

广义相加模型在空气污染与呼吸系统疾病的关系中的应用

米静 张诗荟

河北北方学院, 河北 张家口 075000

摘要: 通过收集河北省石家庄市和张家口市 2013 年到 2019 年两个地市的大气污染、气象及相关疾病数据, 以广义相加模型 (GAM) 为基础, 分析空气污染对居民呼吸系统疾病的影响。

关键词: 广义相加模型 (GAM); 空气污染; 呼吸系统疾病

Application of Generalized Additive Model (GAM) in the Relationship between Air Pollution and Respiratory Diseases

Mi, Jing Zhang, Shihui

Hebei North University, Zhangjiakou, Hebei, 075000, China

Abstract: By collecting the data of air pollution, meteorology and related diseases in Shijiazhuang City and Zhangjiakou City of Hebei Province from 2013 to 2019, based on Generalized Additive Model (GAM) model, this paper analyzes the influence of air pollution on residents' respiratory diseases.

Keywords: GAM; Air pollution; Respiratory diseases

DOI: 10.62639/sspis23.20250201

一、研究目的与意义

在过去的两个世纪里, 工业化和城市化进程的快速推进, 使人类赖以生存的环境遭遇前所未有的生态灾难。现代流行病学研究证实, 大气污染与众多健康效应直接或间接相关, 可对人体多个系统 (呼吸、循环、神经、泌尿和免疫等) 产生不良影响。我国改革开放四十多年来, 经济的高速发展和城镇化速度加快也带来了严重的生态问题, 环境污染对人群健康的危害进入了显现期, 而且未来一段时间内形势会更加严峻。本研究以河北省石家庄市、张家口市为主要研究区域, 收集 2013 年到 2019 年两个地市的大气污染、气象数据, 2018 年到 2019 年相关疾病数据, 利用广义相加模型开展空气污染与呼吸系统疾病的关系的研究。

二、研究数据

(一) 研究数据的选取

本研究收集了河北省石家庄市和张家口市 2018 年 1 月 1 日 ~ 2019 年 12 月 31 日的日空气污染物监测数据, 包括空气质量指数 (Air Quality Index, AQI), PM_{2.5}, PM₁₀, SO₂, NO₂, CO, O₃, 均来自于国家环保部网站 (<http://www.mee.gov.cn>) 发布的同时期每日大气污染数

据。本研究还收集了 2018 年 1 月 1 日 ~ 2019 年 12 月 31 日医院每日呼吸系统疾病门诊人数和住院人数数据。相关数据是通过医院呼吸系统门诊登记和住院登记的病例, 来源于河北省某三甲医院医院。

(二) 研究理论与方法

1. GAM 的基本理论

GAM 应用的前提是应该是具有可加性的, 各分量是平滑的, 通过对部分或全部的解释变量利用平滑函数建立模型。GAM 的基本形式^[1]为:

$$g(\mu_i) = \beta_0 + f_1(x_{1i}) + f_2(x_{2i}) + f(x_{3i}) + \dots \quad (式 1)$$

$g(\mu_i)$ 代表各种连接函数关系, 可以是多种概率分布, 包括: 正态分布、二项分布、Poisson 分布、负二项分布等^[2]。 $f_1(x_{1i})$ 、 $f_2(x_{2i})$ 、 $f(x_{3i})$... 代表各种平滑函数, 如光滑样条 (smoothing spline, S) 平滑、局部回归 (local regression, Loess) 平滑、自然立方样条 (natural cubic spline, ns)、B 样条和多项式 (B spline, B) 等。

2. 基于时间序列的 GAM 分析理论

按照时间顺序排列起来的观测值称为时间序列。GAM 通过一系列函数, 控制时间趋势项和混杂因素影响, 从而正确评价解释序列的效应。时间序列分析的 GAM 通过最小信息准则法选择

(稿件编号: IS-25-1-17013)

作者简介: 米静 (1990-), 女, 汉族, 河北张家口人, 河北北方学院, 讲师, 硕士学历, 硕士研究生学位, 研究方向: 大数据通信。张诗荟 (1990-), 女, 汉族, 河北张家口人, 河北北方学院, 副教授, 硕士学历, 硕士研究生学位, 研究方向: 自动控制系

基金项目: 2023 年度河北北方学院基本科研业务费项目: “基于广义相加模型的空气污染与呼吸系统疾病关系的研究” (项目编号 JYT2023007)。

河北北方学院 2020 年度校级青年基金项目: “广义相加模型在空气污染与呼吸系统疾病的关系中的研究与应用”。

最优模型。

3. GAM 中以二次函数拟合健康效应

由于呼吸系统疾病每日的死亡情况为小概率事件, 从统计学分布角度看, 与泊松分布近似, 所以本研究选择建立泊松分布广义相加模型。本研究应用 GAM 对各污染物指标与每日医院呼吸系统疾病死亡人数之间的关系进行了时间序列分析。

$$\log[E(Y_i)] = \alpha + \sum_{i=1}^m \beta_i X_i + \sum_{i=1}^n \delta_i Z_i + \omega_i(week) \quad (式 2)$$

公式中 Y_i 为观察日当天呼吸系统疾病死亡人数, $E(Y_i)$ 为观察日呼吸疾病死亡人数的数学期望, α 为截距, β 为回归系数, X_i 为颗粒物浓度, δ_j 为非参数平滑函数, Z_j 为对应变量产生非线性影响的变量, $\omega_i(week)$ 为星期哑元变量, 用于解释星期效应。本研究中效应的估计值用相对危险度 (Relative Risk, RR) 表示, 空气污染物指标每升高一个单位, 呼吸系统感染升高值。

对污染物和医院呼吸系统疾病每日门诊量和住院量用 SPSS 24.0 进行分析, 用 R 4.0.1 软件构件 GAM 模型, 分析空气污染物对呼吸系统疾病发生的影响和阈值效应, 其中校验水准 $\alpha = 0.05$ 。

三、结果分析

(一) 大气污染对呼吸系统疾病的影响

本研究在 GAM 在空气污染健康效应分析中的应用, 仅分析了污染的即时效应, 即当日效应, 不考察延迟效应。表 1 为大气污染物 24h 平均浓度每增加 $10 \mu\text{g}/\text{m}^3$ 居民因呼吸系统疾病死亡的相对危险度。

表 1 大气污染物 24h 平均浓度每增加 $10 \mu\text{g}/\text{m}^3$ 居民因呼吸系统疾病死亡的相对危险度

污染物	RR	RR95% 可信限
PM _{2.5}	1.0011	1.0002~1.0020
PM ₁₀	1.0030	1.0005~1.0055
SO ₂	1.0068	1.0018~1.0118
NO ₂	1.0055	1.0024~1.0087

结果表明, PM_{2.5} 每日浓度上升 $10 \mu\text{g}/\text{m}^3$, 对应居民死亡增加 0.11% (0.02% ~ 0.2%); PM₁₀ 每日浓度上升 $10 \mu\text{g}/\text{m}^3$, 对应居民死亡增加 0.30% (0.05% ~ 0.55%); SO₂ 每日浓度上升 $10 \mu\text{g}/\text{m}^3$, 对应居民死亡增加 0.68% (0.18% ~ 1.18%); NO₂ 每日浓度上升 $10 \mu\text{g}/\text{m}^3$, 对应居民死亡增加 0.55% (0.24% ~ 0.87%)。

(二) 气温对呼吸系统疾病的影响

利用 GAM 中的自然立方样条函数控制相

对湿度、星期效应, 以对数线性的形式控制 PM_{2.5}、PM₁₀、SO₂、NO₂ 的影响。以此为前提, 拟合日均气温与居民死亡人数对数的二次函数关系, 分析在过去一周里日均气温对居民死亡的影响。如图 3 所示。

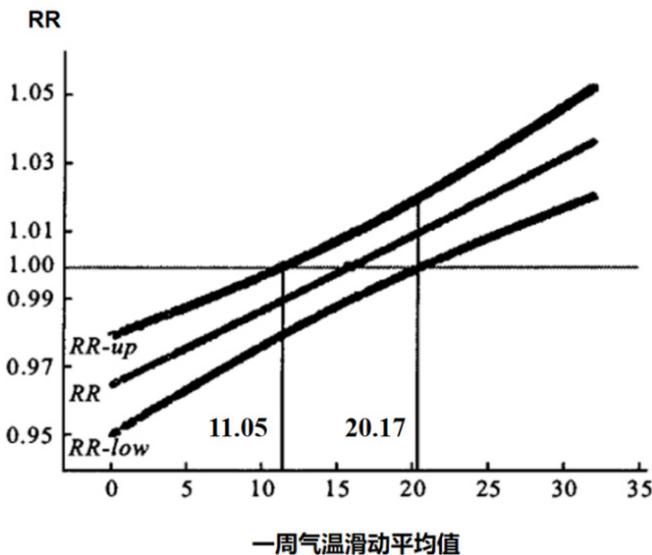


图 3 气温变化—RR 图

当气温 $< 11.05^\circ\text{C}$ 时, RR 均小于 1 (95% 可信区间不包含 1), 当气温 $> 20.17^\circ\text{C}$ 时, RR 均大于 1 (95% 可信区间不包含 1); 在这两个温度区间, 内气温变化引起的 RR 的 95% 可信区间包含 1。因此, 本研究认为, 在 $11.05^\circ\text{C} \sim 20.17^\circ\text{C}$ 范围内, 大气污染对死亡相对危险度无显著影响。

四、结论

结合国内其他地区开展的基于时间序列的大气污染与呼吸系统疾病关联的研究发现, 大气污染物与呼吸系统疾病日门诊量、呼吸系统疾病死亡风险存在正相关。本研究采用 GAM 分析了 PM_{2.5}、PM₁₀、SO₂、NO₂ 对呼吸系统疾病的影响, 结果发现空气中的 PM_{2.5}、PM₁₀、SO₂、NO₂ 会增加呼吸道感染、相关肺炎等呼吸系统疾病发生的风险。本研究还分析了日均气温变化对日死亡率的影响, 用 GAM 通过平滑函数拟合时间趋势项, 通过不同函数控制其他混杂因素对反应变量的作用, 应用灵活方便。

参考文献:

[1] 徐东群, 许群. 空气污染人群健康风险评估方法及应用 [M]. 北京: 人民卫生出版社, 2018

[2] Sun, Xinghua, et al. Forewarn system for college student work based on cloud platform and mobile internet. IPPTA: Quarterly Journal of Indian Pulp and Paper Technical Association[J]. 2018, 30(3):329-335.

[3] Ye, Yongfei, Xinghua Sun, et al. Intelligent on-demand routing protocol for ad hoc network. Journal of Information Processing Systems[J]. 2020, (16)5: 1113-1128.