}(\cdot, \tau_\varepsilon^{-1})$  در نظر گرفته می‌شود. برای مدل‌بندی اثر زمانی ساختاری  $\gamma_t$  نیز هر دو فرآیندهای ICAR و قدم زدن تصادفی<sup>۱</sup> (RW) مرتبه اول و دوم به کار گرفته می‌شوند (لوآن و همکاران، ۲۰۱۶). در روش INLA به‌کارگیری فرآیندهای قدم زدن تصادفی از محبوبیت بیشتری برخوردار است. این نوع فرآیند برای مرتبه اول به صورت  $\gamma_t | \gamma_{t-1} \sim \mathcal{N}(\gamma_{t-1}, \tau_\gamma^{-1})$  است.

اثر متقابل فضایی-زمانی: مدل‌هایی که فقط اثرات زمانی یا فضایی را در بردارند، معمولاً برای مدل‌بندی مشاهداتی که در نواحی مختلف دارای تغییرات زمانی هستند، کارا نیستند و باید اثرات متقابل فضایی و زمانی نیز در مدل وارد شوند. در این مقاله، چهار نوع پیشین برای  $f_{ST}(\cdot)$  در نظر گرفته شده است که هر کدام پذیره‌های متفاوتی را با توجه به ساختار فضایی-زمانی در نظر می‌گیرند (کنور-هلد،

<sup>۱</sup>Random Walk



(۲۰۰۰). برای نمادگذاری ساده‌تر اثر فضایی-زمانی، فرض کنید  $f_{ST}(z_s, t) = \delta_{st}$ .

- اثر متقابل فضایی-زمانی نوع I. در این حالت فرض می‌شود  $\delta_{st}$  برای همه مشاهدات مستقل و هم‌توزیع است و حامل هیچ نوع تغییرپذیری ساختاری فضایی یا زمانی نیست. به عبارتی، به جز اثرات اصلی در نظر گرفته‌شده برای مکان و زمان در مدل (۴)، باقیمانده‌ها هیچ نوع ساختار فضایی یا زمانی ندارند. مدل تصادفی برای این نوع اثر متقابل به صورت  $\delta_{st} \sim \mathcal{N}_{S \times T}(\mathbf{0}, \tau_\delta^{-1} \mathbf{I})$  است، که در آن پارامتر دقت اثر متقابل و  $\mathbf{I}$  ماتریس همانی با بعد مناسب (در اینجا  $ST$ ) را نشان می‌دهد.

- اثر متقابل فضایی-زمانی نوع II. در این نوع اثر متقابل فرض می‌شود زمان دارای اثر ساختاری است و می‌توان آن را به صورت ناپارامتری با استفاده از فرآیندهای ICAR یا RW وارد مدل کرد؛ اما در آن اثر فضایی ساختاری وجود ندارد. این به معنی آن است که اثرات متقابل، ساختار زمانی دارند اما نسبت به مکان‌ها مستقل هستند. بنابراین رابطه  $\delta_{st} \sim \mathcal{N}_{S \times T}(\mathbf{0}, \tau_\delta^{-1} \mathbf{I} \otimes \mathbf{R}_\gamma)$  برقرار است، که در آن  $\otimes$  نماد ضرب کروشه‌کر و  $\mathbf{R}_\gamma$  ماتریس کوواریانس حاصل از فرآیند تصادفی زمانی (مثل RW) را نشان می‌دهد.

- اثر متقابل فضایی-زمانی نوع III. این نوع اثر متقابل از حاصلضرب اثر ساختاری فضایی و اثر غیرساختاری زمانی تشکیل می‌شود. برای اثر ساختاری فضایی، با توجه به شبکه‌ای بودن داده‌ها، مدل معمول، ICAR است. در این حالت، فرض می‌شود اثرات متقابل در طول زمان مستقل اما دارای وابستگی فضایی هستند. بنابراین رابطه  $\delta_{st} \sim \mathcal{N}_{S \times T}(\mathbf{0}, \tau_\delta^{-1} \mathbf{R}_u \otimes \mathbf{I})$  به دست می‌آید، که در آن  $\mathbf{R}_u$  در (۶) داده شده است.

- اثر متقابل فضایی-زمانی نوع IV. در این نوع فرض می‌شود اثر متقابل فضایی-زمانی ترکیبی از هر دو نوع اثرات مکانی و زمانی ساختاری است. بنابراین اثرات متقابل در طول زمان و بر حسب موقعیت‌های مکانی قرار گرفتن مشاهدات، دارای وابستگی هستند. پس  $\delta_{st}$  دارای توزیع  $\mathcal{N}_{S \times T}(\mathbf{0}, \tau_\delta^{-1} \mathbf{R}_u \otimes \mathbf{R}_\gamma)$  است. این نوع اثر متقابل دارای پیچیدگی مضاعف است و برای برازش مدل، از نظر محاسباتی، نسبت به سایر اثرات متقابل زمان بیشتری نیاز دارد.

در مسایل کاربردی، معمولاً، با توجه به شناختی که از ذات مساله در اختیار هست، می‌توان تا حدودی نوع ساختار وابستگی اثر متقابل را قبل از برازش مدل تشخیص داد و نیاز به برازش همه مدل‌ها نیست. این مساله، به‌ویژه، زمانی که حجم داده‌ها بزرگ و برازش مدل‌های پیچیده خیلی زمان‌بر است، مهم خواهد بود.

### ۳.۲ استنباط بیزی مدل

فرض کنید  $\zeta = (u_1, \dots, u_S, \nu_1, \dots, \nu_S, \varepsilon_1, \dots, \varepsilon_t, \gamma_1, \dots, \gamma_t, \delta_{11}, \dots, \delta_{st}, \dots, \delta_{ST})'$  برداری با بعد  $m$  شامل تمام اثرات پنهان، به ازای  $t = 1, \dots, T$  و  $s = 1, \dots, S$ ، در مدل (۴) و  $\theta = (\alpha, \beta, \tau_\nu, \tau_u, \tau_\varepsilon, \tau_\gamma, \tau_\delta)'$  بردار پارامترهای مدل باشد. تابع درستنمایی کناری مدل به صورت

$$L(\theta|\mathcal{D}) = \int_{\mathbb{R}^m} f(\mathbf{y}, \zeta|\theta) d\zeta \quad (۸)$$

است، که در آن  $f(\mathbf{y}, \zeta|\theta)$  تابع درستنمایی توام بردار پاسخ  $\mathbf{y} = (y_{11}, \dots, y_{st}, \dots, y_{ST})'$  و بردار اثرات پنهان مدل را نشان می‌دهد. با توجه به پیچیدگی تابع درستنمایی مدل GC و بزرگ بودن بعد  $m$ ، حتی با فرض گاوسی بودن توزیع اثرات پنهان  $\zeta$ ، کاملاً واضح است که محاسبه تابع درستنمایی (۸) اگر ناممکن نباشد بسیار پیچیده است. روش‌های مختلفی برای برخورد با این پیچیدگی معرفی شده‌اند که همه آن‌ها در حضور بعد بالای اثرات پنهان و تابع درستنمایی پیچیده مشاهدات، ناکارآمد خواهند بود. برای رهایی از پیچیدگی‌های ذکرشده در رهیافت استنباط مبتنی بر درستنمایی، معمولاً، محققین از چارچوب بیزی برای استنباط در این دسته از مدل‌ها استفاده می‌کنند. دلیل آن هم بهره‌گیری از روش‌های مبتنی بر نمونه‌گیری مثل الگوریتم‌های MCMC برای تقریب توزیع پسین مدل است. اما همان‌طور که در مقدمه تاکید شد، این الگوریتم‌ها نیز از مشکلاتی رنج می‌برند که برای مدل‌های پیچیده فضایی جدی‌تر هم می‌شوند. با این توضیحات، رهیافت استنباطی انتخابی در این مقاله مبتنی بر روش INLA است که کارایی محاسباتی بالایی برای برازش مدل پیشنهادی ارائه می‌کند. با در نظر گرفتن توزیع پیشین  $\pi(\theta)$  برای بردار پارامترها، تابع چگالی توزیع پسین مدل به صورت  $\pi(\zeta, \theta|\mathbf{y}) \propto \pi(\theta)f(\zeta|\theta)f(\mathbf{y}|\zeta, \theta)$  محاسبه می‌شود، که در آن  $\mathbf{y}$  بردار مشاهدات پاسخ است. برای انجام استنباط بیزی، لازم است توزیع‌های پسین کناری متغیرهای پنهان و پارامترها به صورت

$$\begin{aligned} \pi(\zeta_i|\mathbf{y}) &= \int f(\zeta_i|\theta, \mathbf{y})\pi(\theta|\mathbf{y})d\theta, \quad i = 1, \dots, m \\ \pi(\theta_j|\mathbf{y}) &= \int \pi(\theta|\mathbf{y})d\theta_{-j}, \quad j = 1, \dots, \dim(\theta) \end{aligned} \quad (۹)$$

محاسبه شوند، که در آن  $\theta_{-j}$  بردار حاصل از حذف درایه  $j$  ام  $\theta$  است. روش INLA بر پایه تقریب‌های لاپلاس و انتگرال‌گیری عددی، تقریب‌هایی برای چگالی‌های پسین (۹) فراهم می‌کند که بسیار سریع محاسبه

می‌شوند و از دقت بالایی برخوردارند. روش INLA برای محاسبه تقریب‌های چگالی‌های پسین کناری سه گام مختلف را اجرا می‌کند که برای اطلاع از جزئیات آن، خواننده علاقه‌مند به [رو و همکاران \(۲۰۰۹\)](#) ارجاع داده می‌شود.

### ۳.۳ توزیع‌های پیشین ابرپارامترهای مدل

برای تحلیل مدل رگرسیونی گامای شمارشی فضایی-زمانی در چارچوب بیزی، لازم است که توزیع‌های پیشین مناسب برای پارامترها و ابرپارامترهای مدل شامل پارامتر پراکندگی توزیع گامای شمارشی،  $\alpha$ ، ضرایب رگرسیونی  $\beta$  و پارامترهای دقت  $(\tau_\nu, \tau_u, \tau_\varepsilon, \tau_\gamma, \tau_\delta)$  تعیین شوند. با پذیرفتن استقلال پیشینی برای پارامترها و ابرپارامترهای مدل می‌توان نوشت  $\pi(\theta) = \pi(\alpha)\pi(\beta)\pi(\tau_\nu)\pi(\tau_u)\pi(\tau_\varepsilon)\pi(\tau_\gamma)\pi(\tau_\delta)$ . با توجه به محدودیت روش INLA ([رو و همکاران، ۲۰۰۹](#))، مولفه‌های بردار ضرایب رگرسیونی  $\beta$ ، به‌صورت مستقل، دارای توزیع پیشین گاوسی با میانگین صفر و یک واریانس ثابت بزرگ هستند که در این مقاله این مقدار برابر ۱۰۰۰ در نظر گرفته شده است. این انتخاب تا حدودی محقق را مطمئن می‌سازد که یک توزیع پیشین سره ناآگاهی‌بخش<sup>۱</sup> برای مولفه‌های بردار ضرایب رگرسیونی انتخاب شده است.

برای ابرپارامترهای دقت اثرات پنهان، اخیراً استفاده از توزیع‌های پیشین جدیدی توصیه می‌شود که توسط [سیمپسون و همکاران \(۲۰۱۷\)](#) معرفی شده است. آن‌ها توزیع‌های پیشین پیشنهادی خود را پیشین‌های با پیچیدگی تاوانیده<sup>۲</sup> (PC) نامیدند. دو مولفه مهم در ساختن این نوع پیشین، تعریف مدل‌های پایه و منعطف است، به‌گونه‌ای که توزیع پیشین از جریمه کردن مدل منعطف، با پیچیدگی بیشتر نسبت به مدل پایه، ساخته می‌شود. به‌عنوان مثال، مدلی را در نظر بگیرید که پیشگوی خطی آن شامل یک اثر تصادفی دارای توزیع نرمال با میانگین صفر و پارامتر دقت  $\tau = 1/\sigma$  است. در این حالت مدل پایه متناظر با ساده‌ترین نسخه از مدل منعطف با اثر تصادفی است؛ یعنی مدل پایه آن است که اثر تصادفی در آن غایب باشد. در واقع در مدل پایه تمام وزن به  $\sigma = 0$  داده می‌شود. پس توزیع پیشین PC طوری تعریف می‌شود که جریمه بیشتری به مقادیر دورتر از صفر  $\sigma$  تحمیل می‌کند.

یکی از جذابیت‌های پیشین‌های PC، این است که با استفاده از عبارت‌های احتمالی در مورد پارامتر ساخته می‌شوند. در چهار اصل طراحی پیشین‌های PC، اصل اول دلالت بر اصل امساک دارد و تأکید آن بر این است که پیشین از مدل پایه پشتیبانی می‌کند مگر این‌که داده‌ها (با استفاده از درست‌نمایی) خلاف آن

<sup>۱</sup>Proper Noninformative

<sup>۲</sup>Penalized Complexity Priors

را نشان دهند. **سیمپسون و همکاران** (۲۰۱۷)، در اصل دوم، برای محاسبه پیچیدگی مدل منعطف نسبت به مدل پایه، استفاده از فاصله کولبک-لایبلر<sup>۱</sup> بین دو مدل را پیشنهاد کردند. در اصل سوم، میزان جریمه بر اساس تابعی از فاصله کولبک-لایبلر با یک نرخ ثابت، توسط یک مدل نمایی، اندازه‌گیری می‌شود. در نهایت، بر پایه اصل چهارم، پیشین PC با تعریف مناسب عبارت‌های احتمالی بر روی پارامترهای مدل، محاسبه می‌شود. برای اطلاعات بیشتر نسبت به جزییات نظری نحوه ساخت این نوع پیشین می‌توانید به **سیمپسون و همکاران** (۲۰۱۷) و **گومز** (۲۰۲۰) مراجعه کنید. استقبال گسترده از پیشین‌های PC به‌ویژه برای پارامترهای دقت اثرات تصادفی (**کلین و کنیب**، ۲۰۱۶؛ **گومز**، ۲۰۲۰)، باعث شده است این پیشین‌ها به توزیع پیشین پیش‌فرض برای برازش مدل‌های فضایی و فضایی-زمانی در روش INLA تبدیل شوند. برای یک پارامتر دقت دلخواه مثل  $\tau$ ، ضابطه تابع چگالی توزیع پیشین PC به صورت

$$\pi(\tau) = \frac{\lambda}{\tau} \tau^{-3/2} \exp(-\lambda \tau^{-1/2}) \quad (۱۰)$$

است، که در آن  $\lambda = -\frac{\ln(a)}{p}$  یک مقدار مثبت و  $(p, a)$  پارامترهای این پیشین را نشان می‌دهد. مقادیر  $(p, a)$  طوری تعیین می‌شوند که رابطه  $P(\sigma > p) = a$ ، برای  $0 < p < 1$  و  $0 < a < 1$ ، برقرار باشد، که در آن  $\sigma = 1/\tau$ . در این مقاله برای تمام پارامترهای دقت مدل، توزیع پیشین (۱۰) در نظر گرفته شده است و مقادیر  $(p, a)$  برابر  $(1, 0.01)$  تعیین شده‌اند. تفسیر مقادیر تعیین‌شده برای  $(p, a)$  بر این اساس است که توزیع پیشین تعریف‌شده، احتمال زیادی را برای میل دادن انحراف معیار  $\sigma$  برای یک اثر تصادفی (مثل اثر فضایی) به سمت صفر در نظر می‌گیرد. این تمایل به معنی عدم وجود اثر فضایی در مدل است که همان مفهوم مدل پایه و پشتیبانی از آن است. اگر اثر فضایی قوی و وجود آن معنی‌دار باشد، مدل منعطف بیشتر توسط داده‌ها پشتیبانی می‌شود.

برای پارامتر پراکندگی توزیع گامای شمارشی،  $\alpha$ ، نیز به پیروی از **نادی‌فر و همکاران** (۲۰۱۹) توزیع پیشین  $\text{Gamma}(c, d)$  با تابع چگالی  $\pi(\alpha; c, d) = \frac{d^c}{\Gamma(c)} \alpha^{c-1} \exp(-d\alpha)$  در نظر گرفته شده است که مقادیر ابرپارامترهای آن برابر با  $(1, 0.005)$  تعیین شدند. لازم به ذکر است که برای مدل‌های رقیب پواسون و NB، از توزیع‌های پیشین مشابه استفاده شده است، با ذکر این نکته که برای پارامتر پراکندگی در مدل NB،  $\phi$ ، توزیع پیشنهادی PC معرفی شده در **گومز** (۲۰۲۰) به کار گرفته شده است.

<sup>1</sup>Kullback-Leibler Distance

## ۴ مطالعه شبیه‌سازی

هدف از این مطالعه، مقایسه مدل رگرسیونی فضایی-زمانی گامای شمارشی با رقبای سنتی، یعنی مدل‌های فضایی-زمانی پواسون و NB، در شرایط مختلف پراکندگی است. به منظور فراهم آوردن یک مطالعه جامع، اثر متقابل فضایی-زمانی نوع IV در مدل شبیه‌سازی شده وارد شده است. برآزش همه مدل‌ها در این بخش و در مثال کاربردی در محیط نرم‌افزار R و به کمک توابع موجود در بسته R-INLA انجام شده است. مشاهدات پاسخ بر اساس ماتریس مجاورت استان‌های ایران (۳۱ استان) و دنباله زمانی ۱۲ تایی از مدل

$$Y_{st}|\eta_{st} \sim \text{GC}(\alpha, \alpha \exp(\eta_{st}))$$

$$\eta_{st} = \beta_0 + \beta_1 x_{st} + u_s + \nu_s + \varepsilon_t + \gamma_t + \delta_{st}, \quad s = 1, \dots, 31; \quad t = 1, \dots, 12$$

تولید شدند، که در آن ضرایب واقعی رگرسیونی  $(\beta_0, \beta_1) = (0/5, 0/9)$  تعیین و مقادیر متغیر تبیینی  $x_{st}$  از توزیع یکنواخت بر بازه  $(-1, 1)$  تولید شدند. برای لحاظ کردن شکل‌های مختلف پراکندگی، شامل بیش‌پراکنش، پراکنش متعادل و کم‌پراکنش، برای پارامتر  $\alpha$  به ترتیب مقادیر  $0/8$ ،  $1$  و  $1/5$  در نظر گرفته شدند. بردار  $u = (u_1, \dots, u_{31})$  به‌عنوان تحقق از اثر ساختاری فضایی از یک مدل ICAR با ماتریس مجاورت  $W$  و ماتریس دقت  $\tau_u^{-1} R_u$  تولید شد، به‌طوری که  $W$  ماتریس مجاورت ساخته‌شده از ۳۱ استان ایران است. پارامتر دقت این اثر برابر  $\tau_u = 2$  در نظر گرفته شد. به‌طور مشابه، بردار  $\gamma = (\gamma_1, \dots, \gamma_{12})$  به‌عنوان تحقق از اثر ساختاری زمانی از یک مدل ICAR با ماتریس مجاورت  $P$  و ماتریس دقت  $\tau_\gamma^{-1} R_\gamma$  شبیه‌سازی شد، که در آن  $P$  ماتریس مجاورت زمانی و  $\tau_\gamma = 2$  تعیین شد. همچنین تحققاتی از اثرات غیرساختاری  $\nu_s$ ،  $s = 1, \dots, 31$  و  $\varepsilon_t$ ،  $t = 1, \dots, 12$ ، از توزیع  $\mathcal{N}(0, 1)$  تولید شدند. در نهایت، اثر متقابل فضایی-زمانی از توزیع گاوسی  $\mathcal{N}_{12 \times 31}(0, \tau_\delta^{-1} R_\gamma \otimes R_u)$  با فرض  $\tau_\delta = 3$  شبیه‌سازی شد. برای در نظر گرفتن عدم قطعیت حاکم بر شبیه‌سازی نمونه‌های تصادفی، تولید مجموعه داده‌ها ۱۰۰ بار تکرار شد. با داشتن مشاهدات  $\{(y_{st}, x_{st}); s = 1, \dots, 31, t = 1, \dots, 12\}$  پارامترهای مدل‌های رگرسیونی گامای شمارشی، پواسون و NB، با روش بیزی تقریبی INLA و لحاظ کردن اثر متقابل فضایی-زمانی نوع IV، برآورد شدند. برای ارزیابی نتایج حاصل از

مدل‌های مختلف، مقادیر ریشه میانگین توان‌های دوم خطا<sup>۱</sup> (RMSE) و اریبی نسبی<sup>۲</sup> (RB) آن‌ها که از رابطه  $E(\frac{\hat{\beta}}{\beta} - 1)$  به دست می‌آید، برای برآوردگرهای پارامترهای رگرسیونی مختلف و پارامتر پراکندگی مدل GC محاسبه و در جدول ۱ ارایه شده‌اند. نتایج حاصل از دو معیار، تقریباً برتری مدل رگرسیونی گامای شمارشی را نسبت به مدل‌های رقیب، در حالت‌های بیش‌پراکنش ( $\alpha = 0.8$ ) و کم‌پراکنش ( $\alpha = 1.5$ ) نشان می‌دهند. برای  $\alpha = 1$ ، نتایج حاصل از دو مدل گامای شمارشی و پواسون بسیار نزدیک به هم و هر دو بهتر از مدل NB هستند. البته این نتیجه مورد انتظار است، زیرا در این حالت مدل GC به مدل پواسون تبدیل می‌شود.

جدول ۱. مقادیر RMSE و RB پارامترهای مدل‌های رگرسیونی شمارشی							
ضریب			پارامتر			پراکندگی	
RB			RMSE				
NB	پواسون	GC	NB	پواسون	GC		
0.189	0.137	0.019	0.176	0.162	0.149	$\beta_0$	0.8
-0.029	-0.031	0.000	0.082	0.083	0.088	$\beta_1$	
		0.080			0.164	$\alpha$	
0.382	0.070	0.030	0.915	0.156	0.154	$\beta_0$	1
-0.010	-0.004	-0.004	0.124	0.084	0.101	$\beta_1$	
		-0.029			0.227	$\alpha$	
-0.013	0.038	0.007	0.605	0.168	0.128	$\beta_0$	1.5
-0.357	-0.229	-0.103	0.632	0.073	0.062	$\beta_1$	
		-0.122			0.274	$\alpha$	

در کنار ارزیابی کارایی برآورد پارامترها، از اهداف دیگر شبیه‌سازی سنجش نیکویی برازش و توان پیشگویی مدل‌ها بوده است. از دو معیار اطلاع انحراف<sup>۳</sup> (DIC) (اشپیگل‌هالتر و همکاران، ۲۰۰۲) و واتانابه-آکاییک<sup>۴</sup> (WAIC) (واتانابه، ۲۰۱۳؛ گل‌من و همکاران، ۲۰۱۴) برای این ارزیابی استفاده شده است. معیار DIC برای ارزیابی پیچیدگی و نیکویی برازش مدل‌ها و معیار WAIC برای بررسی عملکرد توان پیشگویی مدل‌ها در پیشگویی داده‌های جدید به کار گرفته شدند. مدلی که این معیارها برای آن کمترین باشد، مدل برتر است. جدول ۲ مقادیر این دو اندازه را در قالب نرخ برتری نسبی مدل رگرسیونی گامای

<sup>1</sup>Root Mean Squared Errors

<sup>2</sup>Relative Bias

<sup>3</sup>Deviance Information Criterion

<sup>4</sup>Watanabe-Akaike Information Criterion

شمارشى نسبت به مدل‌هاى پواسون و دوجمله‌اى منفى نمايش مى‌دهد. نرخ برترى نسبى برحسب

تعداد دفعات برتر بودن مدل رگرسيونى گامائى شمارشى نسبت به مدل رقيب  
تعداد دفعات تکرار

محاسبه شده است. هر اندازه مقدار اين معيار به ۱ نزديک‌تر باشد، برترى مدل GC را نسبت به مدل‌هاى رقيب نشان مى‌دهد. نتايج جدول را مى‌توان به صورت زير خلاصه کرد:

**الف-** براى حالتى که پراکنش در داده‌ها متعادل است، هر دو معيار یک تعادل تقريبى را بين دو مدل GC و پواسون نشان مى‌دهند. در اين حالت نتيجه مشابهى براى دو مدل GC و NB بر حسب معيار DIC حاصل شده است، ولى معيار WAIC کاملاً برترى مدل GC را نسبت به NB نشان مى‌دهد. لازم به يادآورى است که مدل NB تنها قادر به مدل‌بندى بيش‌پراکنش است.

**ب-** براى داده‌هاى بيش‌پراکنده، يعنى  $\alpha = 0.8$ ، معيار DIC یک برترى نسبى را براى GC در مقابل مدل پواسون نشان مى‌دهد. نتيجه براى مدل NB برعکس است. در اين حالت به نظر مى‌رسد مدل رقيب NB که قادر به مدل‌بندى داده‌هاى بيش‌پراکنده است، برازش بهترى را نشان مى‌دهد. نکته جالب در اين حالت، مقادير معيار WAIC است که توان پيشگويى مدل‌ها را مد نظر دارد. مدل GC از NB کاملاً برتر و برعکس نسبت به مدل پواسون مغلوب است. اين نتيجه براى داده‌هاى بيش‌پراکنده که مدل پواسون قادر به شناسايى بيش‌پراکندگى نيست کمى عجيب است. البته شايد بتوان نتيجه را به پيچيده‌تر بودن مدل GC نسبت به پواسون نسبت داد که مى‌تواند در عملکرد پيشگويى تاثيرگذار باشد.

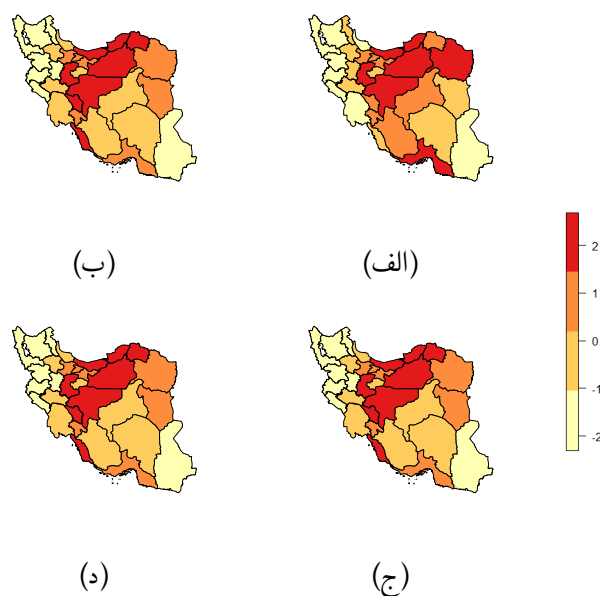
**ج-** در حالت کم‌پراکنش، معيار WAIC تمام قدرت را به مدل GC مى‌بخشد، به طورى که تقريباً ۱۰۰٪ نسبت به دو رقيب خود برترى دارد. اما مساله عجيب ديگر نتيجه معيار DIC است که بر اساس آن، مدل GC از هر دو رقيب شکست مى‌خورد. باز هم ذکر یک نکته در اين جا خالى از لطف نيست و آن اين‌که، معيار DIC یک معيار طلايى براى مقايسه مدل‌ها، به ويژه مدل‌هاى پيچيده فضايى-زمانى، نيست. **گلمن و همکاران (۲۰۱۴)** با توجه به نقاط ضعف اين معيار، استفاده از معيار WAIC را توصيه مى‌کنند.

با نتايج ذکرشده، مى‌توان گفت که مدل GC، به ويژه براى داده‌هاى داراى کم‌پراکنش، برترى خوبي نسبت به رقبائى خود دارد. شکل‌هاى ۱، ۲ و ۳، نقشه پهنه‌بندى اثر ساختارى فضايى شبیه‌سازى شده و نقشه‌هاى متناظر برآورد پسینى اين اثر را، در یک تکرار شبیه‌سازى، براى مدل‌هاى مورد مطالعه و مقادير مختلف پارامتر پراکندگى  $\alpha$  نمايش مى‌دهند. براى دو حالت بيش‌پراکنده و پراکنده متعادل، برآوردهاى پسینى اثر ساختارى فضايى براى هر سه مدل رقيب تقريباً یکسان و به اثر واقعى نیز نزديک هستند. در

جدول ۰۲. نرخ برتری نسبی مدل GC نسبت به مدل‌های رقیب بر اساس معیارهای DIC و WAIC

WAIC		DIC		پارامتر پراکندگی
NB	پواسون	NB	پواسون	
۰/۸۵۹	۰/۱۰۸	۰/۴۳۲	۰/۶۴۹	۰/۸
۰/۸۴۹	۰/۴۳۶	۰/۵۲۶	۰/۵۷۷	۱
۰/۸۸۳	۰/۸۴۹	۰/۲۰۳	۰/۲۲۰	۱/۵

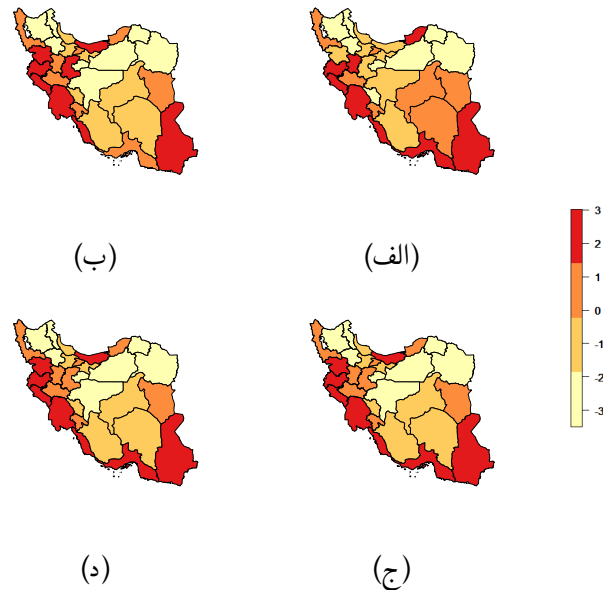
مقابل وقتی که مشاهدات کم‌پراکنده هستند، اثر فضایی ساختاری برآوردشده با مدل‌های GC و پواسون تقریب دقیق‌تری از اثر واقعی را نسبت به مدل NB نشان می‌دهند. نتیجه برای مدل NB با مقدار واقعی تفاوت دارد که به ناتوانی این مدل برای لحاظ کردن کم‌پراکندگی در داده‌ها اشاره دارد. نتیجه برای دو مدل پواسون و GC تقریباً یکسان است. شکل ۴ نیز نمودارهای برآورد اثر ساختاری زمانی (برحسب میانگین



شکل ۱. اثر فضایی اصلی (الف) و برآورد پسینی اثرهای فضایی ساختاری در مدل‌های GC (ب)، پواسون (ج) و NB (د) به‌ازای  $\alpha = ۰/۸$

پسینی) را برای هر یک از مدل‌ها به‌ازای پراکندگی‌های مختلف، در کنار اثر واقعی شبیه‌سازی‌شده، در یک تکرار شبیه‌سازی نشان می‌دهد. با توجه به قطعه (آ)، برای حالت بیش‌پراکنش، اثرات زمانی برآوردشده



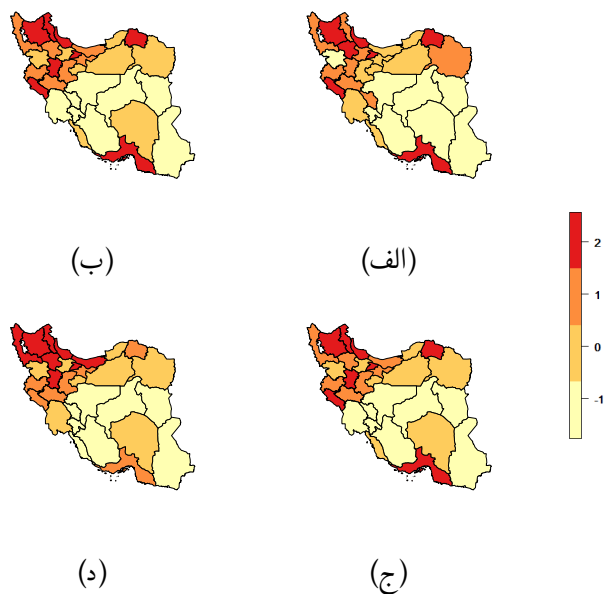


شکل ۲. اثر فضایی اصلی (الف) و برآورد پسینی اثرهای فضایی ساختاری در مدل‌های GC (ب)، پواسون (ج) و NB (د) به‌ازای  $\alpha = 1$

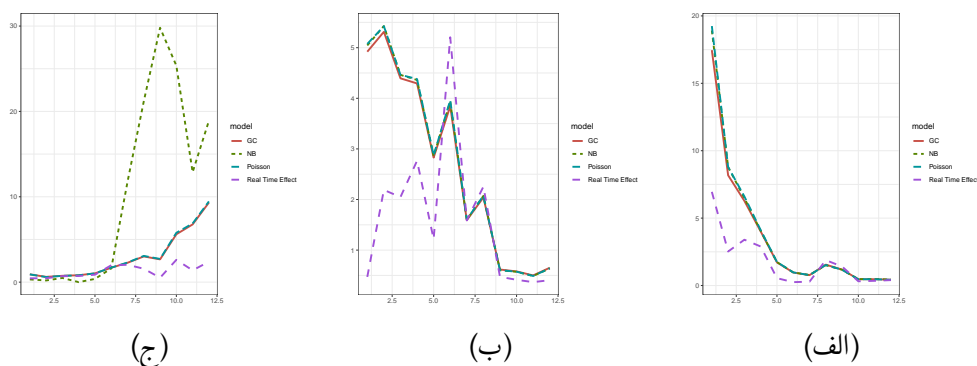
توسط سه مدل، مشابه هم و تقریباً به واقعیت نزدیک هستند. در قطعه (ب)، برای حالت پراکنش متعادل، اثرات زمانی برآوردشده با اثر زمانی واقعی، به‌ویژه در نیمه اول بازه زمانی، تفاوت زیادی دارند. اما در نیمه دوم زمانی، اثرات برآوردشده توسط سه مدل، تقریباً با اثر واقعی یکسان هستند. در قطعه (پ)، برای حالت کم‌پراکنش، مدل‌های GC و پواسون نسبت به NB عملکرد خیلی بهتری در بازایی اثر واقعی زمانی دارند. نمایش تصویری اثرات برآوردشده فضایی و زمانی نیز تاییدی بر برتری مدل پیشنهادی، به‌ویژه، برای حالتی است که داده‌ها دارای کم‌پراکنش هستند.

## ۵ مثال کاربردی

برای نمایش کاربردی مدل پیشنهادی برای تحلیل داده‌های واقعی و مقایسه آن با مدل‌های رقیب، از داده‌های مربوط به تعداد افراد مبتلا به سرطان خون، طی سال‌های ۱۳۹۱ تا ۱۳۹۴ در شانزده شهرستان استان خراسان رضوی استفاده شده است. سرطان خون از انواع سرطان است که معمولاً از مغز استخوان شروع و باعث



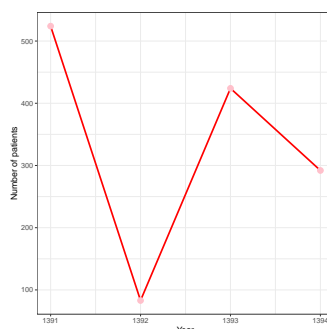
شکل ۳. اثر فضایی اصلی (الف)، و برآورد پسینی اثرهای فضایی ساختاری در مدل‌های GC (ب)، پواسون (ج) و NB (د) به‌ازای  $\alpha = ۱/۵$



شکل ۴. برآورد پسینی اثرهای زمانی ساختاری برای  $\alpha = ۰/۸$  (الف)،  $\alpha = ۱$  (ب) و  $\alpha = ۱/۵$  (ج)

شکل‌گیری تعداد زیادی گلبول سفید غیرطبیعی می‌شود. این گلبول‌های سفید خونی به صورت کامل تشکیل نمی‌شوند و به آن‌ها سلول‌های لوکمی یا سرطان خون گفته می‌شود. بر اساس گزارش‌های مرکز بهداشت

استان خراسان رضوى، نرخ بروز سرطان خون ۴/۳۵ در هر يكصد هزار نفر بوده و هفتمين سرطان شايع در اين استان است. شكل ۵ نمودار تعداد افراد مبتلا به سرطان خون در اين استان را به تفكيك چهار سال مورد نظر نشان مى دهد. مدل در نظر گرفته شده براى تحليل اين داده ها به صورت



شكل ۵. تعداد افراد مبتلا به سرطان خون طى سال هاى ۱۳۹۱ تا ۱۳۹۴ در استان خراسان رضوى

$$Y_{st}|\eta_{st} \sim \text{GC}(\alpha, \alpha E_{st} \exp(\eta_{st}))$$

$$\eta_{st} = \beta_0 + u_s + \nu_s + \varepsilon_t + \gamma_t + \delta_{st}, \quad s = 1, \dots, 16; \quad t = 1, \dots, 4$$

است، كه در آن  $Y_{st}$  متغير پاسخ، يعنى تعداد افراد مبتلا به سرطان خون، در شانزده شهرستان و در چهار نقطه زمانى است. همچنين  $E_{st}$  تعداد افراد مبتلاى مورد انتظار در شهرستان  $s$  و سال  $t$  است كه به صورت

$$E_{st} = \frac{N_{st} Y_{st}}{\sum_{s=1}^{16} \sum_{t=1}^4 N_{st}}$$

محاسبه مى شود، كه در آن  $N_{st}$  جمعيت شهرستان  $s$  در سال  $t$  و  $E_{st}$  به عنوان تعديل كننده وارد مدل شده است. اثرهاى پنهان حاضر در پيشگوى  $\eta_{st}$  همان اثرهاى تشريح شده در بخش ۳ هستند. در اين داده ها اطلاعات ديگرى كه بتوان آن ها را در قالب متغيرهاى تبينى رگرسيونى وارد مدل كرد، وجود نداشت.

مدل فضايى-زمانى با چهار نوع اثر متقابل فضايى-زمانى معرفى شده در بخش ۳، براى سه مدل GC، پواسون و NB به داده ها برازش داده شدند. معيارهاى DIC و WAIC براى همه مدل ها در جدول ۳ گزارش شده اند. با توجه به عدم همگرابى مدل GC براى اثر متقابل نوع IV، مقادير دو معيار در اين

حالت محاسبه و گزارش نشدند. با توجه به نتایج جدول، مدل GC در حالتی که اثر متقابل فضایی-زمانی نوع I انتخاب شود، در بین همه مدل‌ها، کمترین مقادیر را برای هر دو معیار DIC و WAIC داراست. با این نتیجه، در ادامه، داده‌ها بر اساس مدل با اثر متقابل نوع I تحلیل شده‌اند. برای بررسی لزوم وارد کردن اثر متقابل فضایی-زمانی در تحلیل داده‌ها، معیارهای DIC و WAIC برای هر سه مدل رقیب و بدون حضور اثر متقابل نیز محاسبه و در جدول ۳ گزارش شدند. به جز مدل NB، مقادیر این معیارها، با اختلافی قابل ملاحظه، از مقادیر متناظر در مدل‌های با حضور اثر متقابل بزرگتر هستند که نشان‌دهنده لزوم وارد کردن این اثر تصادفی است. فراتر از آن، برای اطمینان از لزوم وارد کردن اثرات فضایی و زمانی، معیارهای برازش برای مدل‌های بدون هر نوع اثر تصادفی فضایی و زمانی نیز محاسبه شدند که به قدرت معنی‌داری این اثرات در هر سه مدل رقیب تایید می‌شوند. برآورد پارامترهای سه مدل، به همراه فاصله‌های

جدول ۳. معیارهای WAIC و DIC برای داده‌های سرطان خون با اثرات پنهان مختلف						
معیار	مدل	نوع اثر متقابل فضایی-زمانی				فاقد اثر متقابل
		I	II	III	IV	
DIC	GC	۳۳۵/۳۷	۴۴۵/۳۰	۳۴۱/۷۷	۳۹۳/۴۸۱	۴۴۵/۶۲۴
	پواسون	۳۳۸/۰۴	۳۳۹/۳۹	۳۴۰/۲۷	۴۳۵/۰۷۸	۷۷۵/۶۳۰
	NB	۳۸۶/۹۷	۳۹۱/۸۴	۳۸۵/۷۷	۳۸۷/۰۴۱	۴۲۸/۲۰۳
WAIC	GC	۳۳۰/۷۲	۳۹۳۸۶/۶۸	۳۳۷/۳۹	۴۰۱/۲۱۷	۴۴۷/۷۷۹
	پواسون	۳۳۲/۰۴	۳۳۱/۱۶	۳۳۶/۴۷	۴۸۳/۹۸۵	۷۸۱/۶۷۵
	NB	۳۸۸/۹۸	۳۸۹/۵۵	۳۸۸/۰۳	۳۸۹/۰۱۷	۴۲۸/۶۸۲

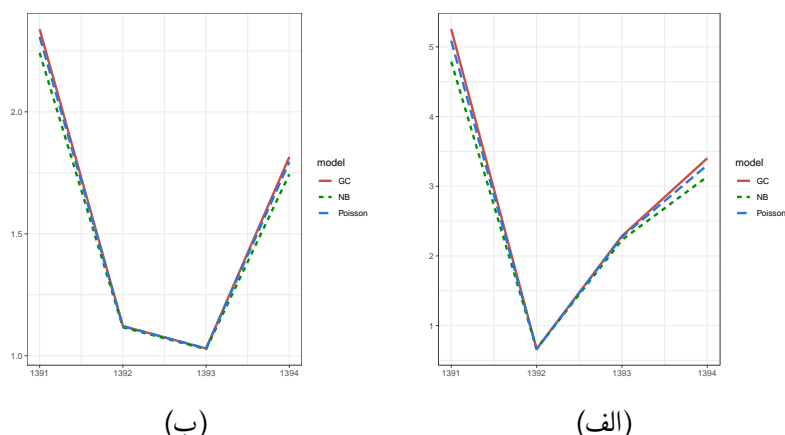
معتبر بیزی ۹۵٪ آن‌ها در جدول ۴ گزارش شده‌اند. نتایج برای هر سه مدل، تقریباً شبیه هم هستند. تنها استثناء برآورد پارامتر دقت اثر متقابل فضایی-زمانی  $T_\delta$  است که اختلاف زیادی بین مدل NB با دو مدل دیگر وجود دارد. در واقع تغییرات متقابل فضایی-زمانی در مدل NB (بر اساس برآورد بزرگ پارامتر دقت اثر متقابل)، بسیار ناچیز برآورد شده است. در حالی که در دو مدل GC و پواسون این تغییرات تقریباً سهمی به اندازه تغییرات فضایی دارد. با توجه به برآورد ۱/۰۴ برای  $\alpha$  و فاصله معتبر شامل ۱ برای آن، در مدل GC، انتظار بر آن است که مدل‌های گامای شمارشی و پواسون عملکرد مشابهی داشته باشند. نتایج جدول نیز تاییدکننده این مطلب است و برآوردهای گزارش شده برای این دو مدل به هم نزدیک‌تر هستند تا مدل NB.

جدول ۴. برآورد پارامترها و فاصله باورمند بیزی ۹۵٪ متناظر برای سه مدل رقیب در داده‌های سرطان خون

مدل												پارامتر
NB				پواسون				GC				
U	L	SE	میانگین	U	L	SE	میانگین	U	L	SE	میانگین	
۲,۳۸	۰/۹۹	۰/۳۵۱	۱/۷۰	۲,۲۷	۰/۸۴	۰/۳۶۲	۱,۵۷	۲,۲۵	۰/۸۵	۰/۳۵۹	۱,۵۶	$\beta_0$
۶,۲۱	۰/۹۰	۱/۵۴	۳/۱۹	۶,۲۵	۰/۸۹	۱/۵۵	۳,۲۱	۶,۳۰	۰/۸۹	۱/۵۷	۳,۲۲	$T_u$
۰/۸۱	۰/۰۰	۰/۲۵	۰/۳۱	۰/۸۴	۰/۰۰	۰/۲۶	۰/۳۲	۰/۸۳	۰/۰۰	۰/۲۶	۰/۳۱	$T_v$
۵۳۲۵۸	۱/۱۲	۱۸۵۰۰	۱۸۷۶۹	۴۹۸۱۲	۰/۴۰	۱۷۲۰۰	۱۷۷۰۸	۵۱۲۳۹	۰/۶۲	۱۷۸۰۰	۱۸۱۰۱	$T_\gamma$
۱۲/۹۰	۰/۰۰	۴/۳۶	۴/۷۷	۱۲/۳۷	۰/۱۱	۴/۲۹	۴/۵۲	۱۲/۵۲	۰/۱۱	۴/۳۴	۴/۵۸	$T_\varepsilon$
۵۳۶۸۵	۱/۲۹	۱۸۵۰۰	۱۸۹۶۹	۷/۳۲	۱/۷۲	۱/۵۴	۴/۲۷	۷/۳۰	۱/۷۵	۱/۵۴	۴/۲۲	$T_\delta$
			— —				— —	۱/۴۲	۰/۷۰	۰/۱۹	۱/۰۴	$\alpha$

SE: خطای استاندارد، L: کران پایین، U: کران بالا

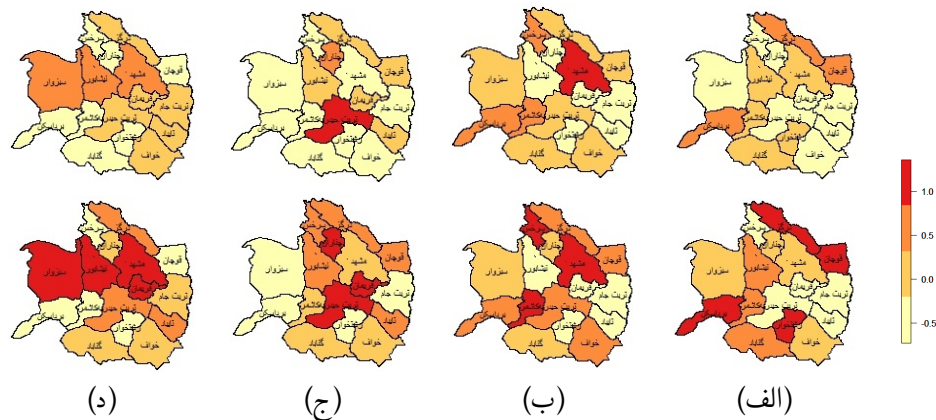
شکل ۶ میانگین پسینی اثرات ساختاری و غیرساختاری زمانی را برای هر سه مدل نشان می‌دهد. در برآورد اثر زمانی، هر سه مدل تقریباً رفتار مشابهی دارند و روند زمانی یکسانی را برآورد کرده‌اند. از روی هر دو اثر می‌توان گفت نرخ ابتلا از سال ۹۱ به ۹۲ به شدت کاهش یافته است، ولی بعد از آن دوباره یک رشد خطی افزایشی با شیبی نسبتاً تند رخ داده است. شکل ۷، نقشه پهنه‌بندی برآورد اثر متقابل فضایی-زمانی



شکل ۶. نمودار پسینی اثرات غیرساختاری (الف) و ساختاری (ب) زمانی برای داده‌های سرطان خون

(بر اساس میانگین پسینی) را برای دو مدل GC و پواسون نشان می‌دهد. با توجه به این‌که در مدل NB این اثر ناچیز برآورد شد، نتیجه متناظر برای این مدل گزارش نشده است. با مشاهده نقشه‌های پهنه‌بندی، اثر زمانی برآوردشده در اثر متقابل فضایی-زمانی نیز مشاهده می‌شود. یعنی در هر دو مدل، نرخ ابتلا از ۹۱ به ۹۲ برای نواحی پرخطر کمتر شده است ولی دوباره برای سال‌های ۹۳ و ۹۴ این نرخ در نواحی پرخطر رشد چشمگیری داشته است. این وضعیت به‌ویژه برای مدل پواسون مشهود است. افزون بر این، به‌طور کلی، برآوردهای مبتنی بر مدل پواسون نرخ بالاتری از ابتلا را نسبت به مدل GC نشان می‌دهند. با توجه به این‌که مدل پیشنهادی GC، بر مبنای معیارهای انتخاب مدل، نسبت به مدل پواسون برتر است، شهرستان‌های پرخطر بر اساس برآوردهای حاصل از مدل GC فهرست شده است. نقشه‌های پهنه‌بندی از سال ۹۱ تا ۹۴ نشان می‌دهند که شهر مشهد، اگرچه در سال ۹۳ وضعیت خوبی داشته است، ولی یکی از شهرهای پرخطر برای ابتلا به سرطان خون است. شهرستان سبزوار نیز از یک منطقه کم‌خطر در سال ۹۱ به یکی از مناطق پرخطر برای ابتلا به این بیماری در سال ۹۴ تبدیل شده است. همچنین، شهرستان‌های واقع در جنوب غربی استان را می‌توان به‌عنوان مناطق کم‌خطر برای ابتلا به شمار آورد. از طرف دیگر،

در مجموع، شهرستان‌های تربیت حیدریه و نیشابور، در کنار سایر شهرستان‌های ذکرشده، به‌عنوان مناطق پرخطر قابل فهرست هستند.



شکل ۷. نقشه پسینی اثر متقابل فضایی-زمانی برای مدل‌های GC (ردیف اول) و پواسون (ردیف دوم) در سال‌های ۱۳۹۱ (الف)، ۱۳۹۲ (ب)، ۱۳۹۳ (ج) و ۱۳۹۴ (د)

## بحث و نتیجه‌گیری

در این مقاله، یک مدل رگرسیونی فضایی-زمانی جدید بر اساس یک مدل پواسون منعطف، با نام گامای شمارشی، برای تحلیل پاسخ‌های شمارشی ارائه شد. ویژگی بارز مدل معرفی‌شده در تحلیل داده‌های شمارشی، توانایی آن برای مدل‌بندی انواع پراکنش موجود در داده‌ها شامل بیش‌پراکنش، پراکنش متعادل یا کم‌پراکنش است. مدل پیشنهادی در قالب یک مطالعه شبیه‌سازی و مثال کاربردی برای ارزیابی نرخ ابتلا به سرطان خون در استان خراسان رضوی، مورد بررسی قرار گرفت. نتایج به‌دست آمده نشان دادند که مدل پیشنهادی نسبت به مدل‌های متداول پواسون و دوجمله‌ای منفی از برتری برخوردار است و به‌ویژه وقتی مشاهدات دارای کم‌پراکنش هستند، کارایی آن کاملاً محسوس است.

## تقدیر و تشکر

نویسندگان مقاله از مرکز بهداشت استان خراسان رضوی برای در اختیار قرار دادن داده‌های سرطان خون، برای استفاده در این پژوهش، قدردانی می‌کنند. همچنین نویسندگان از داوران محترم مقاله، سردبیر و

ویراستار محترم مجله که نظرات آن‌ها باعث ارتقای سطح مقاله شد، قدردانی می‌کنند.

## مراجع

فلاح، ف.، نادى فر، م. و کاظمى، ر. (۱۳۹۲)، مدل رگرسیونى بیزی با متغیر پاسخ پواسون آمیخته متناهی دو متغیره، مجله علوم آماری، ۷، ۷۷-۱۰۲.

Baio, G. (2012), *Bayesian Methods in Health Economics*, CRC Chapman and Hall, London.

Bauer, C., Wakefield, J., Rue, H., Self, S., Feng, Z. and Wang, Y. (2015), Bayesian Penalized Spline Models for the Analysis of Spatio-Temporal Count Data, *Statistics in Medicine*, **35**, 1848-1865.

Berry, S., Carlin, B., Lee, J. and Muller, P. (2011), *Bayesian Adaptive Methods for Clinical Trials*, CRC Chapman and Hall, New York.

Besag, J., York, J. and Mollie, A. (1991), Bayesian Image Restoration, With Two Applications in Spatial Statistics, *Annals of the Institute of Statistical Mathematics*, **43**, 1-59.

Blangiardo, M. and Cameletti, M. (2015), *Spatial and Spatio-Temporal Bayesian Models with R-INLA*, JohnWiley & Sons: West Sussex, UK.

Blangiardo, M., Cameletti, M., Baio, G. and Rue, H. (2013), Spatial and Spatio-Temporal Models with R-INLA, *Spatial and Spatio-temporal Epidemiology*, **7**, 39-55.

Cameron A. C. and Trivedi P. K. (2013), *Regression Analysis of Count Data*, Second Edition, Cambridge University Press, New York.

Cox, D. R. (1962), *Renewal Theory*, Methuen, London.



Cressie, N. and Wikle, C. K. (2011), *Statistics for Spatio-Temporal Data*, Wiley, Hoboken, NJ.

Gelman, A., Hwang, J. and Vehtari, A. (2014), Understanding Predictive Information Criteria for Bayesian Models, *Statistics and Computing*, **24**, 997-1016.

Gelman, A., Lee, D. and Guo, J. (2015), Stan: A Probabilistic Programming Language for Bayesian Inference and Optimization, *Journal of Education and Behavioral Statistics*, **40** , 530–543.

Gómez, R. V. (2020), *Bayesian Inference with INLA*, CRC Press.

Greenland, S. (2006), Bayesian Perspectives for Epidemiological Research: I. Foundations and Basic Methods, *International Journal of Epidemiology*, **35**, 765-775.

Kharrat, T., Boshnakov, G. N., McHale, I. and Baker, R. (2019). Flexible Regression Models for Count Data Based on Renewal Processes, *Journal of Statistical Software*, **90** , 1-35.

Knorr-Held, L. (2000), Bayesian Modelling of Inseparable Space-Time Variation in Disease Risk, *Statistics in Medicine*, **19**, 2555-2567.

Klein, N. and Kneib, T. (2016), Scale Dependent Priors for Variance Parameters in Structured Additive Distributional Regression, *Bayesian Analysis*, **11**, 1071-1106.

Lunn, D., Spiegelhalter, D., Thomas, A. and Best, N. (2009). The BUGS Project: Evolution, Critique and Future Directions, *Statistics in Medicine*, **28** , 3049-3067.

- Lang, S., and Kneib, T. and Brezger, A. (2005), BayesX: Analyzing Bayesian Structural Additive Regression Models, *Journal of Statistical Software*, **14**.
- Li, Y., Brown, P., Rue, H., al Maini, M. and Fortin, P. (2012), Spatial Modelling of Lupus Incidence Over 40 Years with Changes in Census Areas, *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, **61**, 99-115.
- Luan, H., Quick, M. and Law, J. (2016), Analyzing Local Spatio-Temporal Patterns of Police Calls-for-Service Using Bayesian Integrated Nested Laplace Approximation, *International Journal of Geo-Information*, **5**, 162.
- McCullagh, P. and Nelder, J. (1989), *Generalized Linear Models, Second Edition*, Chapman and Hall, London.
- Nadifar, M., Baghishani, H., Fallah, A. and Rue, H. (2019), Statistical Modeling of Groundwater Quality Assessment in Iran using a Flexible Poisson Likelihood. <https://arxiv.org/abs/1908.02344>.
- Riebler, A., Held, L. and Rue, H. (2012), Estimation and Extrapolation of Time Trends in Registry Data - Borrowing Strength from Related Populations, *Annals of Applied Statistics*, **6**, 304-333.
- Rue, H., Martino, S. and Chopin, N. (2009), Approximate Bayesian Inference for Latent Gaussian Models by using Integrated Nested Laplace Approximations, *Journal of the Royal Statistical Society, Series B*, **71**, 319-392.
- Sellers, K. F. and Shmueli, G. (2010), A Flexible Regression Model for Count Data. *The Annals of Applied Statistics*, **4**, 943-961.

- Simpson, D., Rue, H., Riebler, A., Martins, T. G. and Sørbye, S. H. (2017), Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors, *Statistical Science*, **32**, 1-28.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Linde, A. (2002), Bayesian Measures of Model Complexity and Fit (With Discussion), *Journal of the Royal Statistical Society, Series B*, **64** , 583-639.
- Utazi, E. C., Afuecheta, E. O. and Nnanatu, C. C. (2018), A Bayesian Latent Process Spatio Temporal Regression Model for Areal Count Data, *Spatial and Spatio-Temporal Epidemiology*, **25**, 25 - 37.
- Watanabe, S. (2013), A Widely Applicable Bayesian Information Criterion, *Journal of Machine Learning Research*, **14**, 867–897.
- Winkelmann R. (1995), Duration Dependence and Dispersion in Count-Data Models, *Journal of Business and Economic Statistics*, **13**, 467 - 474.
- Zeviani, W. M., Ribeiro Jr, P. J., Bonat, W. H., Shimakura, S. E. and Muniz, J. A. (2014), The Gamma-Count Distribution in the Analysis of Experimental Underdispersed Data, *Journal of Applied Statistics*, **41**, 2616-2626.

## **Analysis of Space-Time Count Data Using the Flexible Gamma-Count Model**

M. Nadifar<sup>1</sup>, H. Baghishani<sup>1</sup>, A. Fallah<sup>2</sup>

<sup>1</sup>Faculty of Mathematical Sciences, Shahrood University of Technology, Iran.

<sup>2</sup>Faculty of Science, Imam Khomeini International University, Qazvin, Iran.

**Abstract:** Many of spatial-temporal data, particularly in medicine and disease mapping, are counts. Typically, these types of count data have extra variability that distracts the classical Poisson model's performance. Therefore, incorporating this variability into the modeling process, plays an essential role in improving the efficiency of spatial-temporal data analysis. For this purpose, in this paper, a new Bayesian spatial-temporal model, called gamma count, with enough flexibility in modeling dispersion is introduced. For implementing statistical inference in the proposed model, the integrated nested Laplace approximation method is applied. A simulation study was performed to evaluate the performance of the proposed model compared to the traditional models. In addition, the application of the model has been demonstrated in analyzing leukemia data in Khorasan Razavi province, Iran.

**Keywords:** Spatial-temporal interaction, Over-dispersion, Penalized complexity prior, Count data, Gamma count model.

**Mathematics Subject Classification (2010):** 62J12, 62F15, 62H11.