arXiv:2302.02110v1 [stat.AP] 4 Feb 2023

用于估计环境暴露短期健康效应的标量对分位数函数方法

Yuzi Zhang * [1]、Howard H. Chang[1], 、Joshua L. Warren[2], 和 Stefanie T. Ebelt[3]

[1]埃默里大学生物统计学与生物信息学系，美国佐治亚州亚特兰大[2]耶鲁大学生物统计学系，美国康涅狄格州纽黑文[3]埃默里大学环境卫生系，美国佐治亚州亚特兰大

摘要

环境流行病学研究通常利用汇总健康结局，估计日益精细的空间分辨率下可获得的短期（例如每日）暴露的效应。然而，通常使用区域平均值来推算人群层面暴露，这无法捕捉感兴趣的空间和时间单元（例如一天或一个邮政编码）内可能出现的暴露空间变异和个体异质性。我们提出一种通用建模方法，通过暴露分位数函数将单位内的暴露异质性纳入健康分析。此外，将暴露分位数函数视为函数型协变量后，我们的方法能够更灵活地刻画不同分位数水平上的关联。我们将所提出的方法应用于一项分析，考察亚特兰大四年间空气污染与急诊就诊的关系。该分析使用随机人体暴露与剂量模拟器模拟的四种交通相关环境空气污染物个人暴露的每日邮政编码级分布。我们的分析发现，随着人群暴露较低分位数的变化，一氧化碳对呼吸系统和心血管疾病急诊就诊的影响更为显著。实现该方法的软件已在 R 软件包 nbRegQF 中提供。

关键词：空气污染、贝叶斯层级建模、函数型数据分析、分位数过程

*yuzi.zhang@emory.edu

1 引言

环境流行病学研究通常利用汇总健康数据评估环境暴露的短期健康效应。例如，城市范围内每日死亡人数、急诊科（ED）就诊人数和早产人数，已被用于研究短期暴露于空气污染和极端温度的影响（Alhanti et al., 2016；Guo et al., 2017；Bekkar et al., 2020；Yoo et al., 2021）。使用汇总健康数据体现了一种生态学设计，其中真实暴露对应于风险人群中个体层面暴露的分布（Richardson and Best, 2003；Sheppard, 2003）。然而，在基于大规模人群的研究中，实际上无法测量个体环境暴露。因此，研究会使用反映个体层面暴露汇总情况（例如均值）的暴露替代指标。这可能导致暴露错分，进而造成偏倚，并错误地描述健康效应估计值相关的不确定性（Dominici et al., 2000；Richmond-Bryant and Long, 2020）。

在空气污染研究中，环境污染物水平通常是关注的暴露对象，以便制定监管政策。在所关注的短期时间尺度内，群体内部可能存在显著的暴露异质性，因为：（1）空气污染物在研究区域内可能呈现空间变异；（2）个体大部分时间处于不同的室内环境中，而各类室内环境的室外污染物渗透特征各不相同。为应对第一个挑战，研究人员开发了时空模型，以细空间分辨率估计污染物浓度，从而实现完整的时空覆盖（Jerrett et al., 2001, 2005）。随后，通过人口加权构建短期暴露值，并将其与感兴趣空间单元（如邮政编码）内汇总的健康结局数据相联系。研究表明，在时间序列研究中考虑空间暴露变异，与使用稀疏布设的固定监测仪测得的暴露指标相比，可得到更大的健康效应

估计值，尤其是对于空间变异程度较高的污染物（如交通相关污染物）（Goldman et al., 2010; Sarnat et al., 2013）。

相比之下，针对个体在不同微环境中停留时间所导致的暴露异质性，相关研究较少。随着可穿戴设备的发展，环境污染物的个人暴露更容易测量（Steinle 等，2013、2015；Sugg 等，2018）。虽然在基于人群的流行病学研究中，测量成本高得令人难以承受，但小规模暴露评估研究的发现已被用于开发概率模型，以模拟人群层面的个人暴露分布；这些模型纳入了时间-活动-地点模式，以及描述室外与室内环境之间关系的参数。将空间精细化室外暴露作为输入时，得到的模拟个人暴露分布既反映了室外浓度的空间异质性，也反映了人群时间-活动-地点模式的异质性。概率模型的例子包括 pCNEM 和随机人体暴露与剂量模拟（SHEDS）（Burke 等，2001；Zidek 等，2005、2007），以及用于降低大规模流行病学研究计算负担的统计模拟器（Chang 等，2012）。

利用此类概率模型模拟的个人暴露，既往关于汇总健康结局的研究将暴露分布汇总统计量（例如日均值或中位数）作为健康模型中的协变量（Calder 等，2008；Chang 等，2012；Sarnat 等，2013）。然而，生态学偏倚仍然存在，因为个人暴露的日均值无法充分体现人群所经历暴露的异质性（Sheppard，2003）。例如，在个人暴露呈正态分布这一特殊情况下，将暴露方差纳入健康模型可以得到

无偏估计（Sheppard，2003；Reich 等人，2009）。然而，环境暴露在人群中的分布往往呈偏态，难以用正态分布准确近似（Leiva 等，2008；Huang 等，2018）。此外，所有既往采用总体汇总统计量的方法都隐含地假设，空气污染的健康效应可以完全由所选汇总统计量表征。换言之，健康风险只取决于该汇总统计量的变化。然而，健康风险也可能取决于暴露分布的变化，而这种变化无法通过所选汇总统计量的变化反映出来。例如，考虑这样一种情景：短期空气污染增加对通常暴露于较低污染水平的个体可能危害更大。在这种情景下，总体层面暴露的下尾上升会导致不良健康事件风险大幅增加，而仅纳入总体平均暴露不足以充分表征暴露-反应关系。

在本研究中，我们提出了一种通用建模框架，通过暴露分位数函数纳入单位内总体暴露异质性。与仅使用单位层面的总体平均值或方差不同，暴露分位数函数能够全面概括整个研究期间单位内的完整暴露分布。在我们的框架中，暴露分位数函数被视为关于分位数水平的函数型协变量。因此，我们进一步允许暴露在不同分位数水平上的效应有所不同。我们在贝叶斯层次建模框架下进行估计和推断，并将暴露分位数函数估计相关的不确定性传递至健康效应估计中。

2 动机数据与应用

环境空气污染暴露已被确定为多种疾病的风险因素（Landrigan，2017年；Manisalidis 等，2020年），并显著加重全球疾病负担（Boogaard 等，2019年）。本研究采用时间序列设计，考察每日急诊就诊与环境空气污染暴露之间的短期关联（Bhaskaran 等，2013年）。我们获得了亚特兰大1999年1月1日至2002年12月31日期间的每日急诊记录计数。我们还按40个邮政编码制表区中的每一个分层统计就诊次数。我们分析了根据《国际疾病分类》第九版（ICD-9）诊断代码确定的三类急诊就诊原因：(1) 呼吸系统疾病（ICD-9代码：460-465、466.0、477、466.1、466.11、466.19、480-486、491-493、496、786.07）；(2) 呼吸系统疾病中的一个子集，仅包括哮喘或喘鸣（ICD-9代码：493、786.07）；以及 (3) 心血管疾病（ICD-9代码：410-414、427-428、433-437、440、443-445、451-453）。

我们考察了四种交通相关空气污染物：空气动力学直径为 < 2.5 微米的颗粒物（$PM_{2.5}$）、一氧化碳（CO）、氮氧化物（NO_x）以及元素碳（EC），后者是$PM_{2.5}$的组成成分。我们分别针对每种污染物，利用SHEDS模型（Burke 等，2001）获得个人暴露的人群层面分布。该模型是一种随机模拟器，可在人口普查区¨层面生成每日个人暴露数据（Ozkaynak 等，1996年；Jenkins，1996年；Dionisio 等，2013年）。为估算空气污染物室外浓度对应的个人暴露分布，SHEDS模型首先针对每个人口普查区，模拟多名虚拟个体的暴露情况；这些个体的人口统计学特征（例如年龄、性别和工作地点）反映了风险人群的特征，并以人口普查数据为依据。通过根据虚拟个体的人口统计学特征，从美国环境保护署（EPA）的综合人类活动数据库中随机分配活动日志，确定每名虚拟个体在不同微环境中的停留时间。

根据其人口统计学特征。随后，将这十三种微环境中的时间加权平均暴露量相加，计算其每日个人暴露量。这些微环境分为四类：室外、车辆、住宅室内和非住宅室内微环境。对于本分析，我们将人口普查区层面的个人暴露分布汇总至邮政编码制表区（ZCTA）层面。

在健康分析中，我们从 Daymet 获取了若干气象变量，以控制气象因素可能造成的混杂（Thornton et al., 2016）。这些变量包括每日最低气温、最高气温和露点温度。

3 方法

3.1 计数结局和暴露均值模型

我们首先介绍用于汇总结局的传统对数线性模型。令 y_i 表示组 $i = 1, \ldots, n$ 观测到的事件数（例如急诊就诊、住院、死亡），令 $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{im_i})^T$ 表示从组 i 的 m_i 名个体收集的暴露向量（例如个人空气污染暴露）。组可以按地理区域或时间段划分。在大多数应用中，组特异性暴露的均值用于过度离散泊松对数线性回归模型，例如：

$$\begin{aligned} Y_i|\lambda_i \sim Poisson(\lambda_i), \quad \lambda_i|\xi, \eta_i \sim Gamma(\xi, \exp(\eta_i)), \text{ and} \\ \eta_i = \alpha\mu_i + \boldsymbol{\gamma}^T \boldsymbol{Z}_i + \epsilon_i, \end{aligned} \tag{1}$$

其中，μ_i 表示组 i 的真实平均暴露，通常用样本均值估计$\sum_{j=1}^{m_i} x_{ij}/m_i$；参数 α 表示平均暴露的效应；$\boldsymbol{Z}_i$ 是一个包含其他协变量的向量，其回归系数为 $\boldsymbol{\gamma}$（包括截距）；ξ 控制过度离散的程度，

而 ϵ_i 表示一个均值为零的空间/时间残差过程。采用这种参数化方式， Y_i 期望值的对数等于 $\log(\xi) + \eta_i$。模型（1）仅描述预期健康结局计数与全人群平均暴露之间的关联，忽略了个体间的暴露异质性。因此，模型（1）可能不足以作为刻画暴露对健康影响的框架。例如，模型（1）无法捕捉暴露分布发生变化但全人群平均值相同所关联的健康效应。

3.2 计数结局与暴露分布模型

为捕捉整个暴露分布的效应，并允许效应随分位数水平变化，我们提出一种将暴露分位数函数作为函数型协变量的模型。所提出的标量对分位数函数过度离散泊松对数线性回归模型如下：

$$\begin{gathered} Y_i|\lambda_i \sim Poisson(\lambda_i), \quad \lambda_i|\xi,\eta_i \sim Gamma(\xi, \exp(\eta_i)), \text{ and} \\ \eta_i = \int_0^1 \beta(\tau) Q_i(\tau) d\tau + \boldsymbol{\gamma}^T \boldsymbol{Z}_i + \epsilon_i, \end{gathered} \tag{2}$$

其中， $Q_i(\cdot)$ 表示组 i 的连续暴露的暴露分位数函数。请注意， $\beta(\tau)$ 表示暴露的第 τ 个百分位数对健康结局均值的效应；其他参数已在模型（1）中说明。

为灵活刻画健康结局与暴露之间的关联，假设系数 $\beta(\tau)$ 是分位数水平的平滑函数，并通过有限个基函数进行建模。具体而言， $\beta(\tau)$ 定义为：

$$\beta(\tau) = \sum_{j=0}^{p} K_{j,p}(\tau)\beta_j, \tag{3}$$

其中 $K_{j,p}=\left(\sqrt{2(p-j)+1}\right)(1-\tau)^{p-j}\sum_{k=1}^{j}(-1)^k\binom{2p+1-k}{j-k}\binom{j}{k}\tau^{j-k}$ 表示定义在区间 $[0,1]$ 上的 p 次正交归一伯恩斯坦多项式（Bellucci，2014）。通过这种基函数展开，式（2）中的所提出的模型可写为：

$$\eta_i=\boldsymbol{\beta}^T\int_0^1\boldsymbol{K}(\tau)Q_i(\tau)d\tau+\boldsymbol{\gamma}^T\boldsymbol{Z}_i+\epsilon_i, \tag{4}$$

其中 $\boldsymbol{\beta}=(\beta_0,\ \ldots\ldots,\ \beta_p)^T$ 是基函数系数向量，$\boldsymbol{K}(\tau)=(K_{0,\ p}(\tau)\ldots\ldots,\ K_{p,\ p}(\tau))^T$ 是基函数向量。值得注意的是，伯恩斯坦多项式的定义域与暴露分位数函数的定义域一致，与 B 样条或高斯核相比，这可能有助于估计 $\boldsymbol{\beta}$。选择伯恩斯坦多项式也因为已有研究表明，只需少量基函数，它们便能准确近似各种光滑函数形式（Bellucci，2014）。我们还注意到一种特殊情况，即暴露在不同分位数水平上的效应相同（即 $\beta(\tau)$ 是 τ 的常数）。此时，所提出的模型（2）将简化为模型（1），因为 $\int_0^1 Q_i(\tau)d\tau=\mu_i$。

3.3 估计与推断

3.3.1 分位数函数已知

已知分位数函数时，式（4）可重新参数化为一个带协变量向量 $\boldsymbol{X}_i^*=\int_0^1\boldsymbol{K}(\tau)Q_i(\tau)d\tau$ 的常规标量对标量模型。例如，在 Reich 等人（2009）的研究中，$Q_i(\tau)$ 被假定为正态分位数函数。通过引入潜在的 Pólya-Gamma 随机变量（Polson 等人，2013年；Neelon，2019），可对过度离散的泊松回归进行高效的全贝叶斯推断。利用 Pólya-Gamma 方法，可通过吉布斯抽样使用马尔可夫链蒙特卡罗（MCMC）估计系数 $\boldsymbol{\beta}$ 和 $\boldsymbol{\gamma}$。对系数 $\boldsymbol{\beta}$ 和引入独立正态先验后

γ，这些回归系数的全条件分布为多元正态分布。

3.3.2 分位数函数未知

对于分位数函数未知这一更现实的情形，我们提出一种两阶段贝叶斯估计程序。第一阶段再次使用基函数展开对每个组 i 的分位数函数 $Q_i(\tau)$ 建模，并根据个体层面暴露（例如，我们应用中的 SHEDS 模拟）对其进行估计。第二阶段，在考虑与第一阶段估计相关的统计不确定性的同时，使用估计的分位数函数拟合健康模型 (2)。

借鉴以往针对连续变量分位数过程建模的半参数贝叶斯方法（Reich，2012年），组 i 中暴露的分位数函数展开如下：

$$Q_i(\tau) = \theta_{0,i} + \sum_{l=1}^{L} B_l(\tau)\theta_{l,i}, \tag{5}$$

其中，$B_l(\tau)$ 是第 l 个基函数，$\theta_{l,i}$ 是基函数系数。基函数 $B_l(\tau)$ 的选择包括分段高斯函数或分段伽马函数，其表达式见补充材料。这两种选择均可灵活地刻画可能存在偏斜的暴露分布。然而，对于严格为正的暴露（例如环境空气污染浓度），建议使用分段伽马函数。使用分段高斯函数或伽马函数时，$\theta_{0,i}$ 表示第 i 组暴露分布的中位数，而 $\theta_{l,i}$ 对于 $l = 1\ldots,$ L 刻画该分布的形状。分位数函数唯一确定密度。令 x_{ij} 表示第 i 组中第 j 个个体测得的暴露水平。当假设 x_{ij} 服从与分位数函数 $Q_i(\tau)$ 相对应的分布时，似然

一组个体层面暴露的似然为

$$L(\boldsymbol{\theta}_{0,\cdot},\{\boldsymbol{\theta}_{l,\cdot}\}_{l=1}^{L};\{\boldsymbol{x}_i\}_{i=1}^{n})=\prod_{i=1}^{n}\prod_{j=1}^{m_i}\left[\frac{dQ_i(\tau)}{d\tau}\right]^{-1}\Bigg|_{\tau^*:Q_i(\tau^*)=x_{ij}}, \tag{6}$$

其中，$\boldsymbol{\theta}_{0,\cdot}=(\theta_{0,1},\ldots,\theta_{0,n})^T$，$\boldsymbol{\theta}_{l,\cdot}=(\theta_{l,1},\ldots,\theta_{l,n})^T$，且 $\boldsymbol{x}_i=(x_{i1},\ldots,x_{im_i})^T$是为组 i 中的 m_i 名个体收集的个人暴露向量。需要注意的是，分位数函数必须是非递减的。为确保这一性质，对于任意 i 和 l，$\theta_{l,i}>0$ 都应成立（Reich，2012年）。在第一阶段估计中，通过引入一个无约束的潜变量 $\theta^*_{l,i}$来施加这一约束。具体而言，$\theta_{l,i}=\max\{\theta^*_{l,i},\nu\}$，其中 ν 是一个较小的常数（例如 0.01）。对于空间数据或时间序列数据，可以在贝叶斯层次建模框架中轻松地为 $\boldsymbol{\theta}_{0,\cdot}$ 和 $\boldsymbol{\theta}^*_{l,\cdot}=(\theta^*_{l,1},\ldots,\theta^*_{l,n})^T$ 引入空间依赖或时间依赖，从而使分位数过程随时间或位置变化。所有模型参数均通过马尔可夫链蒙特卡罗算法进行估计。

在第二阶段，在考虑与估计暴露分位数函数相关的不确定性的情况下拟合健康模型。当使用基函数展开暴露分位数函数时，公式（4）中的健康模型变为：

$$\eta_i=\boldsymbol{\beta}^T\left[\int_0^1\boldsymbol{K}(\tau)\boldsymbol{B}(\tau)^T d\tau\right]\boldsymbol{\theta}_{\cdot,i}+\boldsymbol{\gamma}^T\boldsymbol{Z}_i+\epsilon_i, \tag{7}$$

其中 $\boldsymbol{B}(\tau)=(1,B_1(\tau).....,B_L(\tau))^T$以及 $\boldsymbol{\theta}_{\cdot,i}=(\theta_{0,i},\theta_{1,i},.....,\theta_{L,i})^T$。由于已预先指定 $\boldsymbol{B}(\tau)$，估计的 $\boldsymbol{\theta}_{\cdot,i}$的不确定性会导致分位数函数估计的不确定性。

为适当传递第一阶段估计所产生的不确定性，我们考虑采用环境健康研究中常用的一种方法，以纳入估计暴露中的不确定性（Carroll 等，2006年；Lee 等，2017年）。具体

而言，假设 $\boldsymbol{\theta}_{\cdot\cdot i}$ 服从多元正态（MVN）先验，其均值和方差-协方差矩阵根据第一阶段估计所获得的后验预测分布计算得到。与分位数函数已知的情形类似，我们将 $\boldsymbol{X}_i^* = \left[\int_0^1 \boldsymbol{K}(\tau)\boldsymbol{B}(\tau)^T d\tau\right]\boldsymbol{\theta}_{\cdot,i}$ 视为随机协变量向量。在假设$\boldsymbol{\theta}_{\cdot,i}$服从MVN先验的情况下，$\boldsymbol{X}_i^*$ 的先验同样服从MVN分布。值得注意的是，当假设分位数过程随组而变化时，$\boldsymbol{\theta}_{\cdot\cdot i}$ 的后验分布在各组之间存在相关性。随着组数增加，该方法的计算成本会变得很高，因为马尔可夫链蒙特卡罗算法需要从高维多元正态分布中进行抽样。在模拟研究和真实数据分析中，我们忽略组间相关性以便于计算。我们发现，这不会对健康效应关联的推断产生实质性影响，Comess 等（2022）近期的研究也得出了相同结论。与分位数函数已知的情形一样，系数 $\boldsymbol{\beta}$ 和 $\boldsymbol{\gamma}$ 的估计通过吉布斯抽样完成，并使用Pólya-Gamma 方法为这些系数指定正态先验。关于分位数函数的估计，以及在分位数函数已知和未知情况下进行估计的马尔可夫链蒙特卡罗算法的详细信息，参见补充材料。

4 模拟研究

我们开展了模拟研究，以考察未考虑暴露异质性的影响，这一研究动机源于我们的实际应用，并评估了两阶段估计程序的表现。通过指定不同形式的系数回归函数 $\beta(\tau)$，我们考虑了暴露与结局之间的多种关联。

模拟研究假设健康结局是在 $n = 1000$ 个时间点收集的，且暴露分位数函数在时间上相关。时间依

赖性是通过为无约束的潜在基函数系数定义一阶高斯马尔可夫随机场过程引入的。真实的健康模型和暴露分位数函数如下：

$$Y_i|\lambda_i \sim Poisson(\lambda_i), \quad \lambda_i|\xi,\eta_i \sim Gamma(\xi,\exp(\eta_i)),$$

$$\eta_i = \beta_0 + \int_0^1 \beta(\tau)Q_i(\tau)d\tau,\ i = 1,\ldots,1000,$$

$$Q_i(\tau) = \theta_{0,i} + \sum_{l=1}^{4} B_l(\tau)\theta_{l,i},$$

$$(\theta_{0,1},\ldots,\theta_{0,1000})^T \sim MVN(\theta_0,\Sigma_0),\ \Sigma_0 = \sigma_0^2(D_w - \rho_0 W)^{-1},$$

$$(\theta^*_{l,1},\ldots,\theta^*_{l,1000})^T \sim MVN(\theta_l,\Sigma_1)\ \ \Sigma_1 = \sigma_1^2(D_w - \rho_1 W)^{-1} \text{对 } l = 1,\ldots,4, \theta_{l,i} = \max\{\theta^*_{l,i}, \nu = 0.01\}\ ,$$

其中， $B_l(\tau)$ 是由 4 个基函数构成的分段伽马函数； W 是一个对称矩阵，其第 (i,i') 个元素在时间点 i 和 i' 相邻时取值为 1，否则为 0； D_w 是一个对角矩阵，其第 i个对角元素等于 W中第 i行的行和。对于每个时间点 i，使用分位数函数 $Q_i(\tau)$ 生成样本量为 100 的个体暴露数据。为模拟动机数据中观测到的 $PM_{2.5}$ 右偏分布，我们对 $l = 1,\ldots,4$ 设置 $\theta_0 = 7.2$, $\theta_l = 0.9$，并将控制分位数过程时间相关性的参数（$\sigma_0^2,\rho_0,\sigma_1^2,\rho_1$）设为 (1,0.9, 0.02,0.9)。

我们考察了图1所示的六个系数回归函数：(1) $\beta(\tau) = 0.5$, (2) $\beta(\tau) = \tau$, (3) $\beta(\tau) = 1.5\tau^2$, (4) $\beta(\tau) = \frac{4}{3}\tau I(\tau < 0.5) + \frac{2}{3}I(\tau \geq 0.5)$, (5) $\beta(\tau) = \exp(-\frac{\tau^2}{0.328})$, 以及 (6) $\beta(\tau) = -\tau + 1$。这六个函数涵盖三种效应类型（恒定、递增和递减）。我们注意到，在所有情景下， $\int_0^1 \beta(\tau)d\tau$ 均等于 0.5；该量可解释为暴露分布向右平移一个单位（即均值增加一个单位）时，暴露所对应的效应

）。基于模拟得到的分位数函数 $Q_i(\tau)$，在固定 $\beta_0 = -3.5$ 的情况下，针对 $\beta(\tau)$ 的不同形式分别生成了 100 个健康数据集。我们还拟合了传统“均值”模型，该模型将每个时间点的个体暴露均值作为协变量，以供比较。

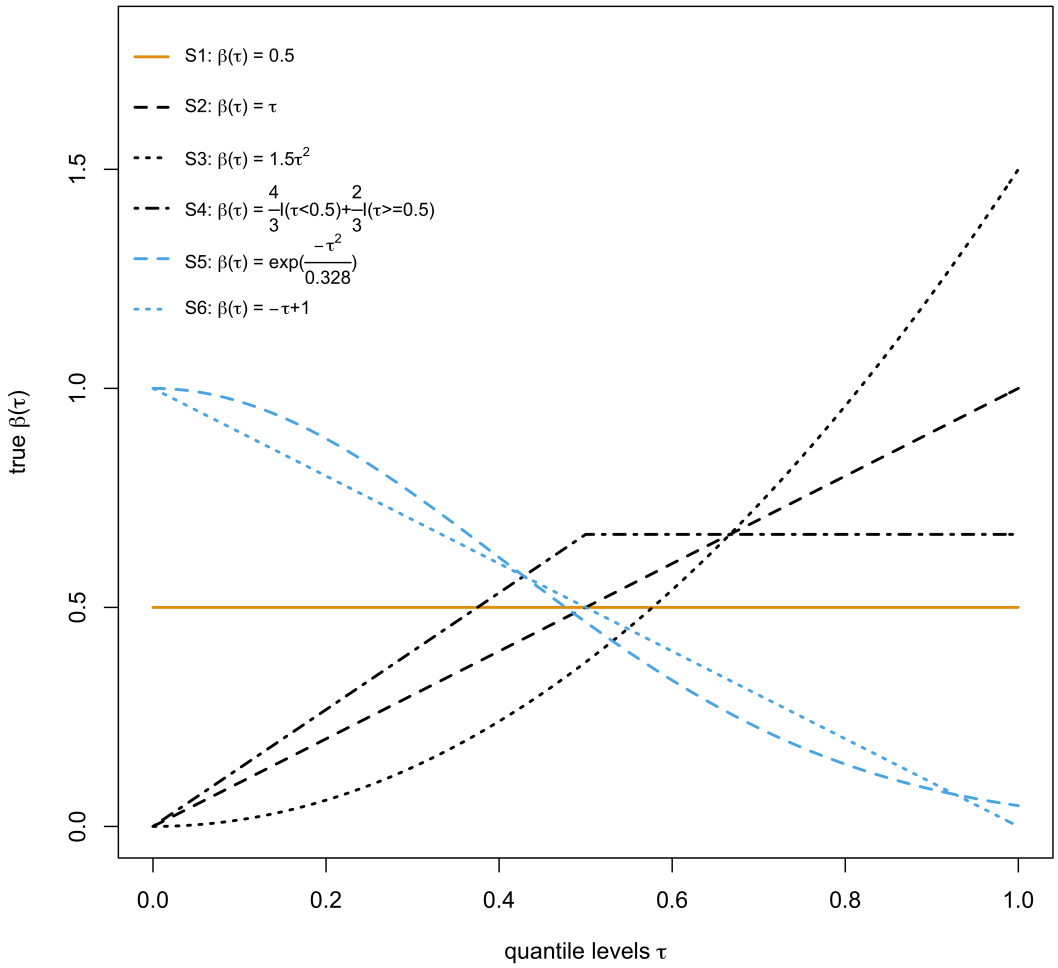

图1：T　模拟研究中指定的真实系数回归函数 $\beta(\tau)$

用于性能评估的三个量为：(a) 暴露分布向右平移一个单位所对应的效应；对于所提出的模型，该效应用 $\int \beta(\tau)d\tau$ 衡量，对于均值模型则用参数 α 衡量；(b) 暴露的预测值；对于所提出的模型和均值模型，分别用 $\int_0^1 \beta(\tau)Q_i(\tau)d\tau$ 和 $\alpha\mu_i$ 衡量，

；(c) 暴露归因事件总数；对于所提出的模型和均值模型，分别计算为 $\sum_{i=1}^{n}\xi\{\exp\left[\beta_0+\int_0^1\beta(\tau)Q_i(\tau)d\tau\right]-\exp(\beta_0)\}$ 和 $\sum_{i=1}^{n}\xi\left[\exp(\beta_0+\alpha\mu_i)-\exp(\beta_0)\right]$。对于这些量，计算了其相对偏倚、均方误差（MSE）和 95% 覆盖概率（CP）。

对于 $\int_0^1\beta(\tau)d\tau$ 和暴露归因事件总数，其相对偏倚、均方误差和覆盖概率通过对模拟结果取平均计算。然而，对于预测值，我们同时对模拟和时间点取平均。具体而言，当估计暴露分位数函数时，暴露预测值的均方误差计算如下：

$$\frac{1}{n\times D}\sum_{i=1}^{n}\sum_{d=1}^{D}\Big[\int_0^1\hat{\beta}^{(d)}(\tau)\hat{Q}_i^{(d)}(\tau)d\tau-\int_0^1\beta(\tau)Q_i(\tau)d\tau\Big]^2 \tag{8}$$

对于所提出的模型，其中下标 d 表示满足 $n=1000$ 和 $D=100$ 的模拟编号；$\hat{}\,\beta^{(d)}(\tau)=\hat{}\,\boldsymbol{\beta}^{(d)^T}\boldsymbol{K}(\tau)$，$\hat{}\,\boldsymbol{\beta}^{(d)}$ 是式（4）中定义的估计基函数系数向量；$\hat{}\,Q_i(\tau)^{(d)}=\boldsymbol{B}(\tau)^T\hat{}\,\boldsymbol{\theta}^{(d)\cdot i}$，$\hat{}\,\boldsymbol{\theta}^{(d)\cdot i}$ 是式（7）中定义的估计基函数系数。相对偏倚和覆盖概率的计算方法类似。

此外，通过对模拟和分位数水平 $\tau_j\in\{0,0.01,\ldots,0.99,1\}$ 取平均，考察了 $\beta(\tau)$ 的估计性能，包括其偏差、均方误差和覆盖概率。例如，$\beta(\tau)$ 的均方误差为 $1/(J\times D)\sum_{j=1}^{J}\sum_{d=1}^{D}\left[\hat{}\,\boldsymbol{\beta}^{(d)^T}\boldsymbol{K}(\tau_j)-\beta(\tau_j)\right]^2$。由于 $\beta(\tau)$ 可能取值为 0，因此未计算相对偏倚。

在这项模拟研究中，我们使用二次正交归一伯恩斯坦多项式（即三个基函数）对 $\beta(\tau)$建模，并为相应的基函数系数指定宽松正态先验 N(0,100)。健康模型中的过度离散参数采用均匀先验，并通过 Metropolis–Hastings（MH）算法更新。为估计暴露分位数函数，我们使用四个分段伽马函数展开分位数函数。系数 $\theta_{0,\,i}$ 和 $\theta^*_{l,\,i}$ 被赋予 N(0, 100) 先验，并且

通过 MH 算法更新； σ_0^2 和 σ_1^2 被赋予逆伽马分布 InvGamma(0.1, 0.1) 先验，并通过吉布斯抽样更新；为 ρ_0 和 ρ_1 指定了离散先验（即 0 到 1 之间等间距的 1000 个值），并通过 MH 算法更新。在估计分位数函数时，我们生成了 10,000 个 MCMC样本，并将前 5,000 个样本作为预热期丢弃；在估计健康效应时，无论使用真实的还是估计的分位数函数，我们都生成了 5,000 个样本，并将前 2,500 个样本作为预热期丢弃。

使用不同暴露协变量（真实平均暴露、真实暴露分位数函数和估计的分位数函数）所得的模拟结果汇总于表1。我们首先关注假定暴露分位数函数和平均暴露已知的情形。在 $\beta(\tau)$ 为常数的情景（即 S1）下，使用暴露分位数函数和均值所得的结果相似。由于 $\beta(\tau)$ 采用基函数展开进行建模，所提出的模型呈现出较大的均方误差（MSE）。在其他情景下，所提出的模型对健康效应的估计在经验上无偏。然而，使用平均暴露时，我们观察到健康效应估计存在偏倚。具体而言，当暴露分布的上尾部分效应较大时（如 S2 和 S3），我们发现正偏倚；当效应随分位数水平升高而降低时（如 S5 和 S6），则发现负偏倚。对于归因于暴露的事件总数和暴露的预测值，观察到的模式恰好相反。在健康效应先增大后保持不变的 S4 中，使用平均暴露的模型恰好得到了近乎无偏的估计，且 $\int_0^1 \beta(\tau)\, d\tau$的95%可信区间（CI）较为保守。然而，正如暴露预测值的较大均方误差和严重覆盖不足所示，该模型未能刻画暴露的效应。还值得注意的是，所提出的模型在这一情景下的表现略差，例如，均方误差

$\int_0^1 \beta(\tau)d\tau$的均方误差高于 S2、S3 和 S5，这可能是因为基函数展开对 S4 中的 $\beta(\tau)$ 近似得不够好。

表1：使用不同暴露协变量的模拟结果

情景	协变量 [a]		$\int_0^1 \beta(\tau)d\tau$ [b]			$\beta(\tau)$			预测值的暴露			暴露归因的事件		
			相对 bias	相对均方误差 [c]	CP (%)	bias	MSE	CP (%)	相对 bias	相对 MSE[c]	CP (%)	相对 bias	相对 MSE[c]	CP (%)
S1		mean	0.004	1.00	94	-	-	-	0.004	1.000	94.00	0.001	1.000	95
		分位数	0.005	1.10	96	0.002	0.016	93.70	-0.002	3.889	93.39	0.000	1.069	95
		分位数含误差	0.011	1.47	92	0.004	0.031	95.48	0.004	6.475	96.41	0.001	1.101	94
S2		mean	0.009	1.00	89	-	-	-	-0.122	1.000	0.01	-0.008	1.000	83
		分位数	-0.002	0.80	96	-0.001	0.015	92.17	-0.002	0.052	93.66	-0.000	0.610	93
		分位数含误差	0.010	1.07	91	0.003	0.032	93.13	0.010	0.108	95.50	0.000	0.620	93
S3		mean	0.019	1.00	75	-	-	-	-0.174	1.000	0.00	-0.010	1.000	82
		分位数	-0.001	0.39	97	-0.000	0.013	92.92	-0.003	0.023	91.74	0.000	0.408	96
		分位数带有误差	0.014	0.74	90	0.005	0.036	90.47	0.017	0.056	94.12	0.001	0.398	96
S4		mean	0.003	1.00	100	-	-	-	-0.081	1.000	0.04	-0.006	1.000	89
		分位数	-0.001	1.01	98	-0.001	0.014	93.61	0.003	0.127	93.56	0.000	0.751	95
		分位数带有误差	0.008	1.38	98	0.001	0.039	90.57	0.015	0.302	94.82	0.001	0.759	94
S5		mean	-0.010	1.00	95	-	-	-	0.182	1.000	0.00	0.017	1.000	77
		分位数	0.002	0.80	95	0.001	0.018	95.90	-0.003	0.057	95.73	0.000	0.361	95
		分位数含误差	0.003	0.85	97	0.000	0.033	97.12	-0.002	0.107	96.02	-0.000	0.381	95
S6		mean	-0.011	1.00	94	-	-	-	0.163	1.000	0.02	0.014	1.000	80
		分位数	0.000	0.75	95	0.000	0.016	95.74	-0.001	0.075	95.67	-0.001	0.537	95
		分位数存在误差	0.002	0.83	96	0.000	0.042	94.42	-0.003	0.127	95.89	-0.002	0.586	93

[a] 均值 = 真实平均暴露，分位数 =真实暴露分位数函数，含误差的分位数 = 估计的暴露分位数函数。
[b] 对于使用平均暴露作为暴露协变量的模型，报告了参数 α。[c] 将使用真实平均暴露的模型作为参照，计算相对均方误差。

我们还根据广泛可用信息准则（WAIC）评估了选择所提出的模型而非传统“平均”暴露模型的决策（?）。在 100 次模拟中，WAIC 判定

均值模型更优的模拟比例为：在 S1 情景下为 82%，此时均值模型是有效且更好的选择。而在其他情景下，WAIC 超过 95% 的情况下判定所提出的模型更优。

当从包含个体层面暴露的模拟数据集中估计暴露分位数函数时，我们采用两阶段估计程序，以处理估计暴露分位数函数时的不确定性。图2(a)展示了连续九天的模拟暴露数据，其中暴露分布呈右偏，并随时间平缓变化。如图2(b)所示，估计的分位数函数很好地描绘了真实情况。与使用真实分位数函数的情形相比，这四个量的均方误差有所增加（表1）。然而，在所有情景下，覆盖概率仍达到或接近名义水平。我们发现，若不通过使用估计的分位数函数的后验均值来考虑估计不确定性，则可能导致偏差和覆盖不足（结果未显示）。

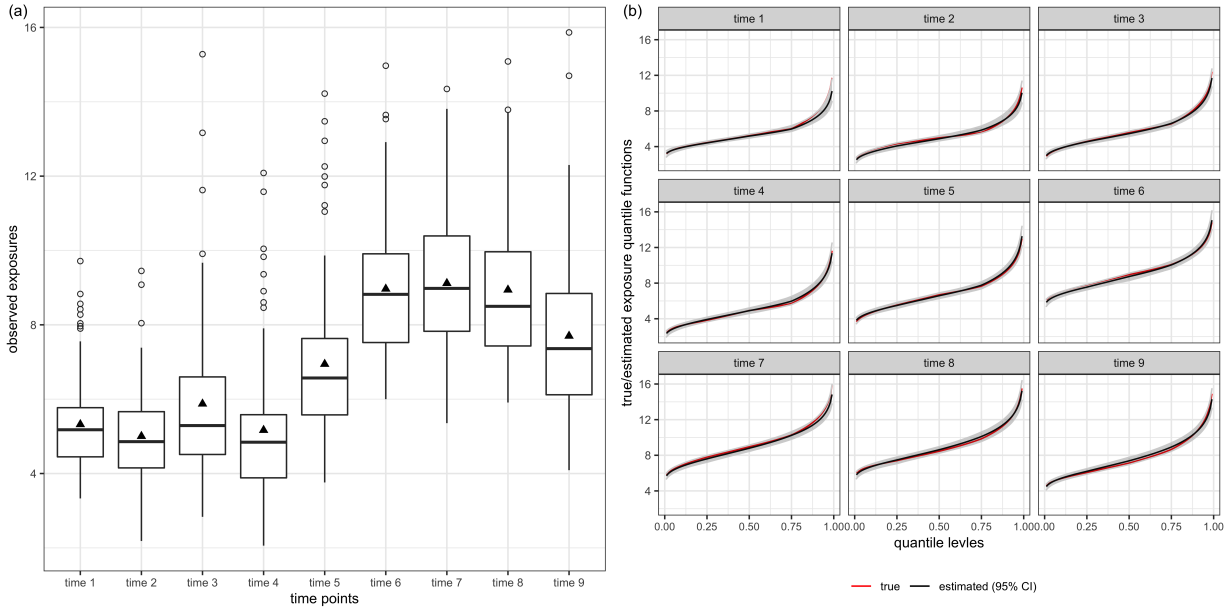

图2：（a）模拟数据中连续九天观测到的暴露箱线图；（b）连续九天的真实/估计暴露分位数函数

5 真实数据分析

本节中，我们使用第2节介绍的动机数据，对所提出的标量对分位数函数模型和传统模型进行了分析；传统模型将空气污染物的平均浓度作为暴露。

5.1 每日暴露分位数函数的估计

图 3(a)–(d) 展示了 SHEDS 中一个具有代表性的邮政编码制表区内四种空气污染物的分布。我们观察到，所有四种空气污染物的暴露分布都呈现不同程度的偏斜。我们注意到，与 NO_x 和 $PM_{2.5}$ 相比，CO 和 EC 的分布在不同日期之间差异很大。因此，在估计分位数函数时，假设 CO 和 EC 的分位数函数在不同日期和 ZCTA 之间相互独立，而对 NO_x 和 $PM_{2.5}$ 的分位数函数则引入了时间相关性。在本分析中，我们使用了与第 4 节相同的先验、马尔可夫链蒙特卡罗迭代次数和预热期。图 3(e)–(h) 展示了使用四个分段伽马函数得到的相应经验分位数和估计的分位数函数。总体而言，正如预期的那样，我们使用基函数充分捕捉了暴露分布，并且在分布的低端和高端存在较大的不确定性。

5.2 ED 就诊健康效应的估计

使用结合估计的分位数函数和不确定性传播的所提出的模型，以及在邮政编码制表区层面使用个体暴露样本均值的传统模型，考察了急诊就诊与当日空气污染浓度之间的关联。按照此前的健康分析（Sarnat 等，2013），所有模型均控制了以下混杂因素：使用日……表示的特定年份时间趋势的非线性效应

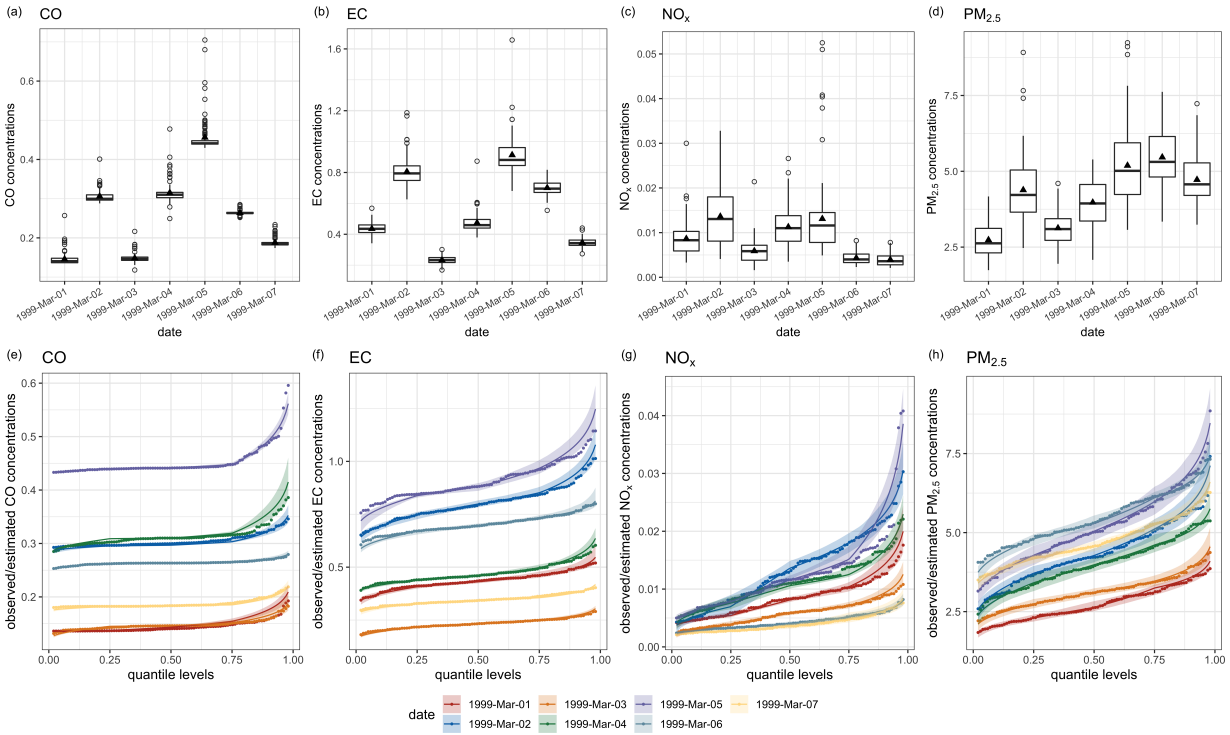

图3：（a）–（d）七个代表性日期在 ZCTA 30032 通过 SHEDS 获得的空气污染物浓度箱线图；（e）–（h）七个代表性日期在 ZCTA 30032 通过 SHEDS 获得的空气污染物经验/估计分位数函数及其 95%可信区间（经验分位数函数以实心点表示）。

年内效应、当日露点温度的非线性效应、星期几指标、联邦假日指标，以及3日移动平均值最低气温的非线性效应（心血管疾病急诊就诊模型中控制了最高气温）。所有非线性效应均采用具有4个自由度的自然三次样条建模。所有模型均纳入可交换的邮政编码制表区特异性随机截距。我们将用于建模 $\beta(\tau)$ 的基函数数量从两个变为三个，并报告广义贝叶斯信息准则较低的模型结果。基函数系数和过度离散参数采用与第4节相同的先验；空间随机效应的方差参数假设服从逆伽马分布(0.1, 0.1)先验。我们共运行7,500次迭代，前3,500次作为预热期舍弃。

图4绘制了不同空气污染物与急诊就诊原因组合下估计的 $\beta(\tau)$ 及其95%置信区间。我们观察到，在某些情况下（例如图4的(b)、(c)、(e)、(f)和(h)面板），广义贝叶斯信息准则明显倾向于所提出的模型。例如，图4(e)显示，元素碳浓度分布的较低和较高尾部对呼吸系统疾病急诊就诊的影响更大。元素碳浓度分布向右平移一个单位所对应的风险百分比增幅，采用所提出的模型和传统模型估计分别为2.58（95%置信区间：0.82–4.31）和0.56（95%置信区间：-0.14–1.26）。在某些情况下，广义贝叶斯信息准则表明，采用分位数函数的模型与采用暴露样本均值的模型对数据的拟合效果相当。例如，图4(j)显示，$PM_{2.5}$ 对哮喘或喘鸣急诊就诊的健康效应估计在不同分位数水平下基本保持不变，与均值模型得出的估计一致。其余组合的风险百分比增幅估计见表S1。

图 5(a)-　　(d) 展示了相应暴露因素的估计的分位数函数，其中

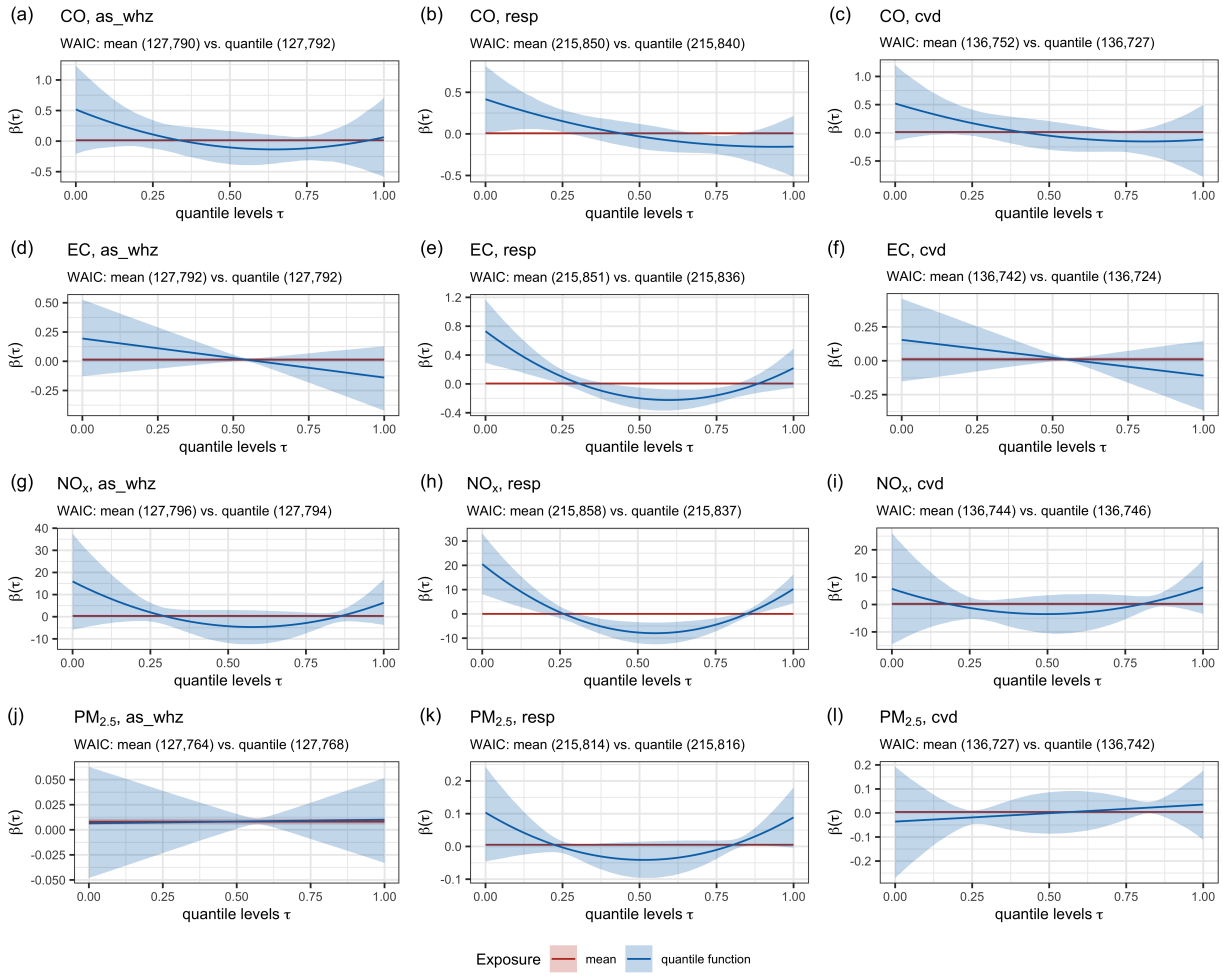

图4：分析动机数据所得的 $\beta(\tau)$ 估计值及其 95%可信区间（as whz = 哮喘或喘鸣急诊就诊, resp = 呼吸系统疾病急诊就诊，cvd =心血管疾病急诊就诊）。

确切地说，中位数分别位于所有 ZCTA 和日期的估计中位数分布的第 25th、 50th、 75th 和 95th 百分位数。虚线表示根据用于估计分位数函数的 SHEDS 数据计算得出的平均暴露。这些分位数函数经过选择，用于代表不同 ZCTA 和日期之间的暴露对比。图 5(e)-(h) 展示了采用所提出模型和传统模型时，与选定的估计分位数函数变化相关的相对风险估计值。我们选择了两种对比来代表不同的暴露效应：(1) 比较中位数位于第75^{th} 和第 25th 百分位数的分位数函数，以代表典型的暴露对比；(2) 比较中位数位于第 95th 和第 50th 百分位数的分位数函数，以代表更极端的暴露效应。

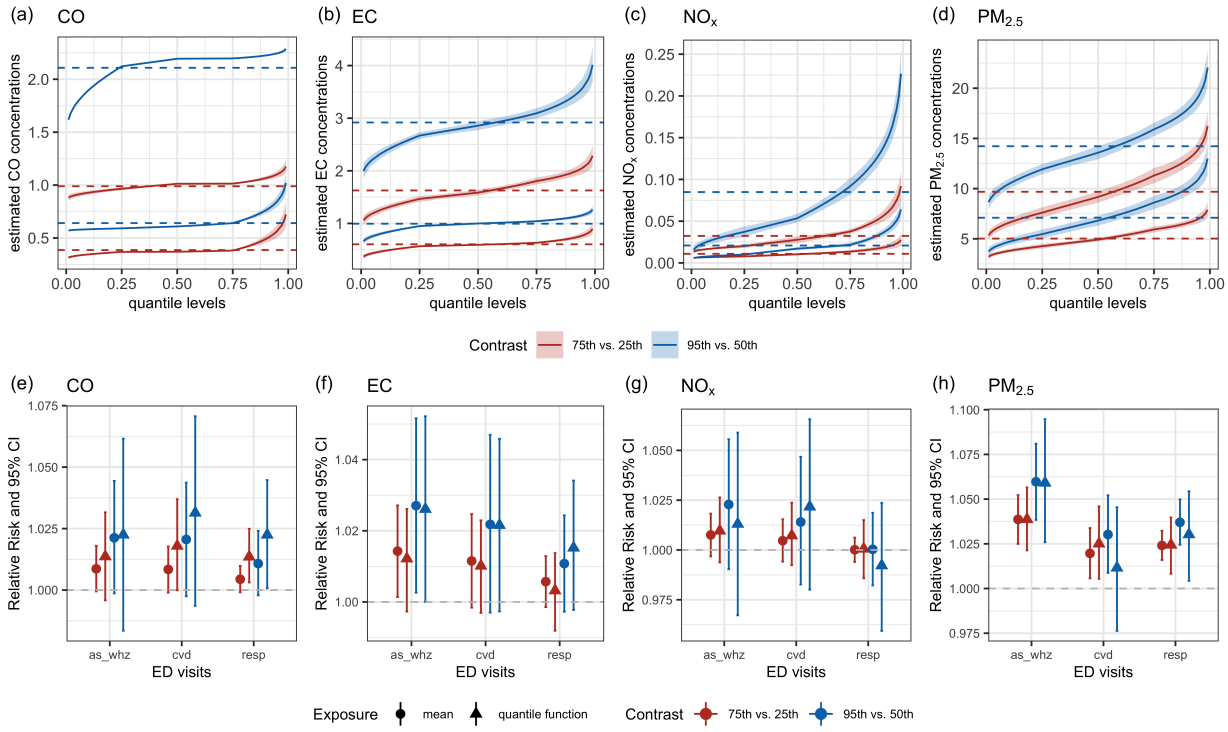

图 5：与两种观测暴露分布相关的急诊就诊短期关联，这两种分布由暴露中位数定义（分别比较75^{th} 与 25th、以及 95th 与 50th，即所有 ZCTA 和日期的估计中位数分布中的百分位数）。同时展示了将平均浓度作为暴露的模型所得结果。as whz= 哮喘或喘鸣，resp = 呼吸系统疾病，cvd = 心血管疾病。

对于短期　一氧化碳室外浓度与呼吸系统急诊就诊之间的关联

以及心血管疾病急诊就诊之间的关联，如图4(b)和(c)所示，包含分位数函数的所提出的模型表现更佳；我们发现，一氧化碳在较低分位数水平上的效应更为显著。两个选定对比对应的短期关联见图5(e)。我们观察到，与考虑整个暴露分布的所提出的模型相比，在健康模型中仅将一氧化碳平均浓度作为协变量，会低估一氧化碳对心血管疾病急诊就诊和呼吸系统疾病急诊就诊的相对风险。暴露于空气污染物所致急诊就诊估计数量也存在差异。例如，图6(a)-(b)展示了当空气污染物的效应随其分位数水平变化时，各邮政编码制表区暴露归因急诊就诊数量的差异。具体而言，采用所提出的模型估计，一氧化碳暴露归因的心血管疾病急诊就诊总数为798（95%置信区间：74–3,240）；而均值模型的估计值为640（95%置信区间：-80–1,343）。相比之下，对于$PM_{2.5}$以及哮喘或喘鸣急诊就诊，如果健康效应不随分位数水平变化，所提出的模型与传统模型得出的结果相近（图6(c)）。

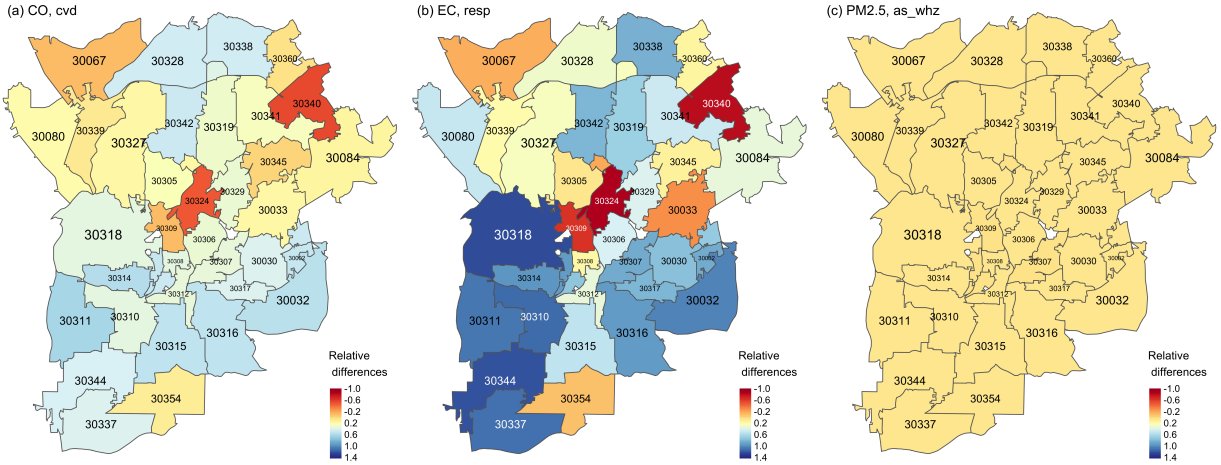

图6：采用所提出的暴露分位数模型与传统平均暴露模型时，邮政编码制表区层面的暴露归因急诊就诊数量相对差异（whz = 哮喘或喘鸣急诊就诊，resp = 呼吸系统疾病急诊就诊，cvd =心血管疾病急诊就诊）。

6 讨论

在本研究中，我们提出了一种标量对分位数函数方法，将暴露分位数函数视为函数型协变量，从而全面刻画环境暴露对汇总健康结局的影响。与仅将个人暴露的汇总统计量作为所关注暴露指标的方法相比，我们的方法能够考虑组内暴露异质性，并允许暴露与汇总健康结局之间存在更灵活的关联。此外，我们的方法不需要对暴露分布作出参数分布假设。借助所提出的贝叶斯两阶段估计程序，可以在估计暴露分位数函数时纳入不确定性，并获得健康效应估计值。与联合建模暴露和健康数据的方法相比，当考察一种暴露与多个健康结局之间的关联时，这一两阶段程序还可减轻计算负担。

将所提出的模型应用于亚特兰大的急诊就诊数据后，我们发现了环境空气污染浓度与急诊就诊之间新颖的短期关联；当每日人口平均暴露与急诊就诊进行关联时，这些关联会被掩盖。例如，结果表明，室外浓度一氧化碳对呼吸系统疾病急诊就诊和心血管疾病急诊就诊的影响，在较低分位数水平下最大。这些新发现可能有助于识别最易受到环境空气污染影响的亚群体。对于大多数污染物与结局组合，我们发现了稳健且呈正向的关联，其结果与传统方法相当，从而进一步支持了空气污染负面健康效应的既有证据。

所提出的模型受到 SHEDS 数据的启发，这些数据提供了亚特兰大四年间不同空气污染物的模拟个人暴露。尽管个人暴

露数据在大规模的基于人群的流行病学研究中并不普遍可得，但所提出的模型可应用于更一般的情境。具体而言，当组内存在暴露异质性且可用分位数函数加以表征时，所提出的方法便适用。例如，在利用时空模型以比健康数据空间分辨率更精细的空间分辨率预测环境暴露的情景下，我们可以结合根据预测暴露和人口密度得出的暴露分位数函数，应用所提出的方法。此外，健康结局不必局限于汇总计数。例如，在 Berrocal 等人（2011）开展的北卡罗来纳州出生体重数据分析中，可以使用我们的方法评估个体暴露分布的影响。

本研究聚焦于单一环境暴露。所提出的方法还可进一步扩展，以同时考察多种暴露对汇总健康结局的影响。一种可能的策略是将暴露的分位数曲面作为函数型预测变量纳入健康模型，从而考虑暴露之间的相关性。在真实数据分析中，我们发现，暴露分位数函数的估计对异常值较为敏感。因此，在分布尾部观察到估计的分位数函数存在较大的不确定性，健康效应的估计也可能受到影响。为了减轻异常值对暴露分位数估计的影响，可以考虑采用参数方法来刻画暴露分布的尾部（Zhou 等人，2012年）。

7 致谢

感谢美国环境保护署的 Lisa Baxter 和 Kathie Dionisio 提供 SHEDS 暴露数据。

参考文献

Alhanti, B. A., Chang, H. H., Winquist, A., Mulholland, J. A., Darrow, L. A., and Sarnat, S. E. (2016). 环境空气污染与哮喘急诊就诊：一项评估年龄效应修饰的多城市研究。《暴露科学与环境流行病学杂志》, 26(2):180–188.

Bekkar, B., Pacheco, S., Basu, R., and DeNicola, N. (2020). 美国空气污染和热暴露与早产、低出生体重及死产的关联：一项系统综述。《*JAMA Network Open*》, 3(6):e208243–e208243.

Bellucci, M. A. (2014). 关于正交伯恩斯坦多项式的显式表示。*arXiv*预印本 *arXiv:1404.2293*。

Berrocal, V. J., Gelfand, A. E., Holland, D. M., Burke, J., and Miranda, M. L.（2011）。利用 pm2. 5 暴露模拟器解释出生体重。《Environmetrics》， 22(4):553–571。

Bhaskaran, K., Gasparrini, A., Hajat, S., Smeeth, L., and Armstrong, B.（2013年）。环境流行病学中的时间序列回归研究。《国际流行病学杂志》， 42(4):1187–1195。

Boogaard, H., Walker, K., 和 Cohen, A. J. (2019). 空气污染：一种主要全球健康风险因素的兴起。《*International Health*》, 11(6):417–421.

Burke, J. M., Zufall, M. J., 和 OeZKAYNAK, H. (2001年). 颗粒物人群暴露模型：美国宾夕法尼亚州费城 PM2.5 案例研究结果。《暴露科学与环境流行病学杂志》, 11(6):470–489.

Calder, C. A., Holloman, C. H., Bortnick, S. M., Strauss, W., 和 Morara, M.（2008年）。利用暴露模拟器研究环境颗粒物浓度水平与死亡率之间的关系。《美国统计学会会刊》，103(481)：137–148。

Carroll, R. J., Ruppert, D., Stefanski, L. A., 和 Crainiceanu, C. M.（2006年）。非线性模型中的测量误差：现代视角。Chapman and Hall/CRC。

Chang, H. H., Fuentes, M., 和 Frey, H. C.（2012年）。利用暴露模拟器分析个人暴露于环境空气污染的时间序列及其与死亡率的关系。《暴露科学与环境流行病学杂志》，22(5)：483–488。

Comess, S., Chang, H.H., 和 Warren, J.L.（2022年）。将暴露不确定性纳入健康分析的贝叶斯框架：在空气污染与死产研究中的应用。*arXiv* 预印本 *arXiv:2203.16627*。

Dionisio, K. L., Isakov, V., Baxter, L. K., Sarnat, J. A., Sarnat, S. E., Burke, J., Rosen- baum, A., Graham, S. E., Cook, R., Mulholland, J., 等（2013年）。在美国佐治亚州亚特兰大开发并评估多种空气污染物暴露评估的替代方法。《暴露科学与环境流行病学杂志》，23(6):581– 592。

Dominici, F., Zeger, S. L., and Samet, J. M.（2000年）。针对空气污染与死亡率时间序列研究的测量误差模型。生物统计学，1(2):157–175。

Goldman, G. T., Mulholland, J. A., Russell, A. G., Srivastava, A., Strickland, M. J., Klein,

M., Waller, L. A., Tolbert, P. E., and Edgerton, E. S. (2010年)。亚特兰大时间序列流行病学研究中的环境空气污染物测量误差：特征及影响。《环境科学与技术》，44(19):7692–7698。

Guo, Y., Gasparrini, A., Armstrong, B. G., Tawatsupa, B., Tobias, A., Lavigne, E., Coelho, M. d. S. Z. S., Pan, X., Kim, H., Hashizume, M., 等（2017年）。热浪与死亡率：一项多国家、多社区研究。《环境健康展望》，125(8):087006。

Huang, G., Lee, D., and Scott, E. M.（2018年）。考虑暴露不确定性的多种空气污染物及其健康效应的多变量时空建模。《医学统计学》，37(7):1134–1148。

Jenkins, P.（1996年）。个人对空气颗粒物和金属的暴露：河滨市（加利福尼亚州）颗粒物研究团队研究的结果。《暴露分析与环境流行病学杂志》，6(1):57。

Jerrett, M., Arain, A., Kanaroglou, P., Beckerman, B., Potoglou, D., Sahsuvaroglu, T., Morrison, J., and Giovis, C.（2005年）。城市内部空气污染暴露模型综述与评估。《暴露科学与环境流行病学杂志》，15(2):185–204。

Jerrett, M.、Burnett, R. T.、Kanaroglou, P.、Eyles, J.、Finkelstein, N.、Giovis, C. 和 Brook, J. R.（2001年）。加拿大汉密尔顿的GIS环境正义分析：颗粒物空气污染。《环境与规划*A*》，33(6)：955–973。

Landrigan, P. J.（2017年）。空气污染与健康。《柳叶刀·公共卫生》，2(1)：e4–e5。
Lee, D., Mukhopadhyay, S., Rushworth, A., and Sahu, S. K.（2017年）。严谨的统计

框架，用于时空污染预测及其对健康的长期影响估计。生物统计学，18(2)：370–385。

Leiva, V., Barros, M., Paula, G. A., and Sanhueza, A.（2008年）。广义 Birnbaum–Saunders 分布在空气污染物浓度中的应用。*Environmetrics*：国际环境计量学会官方期刊, 19(3):235–249。

Manisalidis, I., Stavropoulou, E., Stavropoulos, A., and Bezirtzoglou, E.（2020）。空气污染对环境和健康的影响：综述。公共卫生前沿, 第14页。

Neelon, B.（2019）。基于 Pólya-Gamma 混合的贝叶斯零膨胀负二项回归。贝叶斯分析,14(3):829。

¨Ozkaynak, H., Xue, J., Weker, R., Butler, D., and Koutrakis, P.（1996年）。颗粒物研究（PTEAM）：数据分析。最终报告，第3卷。技术报告，哈佛大学，美国马萨诸塞州波士顿。公共卫生学院。

Polson, N. G., Scott, J. G., 和 Windle, J.（2013年）。使用 p´olya–gamma 潜变量对逻辑回归模型进行贝叶斯推断。《美国统计学会会刊》，108(504):1339–1349。

Reich, B. J.（2012年）。用于检测环境过程分布变化的时空分位数回归。皇家统计学会期刊：*C*辑（应用统计），61(4):535–553。

Reich, B. J., Fuentes, M., and Burke, J. (2009). 在考虑人体暴露的情况下分析超细颗粒物的影响。*Environmetrics*：国际环境计量学会官方期刊，20(2):131–146。

Richardson, S. and Best, N. (2003). 健康—环境效应生态学研究中的贝叶斯层级模型。*Environmetrics*：国际环境计量学会官方期刊，14(2):129–147。

Richmond-Bryant, J. 和 Long, T. C.（2020）。暴露测量误差对不同研究设计的流行病学研究结果的影响。《暴露科学与环境流行病学杂志》，30(3):420–429。

¨Sarnat, S. E., Sarnat, J. A., Mulholland, J., Isakov, V., Ozkaynak, H., Chang, H. H., Klein, M., 和 Tolbert, P. E.（2013年）。在亚特兰大的一项时间序列流行病学研究中应用环境空气污染暴露的替代时空指标。《暴露科学与环境流行病学杂志》，23(6):593–605。

Sheppard, L.（2003）。群体层面研究中关于偏差和信息的见解。生物统计学，4(2):265–278。

Steinle, S., Reis, S., 和 Sabel, C. E.（2013年）。量化人体空气污染暴露——从静态监测迈向时空分辨的个人暴露评估。《整体环境科学》，443:184–193。

Steinle, S., Reis, S., Sabel, C. E., Semple, S., Twigg, M. M., Braban, C. F., Leeson, S. R., Heal, M. R., Harrison, D., Lin, C., 等（2015）。室内外微环境中 PM2.5 的个人暴露监测。《整体环境科学》，508：383–394。

Sugg, M. M., Fuhrmann, C. M. 和 Runkle, J. D.（2018年）。美国东南部户外工作人群个人环境温度的时空变化。《国际生物气象学杂志》，62(8)：1521–1534。

Thornton, P., Thornton, M., Mayer, B., Wei, Y., Devarakonda, R., Vose, R. 和 Cook, R.（2016年）。Daymet：北美洲 1 公里网格的每日地表天气数据，第 3 版。橡树岭国家实验室分布式主动档案中心，橡树岭，田纳西州，美国。美国农业部国家农业统计局，*2019*年。《*2017*年农业普查：摘要与州级数据》，地理区域系列，第 *51* 部分，*AC-17-A-51*。

Yoo, E.-h., Eum, Y., Roberts, J. E., Gao, Q., 和 Chen, K.（2021）。极端温度与精神障碍相关急诊就诊之间的关联：一项在美国纽约州开展的多区域时间序列研究。《整体环境科学》，792：148246。

Zhou, J., Chang, H. H., 和 Fuentes, M.（2012年）。利用校准后的气候模型输出估计气候变化的健康影响。《农业、生物与环境统计学杂志》，17(3)：377–394。

Zidek, J. V., Shaddick, G., Meloche, J., Chatfield, C., 和 White, R.（2007）。预测个人暴露于环境危害的框架。《环境与生态统计学》，14(4)：411–431。

Zidek, J.V., Shaddick, G., White, R., Meloche, J., 和 Chatfield, C.（2005年）。使用概率模型（PCNEM）估计个人暴露于空气污染的程度。*Environmetrics*：国际环境计量学会官方期刊，16(5)：481–493。

8 资助

本研究得到美国国家环境健康科学研究所的 R01ES027892 和 R01ES028346 资助。本文内容仅由作者负责，并不一定代表美国国立卫生研究院的官方观点。