

# تحلیل مدل‌های شمارشی فضایی روی تعداد روزهای هوای ناسالم شهر تهران

امید کریمی<sup>۱</sup>، فاطمه حسینی

تاریخ دریافت: ۹۹/۳/۱۲

تاریخ پذیرش: ۹۹/۱۱/۱

چکیده:

داده‌های شمارشی فضایی در اغلب علوم مانند علوم محیطی، هواشناسی، زمین‌شناسی و پزشکی مشاهده می‌شود. برای تحلیل داده‌های رسته‌ای شمارشی که همبستگی مکانی در آن‌ها مشاهده می‌شود اغلب از مدل‌های خطی تعمیم‌یافته فضایی بر اساس توزیع‌های پواسونی (مدل فضایی پواسون-لگ-نرمال) و دو جمله‌ای (مدل فضایی دو جمله‌ای-لوجیت نرمال) استفاده می‌شود. تابع درستیابی این نوع مدل‌ها دارای پیچیدگی‌های تئوری و محاسباتی است. رهیافت بیزی به واسطه الگوریتم‌های مونت کارلویی زنجیر مارکوف یک راه‌حل برای برآزش این مدل‌ها می‌تواند باشد، هرچند مشکلاتی از لحاظ نرخ پایین پذیرش نمونه‌ها و طولانی شدن زمان اجرای الگوریتم‌ها معمولاً وجود دارد. یک راه‌کار مناسب استفاده از الگوریتم مونت کارلویی همیلتونی (هیبریدی) در رهیافت بیزی است. در این مقاله، روش جدید مونت کارلویی همیلتونی برای تحلیل بیزی مدل‌های شمارشی فضایی روی داده‌های آلودگی هوای شهر تهران مورد مطالعه قرار می‌گیرد. همچنین دو الگوریتم مونت کارلویی معمول زنجیر مارکوفی (گیبز و متروپولیس-هستینگس) و لانجون-هستینگس برای رهیافت بیزی کامل مدل‌ها روی داده‌ها به کار گرفته می‌شوند. در نهایت با ملاک‌های تشخیصی، رهیافت مناسب برای تحلیل داده‌ها و پیشگویی در همه نقاط شهر معرفی می‌شود. **واژه‌های کلیدی:** مدل‌های شمارشی فضایی، مدل‌های خطی تعمیم‌یافته، داده‌های فضایی، الگوریتم مونت کارلویی همیلتونی.

## ۱ مقدمه

الگوریتم‌ها، عدم همگرایی الگوریتم‌ها بیان شده است. برای حل بعضی از این مشکلات از اجرای محاسبات به صورت موازی [۹] و الگوریتم‌های MCMC و لانجون-هستینگس استفاده شده است. در این مقاله به دلیل چالش‌هایی که الگوریتم MCMC دارد از الگوریتم مونت کارلویی همیلتونی<sup>۳</sup> (HMC) استفاده می‌شود. الگوریتم HMC اولین بار توسط [۵] به نام روش مونت کارلویی هیبریدی در مسائل فیزیک کوانتومی شبک‌های معرفی شد. سپس توسط رادفورد [۱۰] در تحلیل‌های بیزی شبکه‌ای وارد متون آماری شد و با توجه به نرخ بالای پذیرش نمونه‌های مونت کارلویی و تولید نمونه‌هایی با خودهمبستگی کم اخیراً مورد توجه خیلی از محققین آمار بیزی مانند: [۱۱]، [۱]، [۲] قرار گرفته است. داده‌های فضایی که در این مقاله مورد مطالعه قرار می‌گیرند از نوع زمین‌آمار است. در بخش دوم مقاله، مدل فضایی پیشنهادی برای داده‌های شمارشی زمین‌آمار معرفی می‌شود. در بخش سوم تحلیل اکتشافی روی داده‌های آلودگی هوا و تعداد روزهای هوای پاک در شهر تهران ارائه می‌گردد. همچنین رهیافت بیزی بر روی داده‌های آلودگی هوا انجام و مقایسه مدل‌ها مورد بررسی قرار می‌گیرد و در نهایت بحث و

داده‌های فضایی شمارشی امروزه به‌طور مرتب در علوم مختلفی مانند هواشناسی (تعداد روزهای ناسالم هوای شهر)، زمین‌شناسی (تعداد لرزه‌های ثبت‌شده)، جنگل‌داری (تعداد درختان آلوده) و اپیدمیولوژی (تعداد بیماران مبتلا به کوید ۱۹) و بسیاری علوم دیگر مشاهده می‌شوند. با توجه به این‌که داده‌ها ناگوسی و دارای همبستگی فضای (از نوع مکان) هستند، اغلب برای مدل‌بندی این نوع داده‌ها از مدل‌های فضایی خطی تعمیم‌یافته<sup>۲</sup> (GLSM) استفاده می‌شود، [۳، ۴، ۶، ۷]. در این مدل‌ها معمولاً تابعی از متغیر مورد نظر تحقیق از یک میدان تصادفی گاوسی می‌باشد و بنابراین تابع درستیابی پیچیده خواهد شد و دارای شکل بسته‌ای نیست و نیازمند الگوریتم‌های عددی برای برآزش مدل است که اغلب از الگوریتم‌های پیشینه‌سازی امید ریاضی (EM) و مونت کارلویی EM استفاده می‌شود، [۱۳]. این الگوریتم‌ها دارای چالش‌هایی در اجرا هستند، شایع‌ترین آن‌ها در [۸] به صورت محاسبات سنگین، زمان‌بر بودن، حافظه زیاد برای اجرای

<sup>۱</sup> دانشگاه سمنان، گروه آمار [omid.karimi@semnan.ac.ir](mailto:omid.karimi@semnan.ac.ir)

<sup>۲</sup> Generalized Linear Spatial Models

<sup>۳</sup> Hamiltonian Monte Carlo

نتیجه‌گیری ارائه می‌شود.

$$\log(\mu_i) = S(s_i), \quad i = 1, \dots, n$$

۲- مدل دوجمله‌ای-لوجیت‌نرمال

$$Y_i | S(s_i) \sim \text{Binomial}(n_i, P_i), \\ \mu_i = n_i P_i, \quad P_i = \frac{\exp\{S(s_i)\}}{1 + \exp\{S(s_i)\}}, \quad i = 1, \dots, n.$$

بنابراین تابع درستنمایی مدل شمارشی فضایی به صورت

$$L(\eta | y) = f(y | \eta) \\ = \int_S \left( \prod_{i=1}^n f(y_i | S(s_i)) \right) f(S | \eta) dS,$$

بیان می‌شود که در آن  $f(y_i | S(s_i))$  تابع احتمال متغیر شمارشی  $Y_i$ ،  $f(S | \eta)$  تابع چگالی متغیر پنهان  $S$  و  $\eta = (\beta, \theta)'$  است. تابع درستنمایی این مدل‌ها دارای پیچیدگی محاسباتی هستند. به وسیله رهیافت بیزی پارامترهای مدل برآورد و مدل‌ها برازش می‌شوند. در این مطالعه پیشین‌های معمول به صورت

$$\pi(\beta) \propto 1, \quad \pi(\sigma) \propto \frac{1}{\sigma}, \quad \pi(\phi) \propto I_{[b_1, b_2]}(\phi), \quad \pi(\kappa) \propto I_{[c_1, c_2]}(\kappa)$$

برای پارامترهای مدل در نظر گرفته شده است، که در آن  $b_1$  و  $b_2$  بر اساس تحلیل اکتشافی داده‌ها تعیین می‌شوند. برای مدل‌های ماترن مقادیر ابرپارامترها  $\kappa = (0.05, 1/95)$  اما معمولاً پارامتر  $\kappa$  با یک تحلیل اکتشافی روی داده‌ها ثابت در نظر گرفته می‌شود که از هزینه محاسباتی کاسته شود. به عنوان مثال  $\kappa = 0.5$  که تابع همبستگی نمایشی حاصل می‌شود.

### ۳ تحلیل اکتشافی داده‌ها

یکی از مسائل مهم در شهرهای بزرگ مثل تهران آلودگی هوا است. تعداد روزهای ناسالم یکی از فاکتورهای مهم برای انتخاب مکان زندگی مناسب افراد زیادی است که مشکلات تنفسی و ریوی دارند. در این تحقیق به مطالعه تعداد روزهای ناسالم در محله‌های مختلف شهر تهران بر اساس اطلاعات حاصل از ایستگاه‌های ثبت سنجش آلودگی هوا است. شاخص‌های مختلفی برای سنجش آلودگی‌ها توسط ایستگاه‌ها مانند  $CO$ ،  $O_3$ ،  $NO_2$ ،  $PM_{2.5}$  و ... ثبت می‌شوند. از شاخص کیفیت هوا<sup>۴</sup> (AQI) که توسط شرکت کنترل کیفیت هوا<sup>۵</sup> ارائه می‌شود، استفاده می‌کنیم.

## ۲ مدل شمارشی فضایی

فرض کنید داده‌های فضایی مورد بررسی به صورت سه تایی مرتب  $(s_i, Y_i, Z_i)$  برای  $i = 1, \dots, n$  در ناحیه  $A \subset R^2$  است، که در آن  $Y_i$  متغیر پاسخ گسسته شمارشی در موقعیت  $s_i$  و  $Z_i$  متغیر کمکی اندازه‌گیری شده در موقعیت نام است، که می‌تواند گسسته یا پیوسته باشد. همچنین موقعیت جغرافیایی به صورت  $s_i = (\text{longitude}_i, \text{latitude}_i) \in A \subset R^2$  در نظر گرفته شده است که در آن  $\text{longitude}_i$  طول جغرافیایی و  $\text{latitude}_i$  عرض جغرافیایی موقعیت نام هستند.

کلاس GLSMها اولین بار توسط [۴] به صورت

$$E(Y_i | S(s_i)) = \mu_i, \quad i = 1, \dots, n \\ \mu_i = g^{-1}(S(s_i)), \quad S = (S(s_1), \dots, S(s_n))' \\ S \sim N_n(X\beta, \Sigma\theta), \quad \Sigma\theta = \sigma^2 \rho_\phi, \quad \theta = (\sigma^2, \phi)' \quad (1)$$

معرفی شده است، که در آن  $g(\cdot)$  تابع پیوند و

$$X_{n \times (p+1)} = (1_n, x_1, \dots, x_p)$$

ماتریس طرح است.  $1_n$  یک بردار  $n$  بعدی با مقادیر یک و  $x_j$  متغیر کمکی  $(x_j(s_1), \dots, x_j(s_n))'$  در موقعیت نام است.  $\beta = (\beta_0, \dots, \beta_p)'$  بردار ضرایب رگرسیونی و  $\Sigma\theta$  ماتریس واریانس کوواریانس با مؤلفه‌های  $\|s_i - s_j\|$  است، که در آن  $\sigma^2 \rho_\phi > 0$  پارامتر واریانس و  $\rho_\phi$  تابع همبستگی همسانگرد با پارامتر  $\phi$  است. به طور معمول در این مدل‌ها تابع همبستگی همسانگرد ماترن به صورت

$$\rho(h) = \frac{1}{\Gamma(\kappa) 2^{2\kappa-1}} \frac{h^\kappa}{\phi} K_\kappa\left(\frac{h}{\phi}\right), \quad h = \|s_i - s_j\|, \quad \phi > 0, \quad \kappa > 0 \quad (2)$$

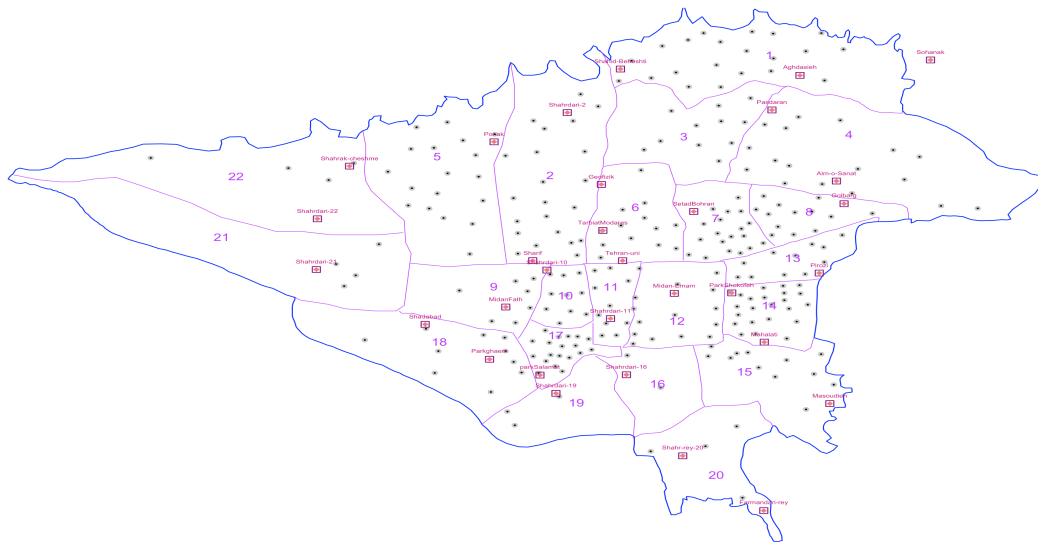
در نظر گرفته می‌شود، که در آن  $\Gamma(\cdot)$  تابع گاما و  $K_\kappa(\cdot)$  تابع بسل با مرتبه  $\kappa$  است. با قرار دادن  $\kappa = \frac{1}{2}$  تابع همبستگی نمایشی  $\rho_\phi = e^{-\frac{h}{\phi}}$  حاصل می‌شود.  $\phi$  پارامتر دامنه است که نرخ پایان همبستگی را کنترل می‌کند و  $\kappa$  پارامتر هموارساز برای کنترل همواری میدان تصادفی  $S(\cdot)$  است. دو مدل فضایی معمول برای داده‌های شمارشی به صورت زیر است:

۱- مدل پواسون-لگ-نرمال

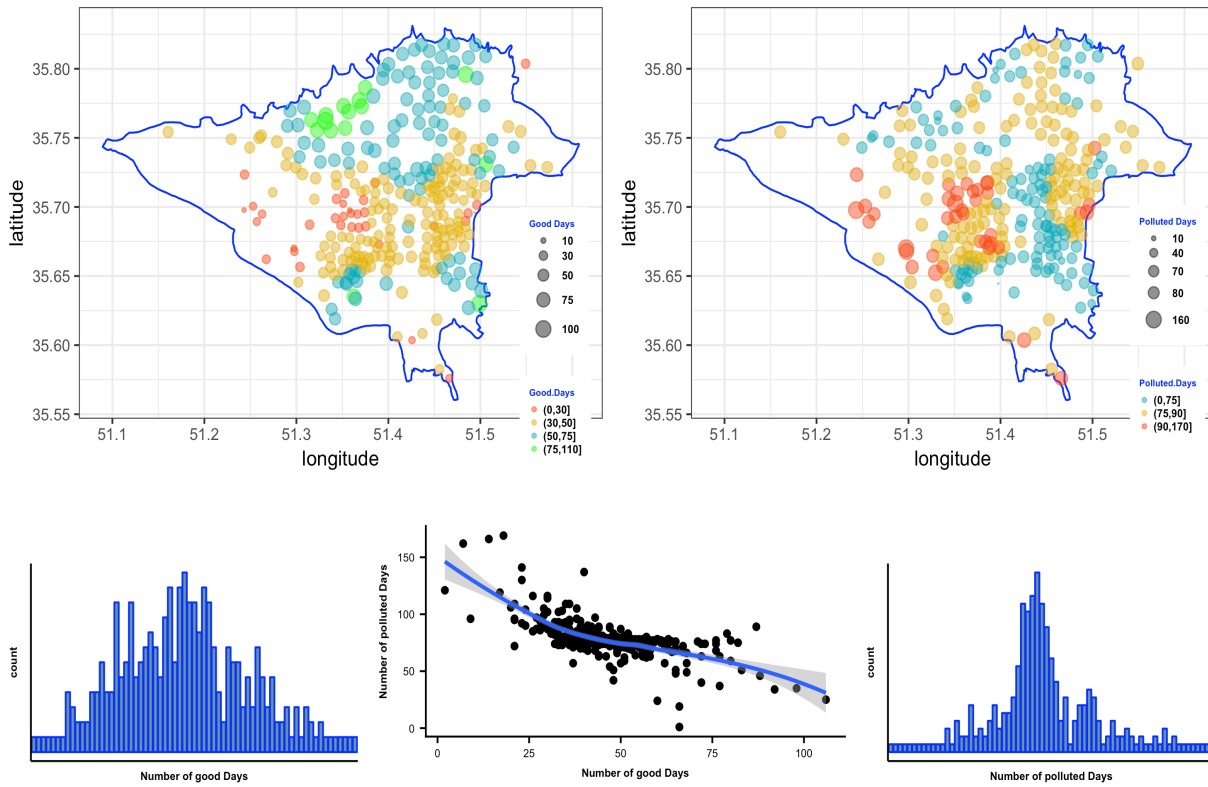
$$Y_i | S(s_i) \sim \text{Poisson}(\mu_i),$$

<sup>4</sup> Air Quality Index

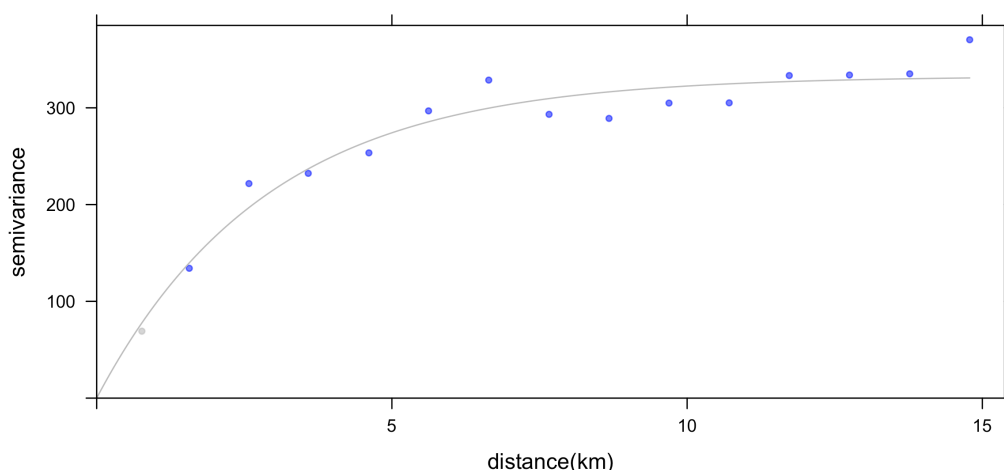
<sup>5</sup> <http://air.tehran.ir>



شکل ۱: موقعیت ۳۱ ایستگاه و ۲۸۶ محله بر روی نقشه شهر تهران



شکل ۲: نمودار حبابی تعداد روزهای ناسالم شهر تهران به تفکیک محله‌های مختلف



شکل ۳: نمودار واریوگرام تجربی تعداد روزهای ناسالم.

تعریف می‌شود، که در آن  $\alpha \in R$  پارامتر ضریب متغیر کمکی  $Z_i$  است و ماتریس طرح به صورت  $x'_j = (1, \text{latitude}_i, \text{longitude}_i)$  فرض شده است که در آن  $\text{longitude}$  و  $\text{latitude}$  به ترتیب طول و عرض جغرافیایی هستند. برای تعیین ساختار همبستگی  $\rho_\phi$  از واریوگرام تجربی داده‌ها استفاده شده است. با توجه به شکل ۳ از ساختار همبستگی همسانگرد نمایی  $\rho_\phi(h) = e^{-\frac{h}{\phi}}$  که یک حالت خاص از ساختار همبستگی ماترن است، استفاده می‌شود. در نتیجه بردار پارامترهای مدل به صورت  $\eta = (\alpha, \beta, \phi, \sigma^2)$  است، که در آن  $\beta = (\beta_0, \beta_1, \beta_2)'$  می‌باشد، که پارامتر  $\beta_0$  ضریب ثابت، پارامتر  $\beta_1$  ضریب طول جغرافیایی و پارامتر  $\beta_2$  ضریب عرض جغرافیایی است. برای برآورد پارامترها و سپس برازش مدل و پیشگویی در نقاط جدید از رهیافت بی‌زی با سه رویکرد استفاده شده است:

۱- استفاده از روش‌های  $MCMC$  مثل روش گیبز و متروپولیس-هستینگس ( $MH$ )

۲- استفاده از روش لانجورین-هستینگس ( $LH$ )

۳- استفاده از روش مونت کارلویی همیلتونی ( $HMC$ )

همچنین دو مدل فضایی پواسون-لگ-نرمال (تابع پیوند لگاریتمی) و دو جمله‌ای-لوجیت-نرمال (تابع پیوند لوجیت) برای تحلیل داده‌های آلودگی هوا در نظر گرفته می‌شود. هر سه روش در ۵۰۰۰ تکرار با مرحله داغیدن ۲۵۰۰ اجرا شده‌اند. درصد نرخ پذیرش نمونه‌های تولید شده توسط الگوریتم‌های  $MCMC$  برای ۵۰۰۰ تکرار در جدول ۱ نشان می‌دهد که الگوریتم نسبت  $HMC$  به بقیه الگوریتم‌ها دارای عملکرد بهتر و همچنین زمان محاسبات کمتری نیاز دارد. نتایج برازش دو مدل (برآورد پارامترهای دو مدل، انحراف

برای تعیین کیفیت هوا، اگر  $AQI > 100$  باشد هوا ناسالم است و اگر  $50 < AQI \leq 100$  هوا قابل قبول و اگر  $AQI \leq 50$  هوا سالم است. این نحوه طبقه‌بندی نیز بر اساس استانداردهای کیفیت هوا است. داده‌ها پس از استخراج و طبقه‌بندی برای سال ۹۸ برحسب محله‌های مختلف شهر تهران در شکل‌های ۱ و ۲ ارائه شده‌اند. شکل ۱ موقعیت ایستگاه‌ها (۳۱ ایستگاه) و محله‌های شهر تهران (۲۸۶ محله) را نشان می‌دهد. تعداد روزهای ناسالم<sup>۶</sup> و تعداد روزهای پاک<sup>۷</sup> برحسب مکان در شکل ۲ به صورت نقاط حبابی شکل و طیف رنگی نمایش داده شده است. این شکل نشان می‌دهد که کدام مکان‌های شهر در طول سال ۹۸ دارای تعداد روزهای بیشتر و کمتر دارای هوای ناسالم و پاک بوده‌اند. همچنین در شکل ۲ نمودار میله‌ای و پراکنش این دو متغیر شمارشی برای مشاهده نوع پراکنندگی و ارتباط بین آن‌ها رسم شده‌اند. مشاهده می‌شود که نمودار پراکنش (همراه با خط هموارساز) یک رابطه در جهت مخالف هم بین این دو متغیر وجود دارد. از این رو در ادامه برای مدل‌بندی تعداد روزهای هوای ناسالم  $Y_i$ ، از تعداد روزهای هوای پاک  $Z_i$  به عنوان یک متغیر کمکی استفاده می‌شود. هر چند می‌توان هر دو را به عنوان متغیر پاسخ دو متغیری نیز مدل‌بندی کرد، که در این مقاله فقط به حالت اول پرداخته می‌شود و حالت دوم به عنوان پیشنهاد برای کارهای آتی مطرح است.

برای داده‌های آلودگی هوا مدل (۱) به صورت

$$E(Y_i | S(s_i)) = \mu_i, \quad i = 1, \dots, 299$$

$$\mu_i = g^{-1}(S(s_i) + \alpha Z_i), \quad S = (S(s_1), \dots, S(s_{299}))'$$

$$S \sim N_{299}(X\beta, \Sigma_\theta), \quad \Sigma_\theta = \sigma^2 \rho_\phi, \quad \theta = (\sigma^2, \phi)'$$

<sup>6</sup>Polluted Days

<sup>7</sup>Good Days

$$P_{WAIC} = \sum_{i=1}^n \text{Var}_{\eta|Y}(\log f(y|\eta)),$$

و  $\text{Var}_{\eta|Y}$  واریانس پسینی بر اساس توزیع پسین  $\pi(\eta|y)$  است. با توجه به مقادیر  $MAPE$  و  $WAIC$  در جدول ۲ مدل پواسون-لگ نرمال مناسب تر از مدل دوجمله‌ای-لوجیت نرمال است. از این رو این مدل برای داده‌های آلودگی هوا انتخاب می‌شود. با توجه به مقادیر فاصله باورمندی ۹۵٪ پارامترها می‌توان نتیجه گرفت که متغیرها کمی هوای پاک تأثیر معنی‌داری با ضریب منفی روی متغیر پاسخ دارد ولی متغیرهای طول ( $lat$ ) و عرض ( $lon$ ) جغرافیایی تأثیر معنی‌داری روی متغیر پاسخ یعنی تعداد روزهای آلوده سال ۱۳۹۸ ندارد.

استاندارد و فاصله باورمندی) فقط برای روش  $HMC$  در جدول ۲ به طور خلاصه ارائه شده است. برای انتخاب مدل مناسب از معیارهای میانگین قدر مطلق خطای پیش‌گویی<sup>۸</sup> ( $MAPE$ ) و معیار اطلاع آکائیکه<sup>۹</sup>  $WAIC$  [۱۲] به صورت

$$MAPE = n^{-1} \sum_{i=1}^n |y_i - \hat{y}_i|,$$

$$WAIC = -2LPPD + 2P_{WAIC},$$

استفاده شده است که در آن‌ها

$$LPPD = \sum_{i=1}^n \log \left( \int f(y|\eta)\pi(\eta|y)d\eta \right),$$

جدول ۱. نرخ پذیرش و زمان اجرای الگوریتم‌های  $MCMC$  در ۵۰۰۰ تکرار برای دو مدل

روش‌های $MCMC$			مدل
$HMC$	$LH$	$MH$	
۸۵	۵۶	۴۳	پواسون (نرخ پذیرش (درصد)
۲۱/۶۵	۶۶/۱۲۵	۷۵/۱۶۰	زمان اجرا (دقیقه)
۸۱	۴۷	۳۵	دوجمله‌ای (نرخ پذیرش (درصد)
۱۳/۷۷	۱۸/۱۳۰	۳۲/۱۸۵	زمان اجرا (دقیقه)

جدول ۲. نتایج برازش مدل‌های شمارشی فضایی توسط روش  $HMC$  برای داده‌های آلودگی شهر تهران

پارامتر	مدل					
	پواسون-لگ نرمال			دوجمله‌ای-لوجیت نرمال		
	برآورد	خطای استاندارد	فاصله باورمندی ۹۵٪	برآورد	خطای استاندارد	فاصله باورمندی ۹۵٪
$\sigma$	۱/۲۰	۰/۲۱	(۰/۲۷, ۲/۷۱)	۱/۱۵	۰/۱۶	(۰/۳۴, ۳/۲۹)
$\phi$	۸/۸۷	۳/۰۴	(۰/۲۳, ۲۸/۵۴)	۸/۵۰	۱/۹۷	(۰/۵۶, ۳۴/۷۶)
$\beta_0$	۵/۰۲	۳/۲۵	(-۸/۳۰, ۳۳/۹۶)	-۲/۹۰	۳/۷۶	(-۲۴/۲۴, ۱۹/۲۹)
$\beta_1$	۰/۶۴	۰/۱۱	(-۰/۰۶, ۱/۵۴)	۰/۴۵	۰/۱۵	(-۰/۵۵, ۱/۳۵)
$\beta_2$	-۰/۹۳	۰/۲۰	(-۲/۰۹, ۰/۰۴)	-۰/۴۳	۰/۲۵	(-۱/۹۱, ۰/۹۰)
$\alpha$	-۰/۰۱	۰/۰۰۰۱	(-۰/۰۱۰۰۲, -۰/۰۱۰۰۱)	-۰/۰۱	۰/۰۰۰۱	(-۰/۰۱۰۰۲, -۰/۰۱۰۰۱)
$MAPE$	۶/۳۴۴			۷/۲۵۳		
$WAIC$	۲۳۷۱/۲			۲۳۸۱/۳		

نقشه پیش‌گویی و انحراف معیار پیش‌گویی در شکل ۴ رسم شده است

که شدت آلودگی محله‌های مختلف شهر تهران را نمایش می‌دهد.

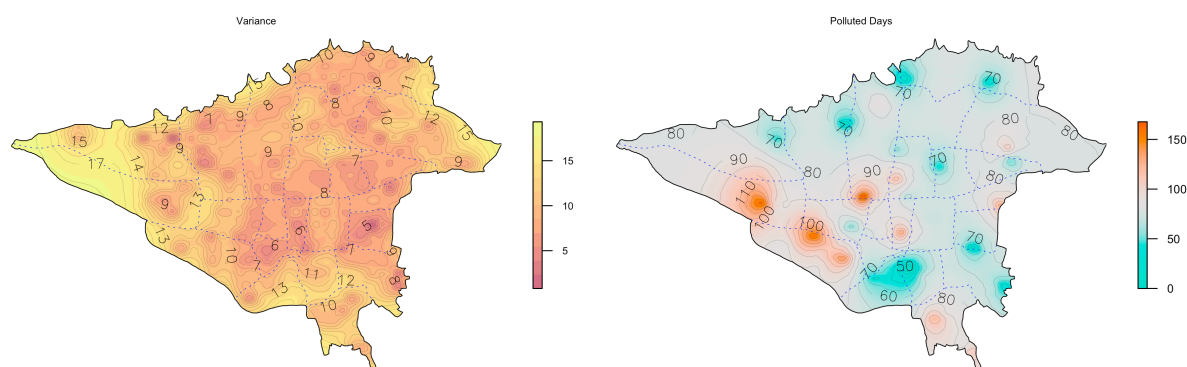
کار گرفته شد. برای برازش مدل‌ها از رهیافت بیزی با الگوریتم کارای  $HMC$  استفاده شد و این الگوریتم با الگوریتم‌های  $MH$  و  $LH$  مقایسه شد. با توجه به نتایج برای این داده‌ها رهیافت بیزی  $HMC$  دارای عملکرد بهتر و

## بحث و نتیجه‌گیری

در این مقاله مدل‌های شمارشی فضایی پواسون-لگ نرمال و دوجمله‌ای-لوجیت نرمال روی داده‌های تعداد روزهای آلوده شهر تهران در سال ۱۳۹۸ به

<sup>8</sup>Mean Absolute Prediction Error

<sup>9</sup>Watanabe Akaike Information Criterion



شکل ۴: نمودار نقشه پیش‌گویی و واریانس پیش‌گویی.

هیچ‌کدام تأثیر معنی‌داری نداشتند. این نتیجه در نقشه پیش‌گویی داده‌ها (شکل ۴) کاملاً مشهود است که روند خاصی در جهت شمال، جنوب، شرق و غرب مشاهده نمی‌شود و بعضی محله‌ها در شرق، جنوب، غرب و مرکز دارای تعداد روزهای آلوده بیشتر نسبت به بقیه محله‌ها هستند. به‌عنوان پیشنهاد، با توجه به وجود دو متغیر شمارشی (تعداد روزهای ناسالم و تعداد روزهای پاک) در این نوع از داده‌های فضایی استفاده از روش کریگیدن دومتغیره (کوکرینگینگ) برای تحلیل آن‌ها می‌تواند مفید باشد.

سریع‌تری بود. برای اجرای روش بیزی  $HMC$  از نرم‌افزار  $Stan$  و برای اجرای الگوریتم‌های  $MH$  و  $LH$  به ترتیب از پکیج‌های  $geoRglm$  و  $geoCount$  در  $R$  استفاده شده است. همچنین با توجه به تحلیل اکتشافی داده‌ها، متغیرهای کمکی تعداد روزهای هوای پاک، طول و عرض جغرافیایی در نظر گرفته شد. نتایج حاصل از هر دو مدل پواسون-لگ-نرمال و دو جمله‌ای-لوجیت-نرمال تأثیر معنی‌دار تعداد روزهای هوای پاک در جهت معکوس (ضریب منفی) روی متغیر پاسخ را تأیید کرده‌اند و متغیرهای کمکی طول و عرض جغرافیایی

## مراجع

- [1] Betancourt, M. (2013). A General Metric for Riemannian Hamiltonian Monte Carlo. *In First International Conference on the Geometric Science of Information (F. Nielsen and F. Barbaresco, eds.)*. Lecture Notes in Computer Science 8085, Springer.
- [2] Betancourt, M. (2018). *A Conceptual Introduction to Hamiltonian Monte Carlo*. <https://arxiv.org/abs/1701.02434>
- [3] Breslow, N. E. and Clayton, D. G. (1993). Approximate Inference in Generalized Linear Mixed Models, *Journal of the American Statistical Association*, **88**, 9-25.
- [4] Diggle, P., Tawn, J.A. and Moyeed, R. A. (1998). Model-Based Geostatistic, *Journal of the Royal Statistical Society, Series C. Applied Statistics*, **47**, 299-350.
- [5] Duane, S., Kennedy, A. D., Pendleton, B.J., Roweth, D. (1987). Hybrid Monte Carlo. *Physics Letters B*. **195** (2), 216-222.
- [6] Hosseini, F. (2016). A New Algorithm for Estimating the Parameters of the Spatial Generalized Linear Mixed Models, *Environmental and Ecological Statistics*, **23**, 205-217.
- [7] Hosseini, F. and Karimi, O. (2019). Approximate Composite Marginal Likelihood Inference in Spatial Generalized Linear Mixed Models, *Journal of Applied Statistics*, **46**(3), 542-558.
- [8] Jing, L. and De Oliveira, V. (2015). geoCount: An R Package for the Analysis of Geostatistical Count Data, *Journal of Statistical Software, Foundation for Open Access Statistics*, **63**(i11), DOI: <http://hdl.handle.net/10.18637/jss.v063.i11>.

- [9] Knaus, J. (2013). **snowfall**: Managing Parallel Execution of R Programs on a Compute Cluster. R package version 1.84-4, URL <http://CRAN.R-project.org/package=snowfall>.
- [10] Neal, R. M. (1995). *Bayesian Learning for Neural Networks*, PhD thesis, University of Toronto.
- [11] Neal, R. M. (2011). *MCMC Using Hamiltonian Dynamics*. In *Handbook of Markov Chain Monte Carlo* (S. Brooks, A. Gelman, G. L. Jones and X.-L. Meng, eds.) CRC Press, New York.
- [12] Watanabe, S. (2013). A Widely Applicable Bayesian Information Criterion *Journal of Machine Learning Research*. **14**, 867-897.
- [13] Zhang, H. (2002). On Estimation and Prediction for Spatial Generalized Linear Mixed Models, *Biometrics*, **58**, 129-136.