






## Bayesian Approach for Modelling of Spatio-Temporal Crime Data

Mohammadian Mosammam, A.<sup>1</sup> , Abbasi, E.<sup>1</sup> , Mateu, J.<sup>2</sup> ,

<sup>1</sup>Department of Statistics, University of Zanzan, Zanzan, Iran.

<sup>2</sup>Department of Statistics, Jaume I University, Spain.

**Corresponding author:** A. Mohammadian Mosammam, a.m.mosammam@znu.ac.ir

Received: 3/3/2022 Revised: 23/8/2022 Accepted and Published Online: 2/9/2022.

### Introduction

Studying crime data has become one of the essential topics in the world due to its connection with human security. Analyzing this type of data can effectively prevent future crimes and identify spatial patterns and factors that facilitate the commission of crimes to control crime-prone areas. Most of the time, crime data has a spatio-temporal structure that causes the formation of different spatio-temporal patterns. Therefore, spatio-temporal monitoring of crime data is essential in identifying factors that cause crime and preventing crime. An important issue in many cities is related to crime events, and the spatio-temporal Bayesian approach leads to identifying crime patterns and hotspots. In Bayesian analysis of spatio-temporal crime data, there is no closed form for posterior distribution because of its non-Gaussian distribution and the existence of latent variables. In this case, we face challenges such as high dimensional parameters, extensive simulation and time-consuming computation in applying MCMC methods.

### Material and Methods

In this paper, we apply INLA to analyze crime data in Colombia. To describe the above concepts, a three-stage hierarchical model is considered. The advantages of this method can be the estimation of criminal events at a specific time and location and exploring unusual patterns in places.

### Results and Discussion

The Bayesian analysis of crime data is usually performed as Bayesian infer-

ence of pure spatial or temporal patterns. However, such spatial or temporal Bayesian analyses are not suitable for crime data. In this article, in a case study, Bayesian hierarchical spatio-temporal analysis of crime data in Colombia was discussed using the INLA approach, which considers spatio-temporal dependence and makes the model more flexible in detecting unusual patterns. Exploratory data analysis is also discussed, detecting areas with unusual behaviour over time. Four different models were fitted to the data, and the best model that includes spatio-temporal interaction was selected using the DIC criterion. The research results identify the most important centre of crime in the Kennedy area of Bogotá, as well as the highest crime rate in the time frame. Then, hierarchical spatio-temporal Bayesian analysis of these data was done with the *INLA* approach.

### Conclusion

The advantage of using this Bayesian approach is that it includes the effects of spatio-temporal correlation in the model and makes the model flexible in detecting areas with abnormal behaviour over time and in different places. For this purpose, four different models, including side effects and spatio-temporal combination, were fitted to the crime data. The best model, including the spatio-temporal interaction effect, was proposed using the deviance information criterion. The comprehensive and scientific comparison of the two Bayesian methods INLA and the MCMC algorithm in terms of accuracy, speed and even accessibility and convenient use for researchers requires independent scientific and practical research because, for example, the various methods of sampling in the MCMC algorithms and sometimes its different methods in INLA make it difficult to compare accuracy. How to use parallel calculations in the application of these two methods is also effective in comparing the speed, and simply comparing the outputs cannot express the advantage of one method over the other.

**Keywords:** Integrated nested Laplace approximation, Bayesian statistics, Spatial-temporal statistics.

**Mathematics Subject Classification (2010):** 62MC, 33C45.



©The Author(s). The Publisher is Iranian Statistical Society.

This is an open access article distributed under the terms and conditions of [\(CC BY-NC 4.0\)](https://creativecommons.org/licenses/by-nc/4.0/)



## کاربست رهیافت بیزی در مدل‌سازی فضایی-زمانی داده‌های جرم

علی محمدیان مصمم<sup>۱</sup>، الناز عباسی<sup>۱</sup>، متیو خورخه<sup>۲</sup>

<sup>۱</sup> گروه آمار، دانشگاه زنجان، زنجان، ایران

<sup>۲</sup> گروه آمار، دانشگاه جومه I، اسپانیا

نویسنده مسئول: علی محمدیان مصمم، a.m.mosammam@znu.ac.ir

تاریخ دریافت: ۱۴۰۰/۱۲/۱۲ تاریخ بازنگری: ۱۴۰۱/۶/۱ تاریخ پذیرش و انتشار: ۱۴۰۱/۶/۱۱

**چکیده:** در تحلیل بیزی داده‌های فضایی-زمانی جرم و جنایت معمولاً به دلیل ناگوسی بودن توزیع متغیر پاسخ و وجود تعداد زیادی متغیر پنهان در مدل تحت بررسی شکل بسته‌ای برای توزیع پسینی وجود ندارد. در این شرایط در استفاده از روش‌های مونت‌کارلوی زنجیر مارکوفی با چالش‌هایی نظیر وجود پارامترهای متعدد در ساختار سلسله‌مراتبی، محاسبات سنگین و زمان‌بر، انجام شبیه‌سازی گسترده، به‌ویژه زمانی که بعد میدان تصادفی بزرگ است و سرانجام عدم همگرایی توزیع پسینی مواجه می‌شویم. برای حل این مشکلات روش تقریب لاپلاس آشیانی جمع‌بسته پیشنهاد شده است. مزیت این روش این است که برآوردهایی از منظر وقوع جرم و جنایت در مکان و زمان معین ارائه کرده و همچنین نواحی با رفتار غیرمعمول را تشخیص می‌دهد. در این مقاله با استفاده همزمان از GIS و روش تقریب لاپلاس آشیانی جمع‌بسته در یک مطالعه موردی به تحلیل داده‌های جرم و جنایت بخشی از کشور کلمبیا می‌پردازیم. واژه‌های کلیدی: تقریب لاپلاس آشیانی جمع‌بسته، تحلیل سلسله‌مراتبی بیزی، آمار فضایی-زمانی. کد موضوع بندی ریاضی (۲۰۱۰): 33C45, 62M30.

## ۱ مقدمه

مطالعه داده‌های جرم (گوتزالز و همکاران، ۲۰۱۶) به دلیل ارتباط با امنیت انسان‌ها به یکی از موضوعات مهم دنیا تبدیل شده است. بررسی این نوع داده‌ها می‌تواند در پیشگیری از جرم‌های آینده و شناسایی الگوهای مکانی



©نویسندگان). ناشر انجمن آمار ایران است.

این مقاله با دسترسی آزاد تحت شرایط و ضوابط (CC BY-NC 4.0) توزیع شده است.

و عوامل تسهیل‌کننده ارتکاب جرایم به منظور کنترل محدوده‌های جرم‌خیز موثر باشد. اغلب اوقات داده‌های جرم دارای ساختار فضایی-زمانی هستند که موجب شکل‌گیری الگوهای فضایی-زمانی متفاوت می‌شود. بنابراین نظارت فضایی-زمانی بر داده‌های جرم و جنایت نقش مهمی در شناسایی عوامل بروز جرم و پیشگیری از وقوع جرم دارد (گربر و فورر، ۲۰۱۵). تحلیل بیزی داده‌های جرم و جنایت معمولاً به صورت استنباط بیزی الگوهای فضایی محض یا الگوهای زمانی محض انجام می‌گیرد. با این حال در مورد داده‌های جرم و جنایت چنین تحلیل‌های بیزی صرفاً فضایی یا زمانی به دلیل اینکه اثر متقابل فضا و زمان را در نظر نمی‌گیرند مدل‌های مناسبی نیستند (لیو و همکاران، ۲۰۲۱). در این مقاله سعی شده است که با استفاده از پهنه‌بندی GIS و مدل‌های بیزی به تحلیل و تفسیر این اثرات متقابل فضایی-زمانی نیز پرداخته شود. در آمار فضایی-زمانی به طور معمول یک میدان تصادفی به عنوان مدل آماری داده‌های فضایی-زمانی در نظر گرفته می‌شود (کوسی، ۱۹۹۳). میدان تصادفی فضایی-زمانی مجموعه‌ای از متغیرهای تصادفی مانند  $Z(\cdot, \cdot) = \{Z(s, t); (s, t) \in D \times T\}$  است، که در آن  $s$  موقعیت فضایی در مجموعه شاخص  $D \subseteq R^d$ ،  $d \geq 1$  و  $t$  لحظه زمانی در مجموعه  $T \subseteq R$  است (محمدزاده، ۱۳۹۸).

روش مرسوم برای تحلیل بیزی داده‌های فضایی-زمانی، یعنی استفاده از روش‌های مونت‌کارلوی زنجیر مارکوفی<sup>۱</sup> (رابرت و همکاران، ۱۹۹۹) مشکلات فراوانی به دلایلی نظیر وابستگی شدید بین مشاهدات و بعد زیاد ابرپارامترها در محاسبات دارد. لذا هدف این مقاله استفاده از روش تقریب لاپلاس آسیانی جمع‌بسته<sup>۲</sup> برای تحلیل چنین داده‌هایی است. روش INLA اولین بار توسط رو و همکاران (۲۰۰۹) ارائه شد و سپس توسط مارتین و همکاران (۲۰۱۳) و رو و همکاران (۲۰۱۷) توسعه پیدا کرد. قلی‌زاده و همکاران (۱۳۹۲) تحلیل فضایی رگرسیون جمعی ساختاری و مدل‌بندی داده‌های جرم شهر تهران را با تقریب لاپلاس آسیانی جمع‌بسته انجام دادند. اقبال و باغشینی (۱۳۹۹) نیز به تحلیل بیزی داده‌های شمارشی در جوامع متناهی با INLA پرداخته‌اند. تحلیل بیزی چندمتغیره فضایی-زمانی روش INLA نیز توسط ویسنت و همکاران (۲۰۲۰) ارائه گردیده است. استاریا و همکاران (۲۰۲۱) تعداد تصادفات جاده‌ای را با استفاده از روش INLA مورد تجزیه و تحلیل قرار داده‌اند. به منظور متناسب بودن حجم مقاله حاضر در این مقاله به مقایسه جامع روش INLA و MCMC پرداخته نمی‌شود. مقایسه روش INLA و MCMC نشان می‌دهد که روش INLA نه تنها زمان محاسبات را کاهش می‌دهد بلکه برآورد بهتری از پارامترها ارائه می‌کند (رو و همکاران، ۲۰۱۳). هرچند مقایسه جامع و علمی دو روش بیزی INLA و الگوریتم MCMC از نظر دقت، سرعت و حتی قابل دسترسی و کاربرد راحت برای پژوهشگران نیازمند یک پژوهش علمی و عملی مستقل است زیرا به عنوان مثال روش‌های متنوع نمونه‌گیری در الگوریتم‌های MCMC و گاه‌ها روش‌های متفاوت آن در INLA مقایسه دقت را مشکل می‌سازد. چگونگی استفاده از محاسبات موازی در استفاده از این دو روش و زبان‌های برنامه‌نویسی بعضاً متفاوت نیز در مقایسه سرعت تاثیرگذار است و صرفاً مقایسه خروجی‌ها نمی‌تواند مزیت یک روش به دیگری را بیان نماید. ادامه ساختار مقاله به صورت زیر است. در بخش ۲ جزئیات روش INLA معرفی می‌شود. در بخش ۳ به تحلیل اکتشافی داده‌های جرم و جنایت می‌پردازیم. در بخش ۴ داده‌های تماس به پلیس برای گزارش جرم و

<sup>1</sup>Markov chain Monte Carlo (MCMC)

<sup>2</sup>Integrated nested Laplace approximation (INLA)

جنایت به روش بیزی مدل بندی می شوند و با استفاده از معیار بیزی اطلاع انحراف<sup>۱</sup> بهترین مدل برازشی انتخاب می شود.

## ۲ تقریب لاپلاس آشیانی جمع بسته

اغلب مدل های بیزی در رگرسیون، آمار فضایی و آمار فضایی-زمانی دارای ساختار گاوسی پنهان هستند. علاوه بر این در کاربست INLA نیازمند یک میدان تصادفی مارکوفی گاوسی<sup>۲</sup> (رو و همکاران، ۲۰۰۵) هستیم. GMRF یک میدان تصادفی گاوسی با خاصیت استقلال شرطی است یعنی  $x_i$  و  $x_j$  به شرط سایر مؤلفه ها مستقل هستند اگر و تنها اگر  $i$  و  $j$  زمین مؤلفه ماتریس دقت  $Q = \Sigma^{-1}$  (رو و همکاران، ۲۰۰۵) که همان معکوس ماتریس کواریانس است صفر باشند. برای توصیف این مفاهیم مدل سلسله مراتبی سه مرحله ای را در نظر بگیرید. فرض کنید در مرحله اول، متغیر پاسخ  $y = (y_1, \dots, y_n)'$  به طور شرطی مستقل و دارای توزیع نمایی خاصی به صورت  $y|x, \theta_1 \sim \prod_{i=1}^n P(y_i|x_i; \theta_1)$  باشد، که در آن  $\theta_1$  پارامتر توزیع است. در مرحله دوم،  $x$  را یک میدان تصادفی گاوسی پنهان با تابع چگالی  $P(x|\theta_2) \propto |Q_{\theta_2}|_+^{1/2} \exp(-\frac{1}{2}x'Q_{\theta_2}x)$  در نظر بگیرید که در آن ماتریس دقت  $Q_{\theta_2}$  یک ماتریس معین مثبت با پارامتر  $\theta_2$  و  $|Q_{\theta_2}|_+$  حاصل ضرب مقادیر ویژه غیر صفر آن است. در مرحله نهایی، فرض کنید که ابرپارامتر  $\theta = (\theta_1, \theta_2)$  دارای توزیع پیشینی  $\pi(\theta)$  است. در نتیجه توزیع پسینی پارامترهای مدل سلسله مراتبی به صورت

$$\begin{aligned} \pi(\eta, \theta|y) &\propto \pi(\theta)\pi(\eta|\theta_2)\prod_{i=1}^n \pi(y_i|\eta_i; \theta_1) \\ &\propto \pi(\theta)|Q_{\theta_2}|_+^{1/2} \exp\{-\frac{1}{2}\eta'Q_{\theta_2}\eta + \sum_{i=1}^n \log P(y_i|\eta_i; \theta_1)\} \end{aligned}$$

خواهد بود. توزیع های پسینی حاشیه ای برای متغیرهای پنهان و ابرپارامترها به صورت

$$\begin{aligned} \pi(\eta_i|y) &= \int \pi(\eta_i|\theta, y)\pi(\theta|y)d\theta \\ \pi(\theta_i|y) &= \int \pi(\theta|y)d\theta_{-i} \end{aligned}$$

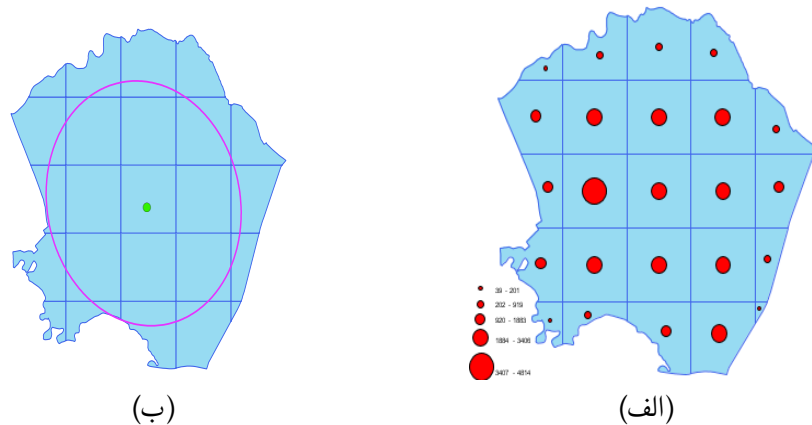
به دست می آیند. محاسبه چنین توزیع های پسینی امر ساده ای نیست. رو و همکاران (۲۰۱۷) یک روش تقریب لاپلاس در INLA برای محاسبه توزیع های حاشیه ای پسینی هر یک از پارامترهای نامعلوم ارائه کردند.

<sup>1</sup>Deviance Information Criterion (DIC)

<sup>2</sup>Gaussian Markov Random Fields (GMRF)

### ۳ تحلیل اکتشافی داده‌های جرم و جنایت

در این بخش به تحلیل اکتشافی داده‌های جرم و جنایت منطقه کندی پرداخته می‌شود. منطقه کندی در جنوب غربی بوگوتا پایتخت کشور کلمبیا واقع شده است. این منطقه یکی از ناامن‌ترین مناطق در بوگوتا است و به دلیل ناامنی و نرخ بالای خشونت بیشترین توجه مسئولین را به خود اختصاص داده است. در کل ۴۲۹۳۰ گزارش جرم و جنایت در طول ۲۴ منطقه و ۱۴۴ واحد فضا-زمان ثبت شده است. این داده‌ها برای سال‌های ۲۰۱۱ تا ۲۰۱۷ در ۶ دوره سالانه طبقه‌بندی شده‌اند. به منظور تحلیل فضایی جرائم ابتدا توزیع نقطه‌ای جرائم در محدوده مورد مطالعه در شکل ۱ نمایش داده می‌شود، سپس با استفاده از مدل‌های گرافیکی و آماری الگوی فضایی این جرائم در منطقه مورد مطالعه را استخراج کرده و در نتیجه کانون‌های جرم‌خیز را شناسایی می‌کنیم. با استفاده از مدل‌های گرافیکی شامل آزمون میانگین مرکزی و توزیع جهت‌دار (بیضی انحراف استاندارد)، توزیع فضایی و مرکز ثقل جرائم را ارزیابی می‌کنیم. این تحلیل، مرکز ثقل مجموعه‌ای از عوارض را شناسایی می‌کند، بیضی انحراف استاندارد این امکان را فراهم می‌کند تا جهت توزیع عوارض را به طور آماری و دقیق شناسایی کنیم. در شکل ۱ میانگین مرکزی و بیضی انحراف استاندارد

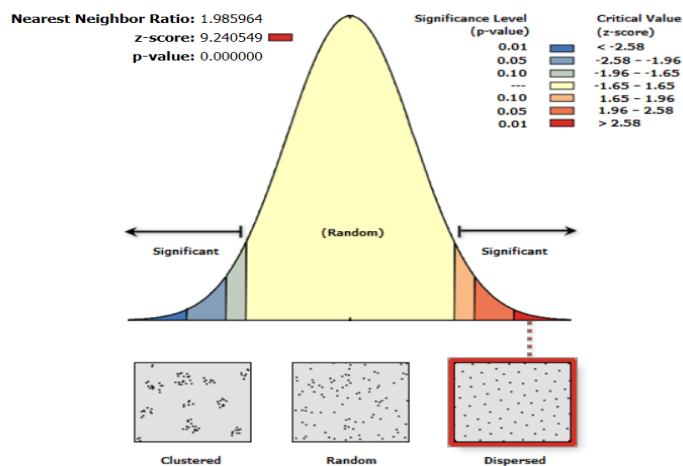


شکل ۱. الف- توزیع نقطه‌ای جرائم منطقه کندی در شهر بوگوتا، ب- میانگین مرکزی و بیضی انحراف استاندارد جرائم در محدوده مورد مطالعه کندی در شهر بوگوتا

جرائم در محدوده مورد مطالعه نشان داده شده است. با توجه به نقشه به‌دست آمده الگوی جهت عوارض و مرکز ثقل جرائم بدین شرح است: مرکز ثقل جرائم تقریباً بر مرکز محدوده مورد مطالعه منطبق است. با توجه به این امر احتمال وقوع جرم در محله‌های نزدیک به مرکز محدوده مورد مطالعه بیشتر است. هم‌چنین بیضی انحراف استاندارد در امتداد جنوب شرقی و شمال غربی کشیده شده است، علت این کشیدگی جرائم ارتكابی در محله‌های جنوب شرقی و شمال غربی است و احتمال وقوع جرائم در جهت کشیدگی بیضی بیشتر از سایر نقاط است.

### ۳.۱ خوشه‌بندی جرائم

کشف روندهای موجود در داده‌های فضایی و شناخت الگوها در آمار فضایی از اهمیت ویژه‌ای برخوردار است، زیرا در تشخیص چگونگی توزیع داده‌ها در فضا و همچنین شناخت الگوی پیروی شده توسط توزیع این داده‌ها به ما کمک می‌کند. از جمله مهم‌ترین آزمون‌ها در شناسایی الگوهای فضایی آزمون نزدیک‌ترین همسایگی است. با انجام این آزمون می‌توان پی برد که آیا توزیع داده‌ها خوشه‌ای است یا پراکنده، به طوری که در این آزمون اگر نسبت میانگین نزدیک‌ترین همسایگی کمتر از یک باشد، داده‌های مورد مطالعه دارای الگوی خوشه‌ای و اگر شاخص محاسبه شده بزرگتر از یک باشد، داده‌ها دارای الگوی پراکنده هستند (عسگری، ۱۳۹۰). نتیجه آزمون به صورت گرافیکی و عددی به ترتیب در شکل ۲ و جدول ۱ آورده شده است. نسبت نزدیک‌ترین همسایه به دست آمده نشان می‌دهد که



شکل ۲. آزمون نزدیک‌ترین همسایگی داده‌های جرم و جنایت منطقه کندی در شهر بوگوتا

جدول ۱. نتایج عددی آزمون نزدیک‌ترین همسایگی داده‌های جرم و جنایت منطقه کندی در شهر بوگوتا

۱۱۷۶۶۰۲۲	میانگین فاصله مشاهده شده
۵۹۲۴۵۸۸	میانگین فاصله مورد انتظار
۱٫۹۸۵۹	نسبت نزدیک‌ترین همسایه
۹٫۲۴۰۵	امتیاز استاندارد
< ۰٫۰۰۰۰۱	p - value

الگوی توزیع کل داده‌های جرم در فضای مورد مطالعه به طور یکسان یا منظم است، در صورتی که نتایج تحلیل بیزی فضایی-زمانی، بیانگر وجود الگوهای با کانون‌های جرم‌خیز در داده‌ها در مکان‌ها و زمان‌های خاص است.

### ۳.۲ شناسایی کانون‌های جرم‌خیز

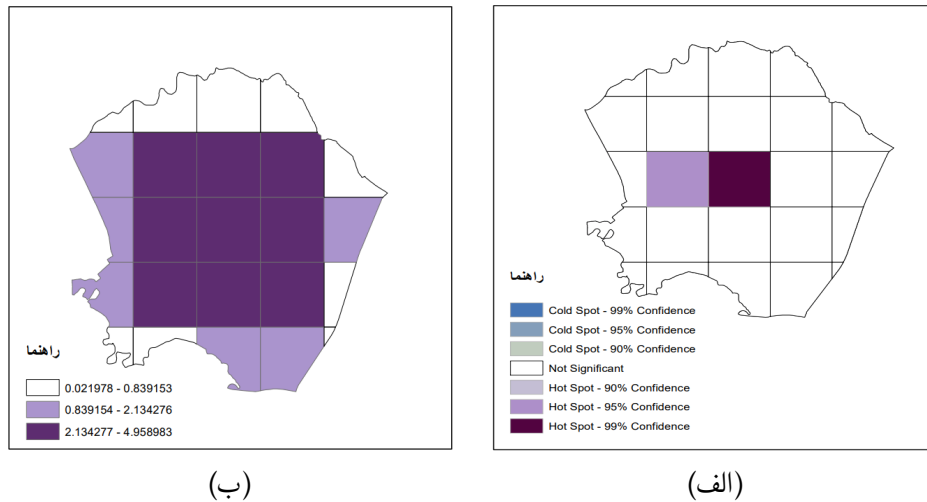
طبق بررسی‌های انجام شده ساختار برخی از فضاها طوری است که مانع ارتکاب جرم می‌شود و برخی از فضاها ساختار کم خطری برای ارتکاب جرم دارند. بنابراین در سطح شهرها حوزه‌هایی شکل می‌گیرد که دارای فرصت‌های مجرمانه بیشتر و راحت‌تر و به دنبال آن تعداد جرائم بیشتر هستند که کانون‌های جرم‌خیز نامیده می‌شوند (گرب و فور، ۲۰۱۵). در این مکان‌ها تعداد جرم بالاست و به احتمال زیاد دوباره جرم تکرار خواهد شد. بنابراین شناسایی و تحلیل کانون‌های جرم‌خیز از جمله مفاهیم معتبر در زمینه پیشگیری از وقوع جرم و مقابله با آن محسوب می‌شود. برای شناسایی خوشه‌های عوارض با مقادیر زیاد و خوشه‌های عوارض با مقادیر کم از شاخص  $Getis - OrdGi$  (گتیس و اورد، ۲۰۱۰) استفاده می‌شود. این شاخص به برنامه‌ریزان و تحلیلگران کمک می‌کند تا مکان‌هایی را که تعداد زیادی جرم در آن‌ها رخ می‌دهد و به اصطلاح مکان‌های جرم‌خیز نامیده می‌شوند شناسایی کنند تا با تخصیص درست منابع، سریعتر و موثرتر نسبت به پیشگیری از جرم و کشف آن اقدام کنند. نقشه کانون‌های جرم‌خیز برای دوره زمانی ششم در شکل ۳ آورده شده است. این دوره بیشترین اثرات اصلی زمان را دارد و با توجه به نقشه کانون‌های جرم‌خیز محدوده‌های تیره محدوده‌هایی هستند که ۹۹ درصد احتمال دارد در آن مناطق جرم به وقوع بپیوندد. محدوده‌های روشن مناطقی هستند که با احتمال ۹۵ درصد در آن‌ها تعداد زیادی جرم متمرکز شده است. هدف از شناسایی این کانون‌ها تشخیص عوامل مکانی موثر و خصوصیات فرهنگی و اجتماعی ساکنان و استفاده‌کنندگان این محدوده‌ها است که بتوان با استفاده از امکانات و برنامه‌ریزی درست این مناطق را به مناطق امن‌تر در برابر جرم تبدیل کرد. شکل ۳ اثرات اصلی فضایی در دوره ششم و الگوی فضایی عمومی تماس برای خدمات را نشان می‌دهد. مناطقی با اثرات فضایی زیاد، به‌طور مداوم تعداد زیادی تماس برای خدمات دارند و مناطقی هستند که در آنجا باید منابع پلیس بیشتر به‌کار گرفته شود.

### ۴ مدل‌سازی تعداد تماس‌ها برای خدمات پلیس

اخیراً پژوهش‌های زیادی برای تحلیل بیزی داده‌های جرم و جنایت به‌ویژه درمورد تعداد تماس‌های تخلفات به پلیس ارائه شده است (لوان و همکاران، ۲۰۱۶). از ویژگی‌های مهم چنین داده‌هایی داشتن ساختار GMRF بوده که ساختار ماتریسی تنک دارند (رو و همکاران، ۲۰۰۹). در این بخش برای مدل‌سازی فضایی-زمانی بیزی چنین داده‌هایی از مدل سلسله‌مراتبی سه مرحله‌ای و رهیافت INLA استفاده می‌شود. در مرحله اول چون تعداد جرم و جنایت کم است، لذا تعداد جرم‌ها در منطقه  $i$ ،  $24, 1,000, i = 1, \dots$  و دوره زمانی از سال ۲۰۱۲ تا ۲۰۱۷ که به ترتیب دوره اول تا ششم در نظر گرفته شده اند  $t = 1, 6, \dots, 1,000, t$ ، دارای توزیع پواسن به صورت  $y_{it} | \mu_{it} \sim \text{Poisson}(\mu_{it})$   $i = 1, \dots, 24, t = 1, \dots, 6$  فرض شده است. میانگین توزیع تعداد تماس‌ها را به صورت

$$\log(\mu_{it}) = \log(E_{it}) + \alpha + u_i + s_i + \gamma_t + \phi_t + \psi_{it}$$





شکل ۳. الف- مناطق جرم‌خیز در دوره زمانی ششم، ب- اثر فضایی در دوره زمانی ششم

در نظر بگیرید که در آن تعداد مورد انتظار تماس‌ها بوده و ریسک نسبی به مولفه‌های ریسک کلی یا اثر ثابت  $(\alpha)$ ، اثرات تصادفی فضایی  $(u_i + s_i)$ ، اثرات تصادفی زمانی  $(\gamma_t + \phi_t)$  و اثر تصادفی فضایی- زمانی  $(\psi_{it})$  تجزیه شده است. در مرحله دوم فرض کنید

$$u_i \sim \text{Normal}(0, \sigma_u^2), \quad s_i \sim \text{ICAR}(W, \sigma_s^2) \quad i = 1, \dots, 24,$$

$$\gamma_t \sim \text{Normal}(0, \sigma_\gamma^2), \quad \phi_t \sim \text{ICAR}(P, \sigma_\phi^2) \quad t = 1, \dots, 6$$

که در آن  $\sigma_u^2, \sigma_s^2, \sigma_\gamma^2, \sigma_\phi^2$  ابرپارامترها هستند. همان‌گونه که ملاحظه می‌گردد اثرات تصادفی فضایی و اثرات تصادفی زمانی به دو مؤلفه بدون ساختار و ساختاریافته تجزیه شده‌اند که مولفه‌های دارای ساختار فضایی از توزیع پیشین اتورگرسیو شرطی ذاتی ICAR که حالت خاصی از مدل اتورگرسیو شرطی است (بسپج، ۱۹۹۱)

تبعیت می‌کند به عبارتی دیگر به صورت  $x_i | x_{-i} \sim \text{Normal}(\bar{x}(i), \frac{\sigma_s^2(i)}{n(i)})$  بیان می‌شود، که در آن  $\bar{x}(i) =$

$$W = (w(i, j))_{i,j=1}^n \text{ واریانس و میانگین، } \sum_{j:i \sim j} \frac{w(i, j)x(j)}{n(i)}$$

$$\begin{cases} w_{i,j} = 1 & i \sim j \\ w_{i,j} = 0 & o.w. \end{cases}$$

است،  $n(i) = \sum_{j:i \sim j} w(i, j) + w(i, i) = 0$  تعداد همسایه‌های منطقه  $i$  است.

۴۴۴ ..... کاربرد رهیافت بیزی در مدل‌سازی فضایی-زمانی داده‌های جرم

برای مرحله سوم تحلیل بیزی سلسله‌مراتبی توزیع پیشینی تمام ابرپارامترها را  $(\gamma, \delta, \sigma^2)$  در نظر می‌گیریم. به روش INLA چهار مدل متفاوت برای تحلیل بیزی داده‌ها در نظر گرفته شده است:

- ۱-  $\psi_{it}$ ها مستقل و هم توزیع هستند.
- ۲-  $\psi_{it}$ ها دارای همبستگی فضایی و ناهمبستگی زمانی هستند.
- ۳-  $\psi_{it}$ ها دارای ناهمبستگی فضایی و همبستگی زمانی هستند.
- ۴-  $\psi_{it}$ ها دارای همبستگی فضایی و زمانی هستند.

#### ۴.۱ انتخاب مدل

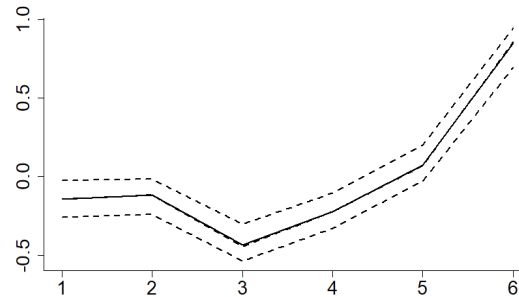
در این مقاله برای تشکیل پایگاه داده‌ای و برای مدل‌سازی با روش INLA از نرم‌افزار  $R$  با استفاده از تابع  $INLA$  در بسته آماری  $INLA$  که از طریق سایت <http://www.r-inla.org/> در دسترس است برای ۴ مدل فوق استفاده شده است. نقشه‌ها با نرم‌افزارهای سیستم اطلاعات جغرافیایی (GIS) تهیه شده‌اند. برای انتخاب مدل از معیار  $DIC$  استفاده می‌شود.  $DIC$  معیار بیزی برای نیکویی برازش و تعیین پیچیدگی مدل است. **اشپیکل و همکاران (۲۰۰۲)** پیشنهاد دادند که امید ریاضی پسین آماره انحراف  $\bar{D}$  و تعداد پارامترهای مؤثر در مدل  $pD$  اطلاعات مدل را خلاصه می‌کنند. بنابراین معیار اطلاع انحراف  $DIC$  به صورت  $DIC = \bar{D} + pD$  است. مدلی که کمترین  $DIC$  را نسبت به بقیه مدل‌ها داشته باشد مدل بهتری است. مقادیر  $DIC$  و تعداد پارامترهای مؤثر برای ۴ مدل فضایی-زمانی و تعامل فضا-زمان در جدول ۲ آورده شده است. طبق جدول ۲ مدل چهارم با کمترین  $DIC$  بهترین مدل انتخاب شده است. نمودار ۴ نشان‌دهنده اثرات تصادفی زمانی است. همانطور که در

جدول ۲. مقایسه کارایی مدل‌ها براساس معیار اطلاع انحراف  $DIC$

تعامل فضا-زمان	پارامترهای تعامل	اطلاع انحراف	پارامتر مؤثر
مدل اول	$\gamma_t$ و $u_i$	۱۳۱۰۲۵	۱۲۵۸۲
مدل دوم	$\phi_t$ و $u_i$	۱۲۹۴۳۳	۱۱۳۸۲
مدل سوم	$\gamma_t$ و $s_i$	۱۳۱۵۳۳	۱۲۴۸۲
مدل چهارم	$\phi_t$ و $s_i$	۱۲۹۳۸۱	۱۱۶۵۹

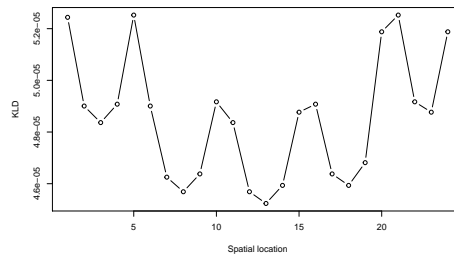
نمودار ۴ مشاهده می‌شود بیشترین تأثیرات اصلی زمانی در دوره زمانی ششم یعنی سال ۲۰۱۷ است. بدین معنی که بیشترین تعداد جرم در این دوره زمانی رخ داده است. بنابراین تخصیص منابع انسانی پلیس، در این دوره زمانی بیشتر از سایر دوره‌ها مورد نیاز بوده است. معیار واگرایی کولبک لیبر<sup>۱</sup> روشی برای مشخص کردن میزان تفاوت بین چگالی پسینی حاشیه‌ای و تقریب لاپلاس گاوسی است، هرچه مقدار شاخص KLD کوچکتر باشد میزان مطابقت بین دو توزیع بیشتر است (کولبک، ۱۹۹۷). نمودار ۵ نشان‌دهنده معیار واگرایی کولبک لیبر برای دوره‌های زمانی و موقعیت‌های فضایی است. همانطور که ملاحظه می‌شود دوره زمانی ششم بیشترین انحراف را دارد بنابراین کمترین

<sup>1</sup> Kullback – Leibler Divergence (KLD)

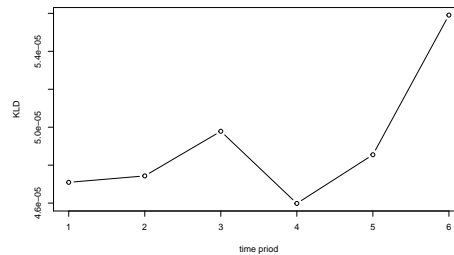


شکل ۴. اثرات تصادفی زمانی وقوع جرم در منطقه کندی شهر بوگوتا با فاصله باورمندی ۹۵ درصد

شبهات مربوط به این دوره است و بیشترین شبهات مربوط به دوره زمانی چهارم است. همچنین موقعیت فضایی پنجم در جنوب شرقی منطقه مورد مطالعه و موقعیت بیست و یکم در شمال شرقی منطقه مورد مطالعه بیشترین انحراف و کمترین شبهات را دارند و موقعیت دوازدهم در نزدیکی مرکز منطقه مورد مطالعه دارای بیشترین شبهات است.



(ب)



(الف)

شکل ۵. معیار واگرایی کولیک-لیبر برای الف- دوره‌های زمانی، ب- موقعیت‌های فضایی

## بحث و نتیجه‌گیری

یکی از مشکلات اصلی کشورها به‌ویژه در کلان شهرها پیش‌بینی و پیش‌گیری از وقوع ناهنجاری‌ها به ویژه جرم و جنایت است. شناسایی مکان‌های جرم‌خیز و همچنین زمان‌های با فراوانی بیشتر وقوع جرم بینش عمیقی به تامین‌کنندگان امنیت این شهرها در ارائه خدمات پلیس فراهم می‌سازد. در این مقاله، در یک مطالعه موردی ابتدا به تحلیل اکتشافی داده‌های جرم و جنایت کشور کلمبیا به منظور شناسایی نواحی با رفتار غیرمعمول و ناهنجار پرداخته

شده است. نتایج تحقیق مهم‌ترین کانون جرائم در منطقه کندی شهر بوگوتا و همچنین بیشترین میزان وقوع جرائم در محدوده زمانی را شناسایی می‌کند. سپس به تحلیل بیری فضایی-زمانی سلسله‌مراتبی این داده‌ها با رهیافت INLA پرداخته شد. مزیت کاربست این رهیافت بیزی این است که اثرات همبستگی فضایی-زمانی را در مدل لحاظ کرده و موجب انعطاف‌پذیری مدل برای تشخیص نواحی با رفتار ناهنجار در طول زمان و مکان‌های مختلف می‌گردد. برای این منظور چهار مدل مختلف شامل اثرات جانبی و توام فضایی-زمانی به داده‌های جرم و جنایت برازش داده و بهترین مدل که شامل اثر متقابل فضایی-زمانی است با استفاده از معیار اطلاع انحراف پیشنهاد گردید.

## تقدیر و تشکر

نویسندگان مقاله کمال قدردانی و تشکر را از پیشنهادات ارزنده داوران، سردبیر و ویراستار محترم مجله که باعث ارائه بهتر و افزایش سطح کیفی مقاله شده است، دارند.

## مراجع

- اقبال، ن.، باغیشنی، ح. (۱۳۹۹)، تحلیل بیزی داده‌های شمارشی فضایی در جوامع متناهی با رهیافت معادلات دیفرانسیل جزئی تصادفی، مجله علوم آماری، ۱۴(۲)، ۳۰۷-۳۳۴.
- عسگری، ع. (۱۳۹۰)، تحلیل‌های آمار فضایی با ArcGIS، چاپ اول، انتشارات سازمان فناوری اطلاعات و ارتباطات شهرداری تهران، تهران.
- قلی‌زاده، ک.، محمدزاده، م. و قیومی، ز. (۱۳۹۲)، تحلیل فضایی رگرسیون جمعی ساختاری و مدل‌بندی داده‌های جرم شهر تهران با تقریب لاپلاس آشیانی جمع بسته، مجله علوم آماری، ۷(۱)، ۱۰۳-۱۲۴.
- محمدزاده، م. (۱۳۹۸)، آمار فضایی و کاربردهای آن، چاپ سوم، مرکز نشر آثار علمی دانشگاه تربیت مدرس، تهران.
- Besag, J., York, J., and Mollié, A. (1991), Bayesian Image Restoration, with Two Applications in Spatial Statistics, *Annals of the Institute of Statistical Mathematics*, **43**(1), 1-20.
- Cressie, N. (1993), *Statistics for Spatial Data*, John Wiley, New York.
- Gerber, F., and Furrer, R. (2015), Pitfalls in the Implementation of Bayesian Hierarchical Modeling of Areal Count Data, An Illustration using BYM and Leroux Models, *Journal of Statistical Software*, **63**, 1-32.

Getis, A., and Ord, J. K. (2010), *The Analysis of Spatial Association by use of Distance Statistics*, In Perspectives on Spatial Data Analysis (pp. 127-145), Springer, Berlin, Heidelberg.

González, J. A., Rodríguez-Cortés, F. J., Cronie, O., and Mateu, J. (2016), Spatio-Temporal Point Process Statistics: A Review, *Spatial Statistics*, **18**, 505-544.

Kullback, S. (1997). *Information Theory and Statistics*, Courier Corporation.

Lindgren, F., Rue, H., and Lindström, J. (2011), An Explicit Link Between Gaussian Fields and Gaussian Markov Random Fields: the Stochastic Partial Differential Equation Approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **73(4)**, 423-498.

Liu, D., Song, W., Xiu, C., and Xu, J. (2021), Understanding the Spatio-Temporal Pattern of Crimes in Changchun, China: A Bayesian Modeling Approach, *Sustainability*, **13(19)**, 10500.

Luan, H., Quick, M., and Law, J. (2016), Analyzing Local Spatio-Temporal Patterns of Police calls-for-service using Bayesian Integrated Nested Laplace Approximation, *ISPRS International Journal of Geo-Information*, **5(9)**, 162.

Martins, T. G., Simpson, D., Lindgren, F., and Rue, H. (2013), Bayesian Computing with INLA: new features, *Computational Statistics and Data Analysis*, **67**, 68-83.

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 (Statistical Methodology)*, **71(2)**, 319-392.

Rue, H., Riebler, A., Sørbye, S. H., Illian, J. B., Simpson, D. P., and Lindgren, F. K. (2016), Bayesian Computing with INLA: A Review, *arXiv preprint arXiv:1604.00860*.

Rue, H., and Held, L. (2005), *Gaussian Markov Random Fields: Theory and Applications*. Chapman and Hall/CRC.

مراجع ..... ۴۴۸

Blangiardo, M., Cameletti, M., Baio, G., and Rue, H. (2013), Spatial and Spatio-Temporal Models with R-INLA, *Spatial and Spatio-Temporal Epidemiology*, **4**, 33-49.

Robert, C. P., Casella, G., and Casella, G. (1999), *Monte Carlo Statistical Methods (Vol. 2)*, New York: Springer.

Satria, R., Agüero-Valverde, J., and Castro, M. (2021), Spatial Analysis of Road Crash Frequency using Bayesian Models with Integrated Nested Laplace Approximation (INLA), *Journal of Transportation Safety and Security*, **13**(11), 1240-1262.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002), Bayesian Measures of Model Complexity and Fit, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **64**(4), .583-639

Vicente, G., Goicoa, T., and Ugarte, M. D. (2020), Bayesian Inference in Multivariate Spatio-Temporal Areal Models using INLA: Analysis of Gender-based Violence in Small Areas, *Stochastic Environmental Research and Risk Assessment*, **34**(10), .1421-1440