arXiv:2305.01906v1 [stat.ME] 3 May 2023

一种识别时空有序分类数据中变点的贝叶斯方法：COVID-19 数据的应用

SiddharthRawat[1],**Abe Durrant**[2],**Adam Simpson**[2],**Grant Nielson**[2], **CandaceBerrett**[2] **以及Soudeep Deb**[1],

[1]*Indian Institute of Management Bangalore*Bannerghatta Main Rd, Bangalore, KA 560076, India.Corresponding author: S. Rawat, Email: e-mail: siddharth.rawat19@iimb.ac.in

[2]*Department of Statistics, Brigham Young University, Provo, UT, USA*

摘要： 尽管已有大量文献研究连续时空过程中的结构变化识别问题，但针对分类时空数据的相关研究却并非如此。本文填补了这一研究空白，提出了一种新颖的时空模型，用于识别有序分类数据中的变点。该模型利用加性均值结构，并通过可分离的高斯时空过程对潜在变量进行建模。我们提出的方法不仅能够检测均值结构中的显著变化，还能识别时空协方差结构中的显著变化。我们通过贝叶斯框架实现该模型，相较于传统方法，该框架具有计算优势。从实际应用的角度来看，我们的方法能够处理有序分类数据，这为其在实际应用中的表现提供了额外优势。本文以美国纽约州各县的 COVID-19 数据为例进行说明，这些数据根据美国疾病控制与预防中心（CDC）的指南转换为分类数据。我们的模型识别出 COVID-19 传播水平中的三个变点，这些变点确实与大流行期间不同特定变异株所引发的“疫情浪潮”相吻合。研究结果还揭示了疫苗接种的影响，以及大流行不同阶段时空依赖程度的有趣见解。

关键词和短语： 大流行建模、吉布斯抽样、切片抽样、冠状病毒、过程变化、马尔可夫链蒙特卡洛。

1. 引言

关于变点问题，现有文献十分丰富。最早的研究由Chernoff 和 Zacks (1964)分析了独立同分布（i.i.d.）高斯分布中的均值结构变化。后续研究考察了非高斯情形下的方差变点和结构变化（例如，参见Zacks, 1983；Krishnaiah 和 Miao，1988）。从独立同分布的高斯框架转向其他情形后，变点通常在时间序列数据中建模或检验。在这种情况下，变点意味着时间上的特定时点之后，时间依赖性发生变化。处理时间变点最简单的情形，是考虑变点之后均值结构发生变化，而自回归结构或依赖结构保持不变（参见Brockwell 和 Davis，1991年）。Taylor (1994)提出了一类用于处理随机波动率的新模型；Kim、Shephard 和 Chib(1998)则在自回归条件异方差（ARCH）框架下分析了类似的研究主题。

事实上，关于时间序列数据中变点检测方法的研究在现有文献中已经十分丰富。由于本文关注的是时空数据集，因此我们在此对相关文献作简要回顾。Majumdar、Gelfand 和 Banerjee （2005年）提出了一种基于模型的贝叶斯框架方法，并估计了均值和依赖结构中的单个变点。Yu 等 （2008）以及Xu 等 （2012年）利用条件自回归（CAR）结构构建了多层时空双变点模型。他们将该模型应用于考察酒类销售网点管控政策对攻击性暴力发生率的影响。Altieri 等 （2015年）针对时空点过程提出了贝叶斯变点模型，并借助积分嵌套拉普拉斯近似（INLA）拟合该模型，以检测多个变点。Altieri 等 （2016年）将此前的工作扩展到贝叶斯 P 样条框架，用于研究地震点过程。Fiedler 等 （2018）研究了时空地震数据中的多变点模型，并通过贝叶斯方法实现了该模型。最近，Berrett

等 （2023年）采用模型选择框架，为高斯时空数据在任意给定位置选择变化的数量。

尽管在实际问题中，针对连续数据的变点建模最为常见，但响应变量有时会以有序类别的形式记录。一个典型例子是调查回答。此外，即使响应变量是连续的，考虑到数据收集中的误差以及计数数据建模可能存在的困难，将其转换为有序分类变量通常也是合理的（参见 Taylor，1961年；Mullahy，1997年）。这对于本文的研究背景，即 COVID-19 病例计数数据，尤为重要。已有研究表明，这些病例计数数据噪声较大，通常存在漏报，并且往往不可靠，主要原因是当地检测数据收集和上游报告中的错误（Dubrow，2022年）。因此，将其建模为有序分类响应变量更为合适。在本研究中，我们依据美国疾病控制与预防中心的指南（美国疾病控制与预防中心，2022年），根据县级每周新增病例数定义 COVID-19 的“传播水平”类别。这类疾病的传播取决于许多无法测量或难以精确测量的量——例如不同病毒变异株的出现以及居家检测的可及性——从而导致真实 COVID-19 病例与测量值之间的关系发生变化。我们计划在考虑适当协变量的情况下，研究传播水平之间不断变化的关系，其中包括疫苗接种这一重要因素。

分析 COVID-19 传播面临的一项特殊挑战是时空依赖模式，如 Rawat和Deb (2021年所示。这一现象使病例数计数数据的协方差结构变得复杂。关于这一背景下的其他挑战，参见Bertozzi 等人(2020年)。有趣的是，尽管采用上述有序响应变量的表述会导致精确度降低，但时空依赖仍然存在，正如我们在第 2节的探索性分析中所展示的。因此，有必要在有序响应变量中考虑时空依赖。

关于有序响应变量的多项回归模型已有大量文献，这些模型是逻辑回归和 Probit 回归的扩展（例如，Campbell and Donner, 1989年；Benoit, 2012年；Liang、Bi 和 Zhan, 2020年）。采用 Probit 链接函数的贝叶斯模型使用潜变量来促进吉布斯抽样，如 Albert and Chib (1993) 所示。Tanner and Wong (1987) 还研究了贝叶斯估计方法及其计算发展，探讨了如何使用近似抽样方法计算后验分布以进行推断。Cowles (1996) 讨论了马尔可夫链蒙特卡罗（MCMC）近似，并通过更新吉布斯抽样步骤来加速贝叶斯估计过程的收敛（另见 Imai and van Dyk, 2005年；Berrett and Calder, 2012年；Heiner 等人, 2022年）。此外，Nandram and Chen (1996) 提出了广义线性模型（GLM）的重新参数化方法，以加快吉布斯采样器的收敛速度。出于同样目的，Liu and Wu (1999) 提出了参数扩展数据增广（PX-DA）方法，该方法利用辅助变量来加速收敛。Chen and Dey (2000) 在具有信息先验分布的贝叶斯框架下，采用多元正态链接函数的尺度混合，处理了相关有序数据的广义线性模型（GLM）估计问题。Roy and Hobert (2007) 比较了 PX-DA 与 Albert and Chib (1993) 提出的贝叶斯数据增广方法在性能和收敛速度方面的表现。

尽管这些方法很有价值，但它们并未提出如何检测分类时空数据集中的均值、方差或相关模式的任何变化。因此，亟需开发一种变点检测方法，将这些建模方法结合起来并加以拓展，以量化存在时空依赖时的关系。这正是我们旨在填补现有文献空白之处。具体而言，我们提出一种新方法，采用贝叶斯方法对时空有序分类数据中的变点进行建模。在所提出的模型中，使用潜变量对分类观测进行建模，并允许变点之后均值和方差-协方差结构发生变化，从而使时空关联也能够改变。该模型通过贝叶斯框架实现，相较于经典方法具有计算优势。为此，我们采用吉布斯抽样（Geman 和 Geman，1984年）和单变量切片抽样算法（Neal，2003年）。

我们的方法在三个方面为相关文献作出了重要贡献。首先，据我们所知，这是首次尝试开发一种用于检测有序分类时空数据中变点的方法。其次，我们提供了一种从统计学角度识别变点是否存在的方法，并将所提出的方法与二分分割技术相结合，扩展为一种能够检测数据中多个变点的方法。第三，我们的模型采用高效的估计算法来估计模型中的所有参数，并减轻了类似连续型时空模型中普遍存在的一些计算负担

响应变量。我们将在下文第3.2 节对此作详细阐述。

本文其余部分安排如下。第 2 节简要介绍并概述数据，同时阐述考虑时空模型的动机。我们提出的方法及相关讨论见第 3节。第 4节介绍了该方法在美利坚合众国（美国）纽约州 COVID-19 数据中的应用。最后，我们在第 5节进行讨论并提出相关意见。附录 A提供了一些额外的探索性分析，而该方法中所用贝叶斯步骤的详细数学推导则见附录 B。

2. 数据

用于分析的数据取自由 JHU-CSSE (2020年) 维护的 GitHub 存储库。数据以县为单位提供。本研究关注纽约州所有县的数据。我们按周粒度查看数据，时间范围为 2020年1月20日 至 2022年5月16日所在周。这涵盖了纽约大部分 COVID-19“浪潮”。请注意，数据按周整理主要有两个原因。首先，这可以处理周末数据报告带来的问题（参见 Ricon-Becker等人，2020年）。其次，将数据从每日汇总为每周数据，在不会损失太多数据收集方面信息的同时，显著降低了计算成本。因此，该数据集包含来自 62 个县、跨越 122 周的 7564 个观测值。如第 1 节所述，首先将每周数字转换为反映不同传播水平的分类数据。我们遵循美国疾病控制与预防中心的指南来定义这些水平，具体见表 1。

TABLE 1
CDC guidelines used to define the ordered categories of COVID-19 transmission levels.

一周内每10万人新增病例数	类别
0 至 9.99	1
10 至 49.99	2
50 至 99.99	3
$\geqslant$ 100	4

作为探索性分析的第一步，图1 展示了数据集整个时间段内四个不同周的纽约 COVID-19 传播水平类别。在每个时间点，这些类别都呈现出明显的空间依赖。地图显示，COVID-19 传播水平相同的县往往聚集在一起，这表明任何建模方法都需要考虑或量化空间依赖。

我们观察到，各县的传播水平随时间呈现类似的依赖关系。由于响应变量具有类别性质，通过更平滑、噪声更少的图表来考察各县之间的相似性更为容易。具体而言，图2 展示了观测时段内以30周为窗口计算的平均传播水平。每条线代表一个不同的县，其中四个县——奥尔巴尼（红色）、尼亚加拉（绿色）、皇后区（蓝色）和萨福克（紫色）——以不同颜色突出显示，以供说明。显然，在整个时段内，各县的传播水平相似。

我们的模型还纳入了两个协变量：前一周该县发生的死亡人数（反映病毒严重程度），以及接种第一剂疫苗者的比例（反映预防措施的实施程度）。关于使用这些协变量的理由，我们在第3.3节中进行了更详细的讨论。对于第一个协变量，沿用Rawat和Deb (2021年)的类似做法，我们对前一周的新增死亡数据进行对数变换，并将其记为 $\log\ d(s_i, t-1)$，表示第 i^{th} 县在第 t^{th} 周的情况。对于疫苗接种协变量，我们将疫苗接种率 $x(s_i, t)$定义为

$$x(s_i, t) = \frac{vc_{s_i,t}}{p_{s_i,t}}, \tag{1}$$

其中，$vc_{s_i,t}$ 表示截至第 t^{th} 周，在 i^{th} 县接种的 COVID-19 疫苗首剂累计数量，而$p_{s_i,t}$ 表示相应的人口数量。在我们的模型中，对于这两个协变量，我们均采用标准化数值，使其均值为 0，标准差为 1。

<escape>为了了解协变量值随时间的变化情况，我们查看图 3。同样，每条线代表一个独立的县，其中四个具有代表性的县以不同颜色突出显示。与传播水平类似，所有县的协变量总体上呈现出相似的模式，但各县的取值存在相当大的差异。</escape>

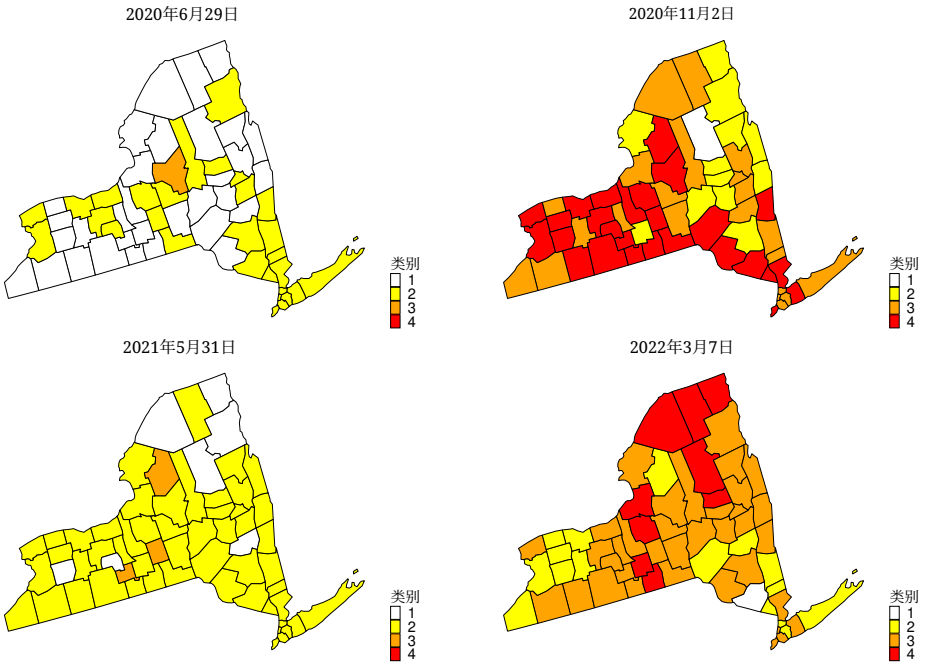


FIG 1. *COVID-19 transmission level category at four different time points for New York state counties.*

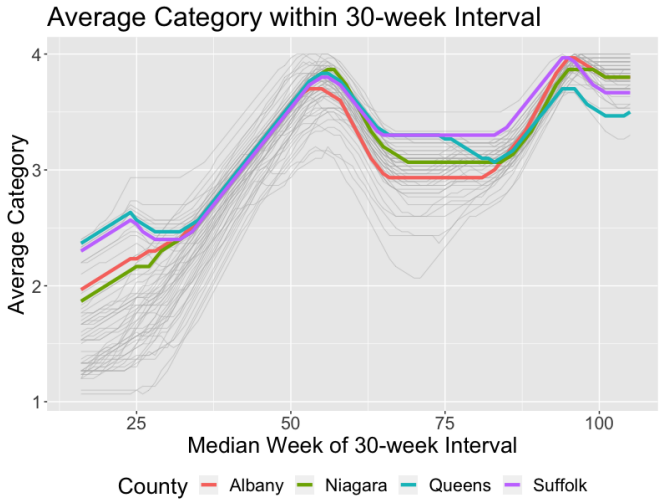
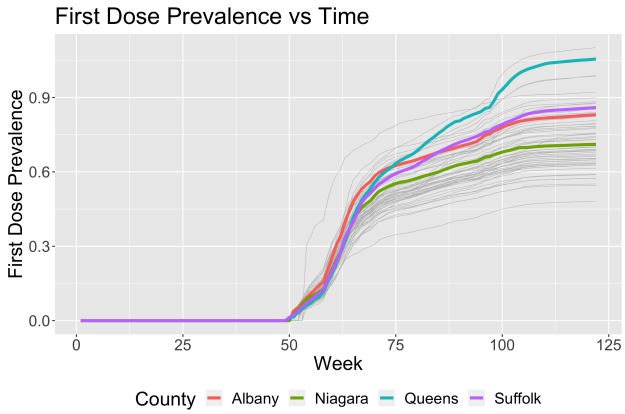


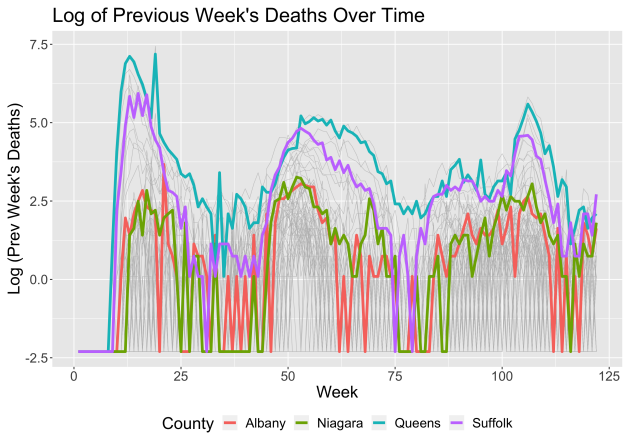
FIG 2. *Average transmission level for 30-week windows for all counties (gray) with four counties from across the state highlighted in the four colors.*

协变量值随时间变化的趋势不同于传播水平的趋势；例如，疫苗接种量不会减少，但传播水平在整个时间段内既有上升也有下降。因此，必须采用一种能够考虑协变量关系随时间变化的模型。由于在模型中引入了变点，我们的方法能够恰当地处理这种关系变化。

值得一提的是，我们也尝试了线性趋势函数和二次趋势函数，但由于 COVID-19 数据呈现出类似“波浪”的模式，因此发现它们并不适用于本研究。此外，我们还探究了 COVID-19疫苗第二剂接种率作为额外协变量的作用。结果



(a)



(b)

FIG 3. *Plots showing the values of vaccine prevalence (top) and log of the previous week's new deaths (bottom) across time. The different lines represent the different counties with four counties highlighted in color.*

发现其与第一剂的关系相似，因此认为无需将其纳入模型。

在美国，各州、各县乃至各城市对 COVID-19 疫情的应对和政策都存在很大差异。因此，尽管我们的分析旨在共同考察纽约州所有县的传播水平，我们还是首先探讨了单独分析各县是否更为合适。由于篇幅所限，本文的这部分分析详见附录 A。有趣的是，我们发现，单个县的数据不足以进行估计。例如，我们首先发现，部分县（约一半）的模型即使在后验样本数量非常大的情况下也无法收敛。此外，单个县的数据匮乏意味着我们也只能拟合这些模型来估计一个变点，因为信息不足以估计多个（如果能估计的话）变点。第三，我们发现，对于能够收敛的县，变点和系数的估计不确定范围很大，且与其他县的估计范围高度重叠。基于这些原因，能够借用邻近地区信息的时空模型绝对必不可少。

3. 方法

3.1. 模型开发

考虑位置 $s_i \in \mathcal{S} \subset \mathbb{R}^2$, $i = 1, \ldots, n$ 和时间点$t \in \Gamma$的分类响应变量 $y(s_i, t)$。虽然我们主要使用规则间隔的数据集 $\Gamma = \{1, \ldots, T\}$，但需要强调的是，该方法同样适用于非等间隔数据。同样，尽管 COVID-19 数据并非

点级，我们发现该方法对于区域汇总数据效果良好，延续了其他研究人员的类似工作（Jin、Carlin和Banerjee，2005年；Bradley、Holan和Wikle，2015年）。

对于响应变量，令可能类别的总数表示为 m。为方便起见，我们将这些类别记为1、2、.....、 m。鉴于响应是一个有序变量，我们采用有序Probit模型的结构。关于标准设定下此类模型的综述，参见Daykin和Moffatt(2002年)。为定义本研究中的模型，令 $\pi(s_i、\ t)$为一个潜在随机变量，使得对于 $j = 1、2、.....、\ m$，

$$y(s_i, t) = j \ \text{ if } \delta_{j-1} < \pi(s_i, t) \leqslant \delta_j, \tag{2}$$

其中， $-\infty = \delta_0 < \delta_1 = 0 < \delta_2 < \ldots < \delta_{m-1} < \delta_m = \infty$ 为决定分类响应分布的切点。

接下来，令 t_0 为变点，潜变量的模型定义为

$$\pi(s_i, t) = \begin{cases} \mu(s_i, t) + u(s_i, t) + \epsilon(s_i, t), & \text{for } t \leqslant t_0, \\ \mu^*(s_i, t) + u(s_i, t) + v(s_i, t) + \epsilon^*(s_i, t), & \text{for } t > t_0. \end{cases} \tag{3}$$

在该模型中， $\mu(\cdot)$和 $\mu^*(\cdot)$分别构成变点前后的均值结构。两个时期各自的协方差结构写成时空相关过程与白噪声过程之和。项 $u(s_i、\ t)$表示变点前的时空效应，而 $v(s_i、\ t)$表示结构突变后该效应的偏差。假设这两个过程均服从高斯分布。最后，潜变量的白噪声过程定义为 $\epsilon(s_i、\ t)$和$\epsilon^*(s_i、\ t)$，并假设它们分别为均值为零、标准差为 σ_ϵ和 σ_{ϵ^*}的独立同分布高斯随机变量。下文将使用记号 $N_r(\cdot、\ \cdot)$表示 r元高斯分布。

在模型中， $\mu(\cdot)$ 和 $\mu^*(\cdot)$ 都被视为协变量的线性函数，其中一些协变量的效应可以是空间变异的。用数学形式描述如下：令 $x_g(s_i,\ t)$ 表示对所有位置均具有固定效应的 g^{th} 协变量 $(1 \leqslant g \leqslant G)$，也就是说，该效应不随空间变化。类似地，对于 $1 \leqslant h \leqslant H$，令$x_{g+h}(s_i,\ t)$ 表示效应被视为空间变异的协变量。于是，均值函数可写为

$$\begin{aligned} \mu(s_i, t) &= \beta_0 + \sum_{g=1}^{G} \beta_g x_g(s_i, t) + \sum_{h=1}^{H} \gamma_h(s_i)\, x_{g+h}(s_i, t), \quad \text{for } \ t \leqslant t_0, \\ \mu^*(s_i, t) &= \beta_0^* + \sum_{g=1}^{G} \beta_g^* x_g(s_i, t) + \sum_{h=1}^{H} \gamma_h^{*'}(s_i)\, x_{g+h}(s_i, t), \quad \text{for } t > t_0; \end{aligned} \tag{4}$$

其中， β_g 和 β_g^* 表示变点前后固定效应协变量的系数，$\gamma_h(s_i)$ 和 $\gamma_h^*(s_i)$ 则是两个区段中的位置特异系数。换言之， $\boldsymbol{\gamma}_h = (\gamma_h(s_1).....,\ \gamma_h(s_n))'$
表示用于刻画相应回归变量效应中空间变异的系数向量。向量 $\boldsymbol{\gamma}_h^*$ 的解释也相同。我们假设这些效应具有空间相关性，并遵循高斯框架下的指数衰减协方差结构。具体而言，为了在贝叶斯算法中估计效应大小（将在第3.2节详细讨论），我们设 $\boldsymbol{\gamma}_h \sim N_n(0,\ k\Omega_s^h)$，且 $\boldsymbol{\gamma}_h^* \sim N_n(0,\ k\Omega_s^{*h})$。值得一提的是，向量 $\boldsymbol{\gamma}_1$，$\boldsymbol{\gamma}_2$，.....，$\boldsymbol{\gamma}_H$ 彼此独立。现在，离散矩阵Ω_s^h 和Ω_s^{*h}定义如下：

$$(\Omega_s^h)_{ij} = \exp\{-\omega_s^h ||s_i - s_j||\},\ (\Omega_s^{*h})_{ij} = \exp\{-\omega_s^{*h} ||s_i - s_j||\}; \tag{5}$$

其中， ω_s^h 和 ω_s^{*h}分别是变点前后的衰减参数。请注意，这是 Mat´ern 协方差函数类的一个特例，其中参数反映了相关性的衰减程度。在计算中，项 $||s_i - s_j||$ 通过两个位置 s_i 和 s_j之间的大圆距离计算得出。

对于时空过程，与相关应用领域的其他研究类似（例如，参见 Rawat和Deb，2021年），我们假设相关函数具有可分离性。因此，空间相关函数和时间相关函数均被认为遵循指数衰减结构。对于空间相关性，我们再次使用大圆距离，而对于时间相关性，则使用时间差

$|t_1 - t_2|$ 通过时间点之间的滞后计算得出。换言之，对于衰减参数 ϕ_{us}, ϕ_{ut},ϕ_{vs}, ϕ_{vt}，可以表示为

$$\begin{aligned} \text{Corr}\left(u(s_i,t_1),u(s_j,t_2)\right) &= \exp\left\{-\phi_{us}\left\|s_i-s_j\right\|\right\} \times \exp\left\{-\phi_{ut}\left|t_1-t_2\right|\right\}, \\ \text{Corr}\left(v(s_i,t_1),v(s_j,t_2)\right) &= \exp\left\{-\phi_{vs}\left\|s_i-s_j\right\|\right\} \times \exp\left\{-\phi_{vt}\left|t_1-t_2\right|\right\}. \end{aligned} \tag{6}$$

在本文中，我们用 $M = nT$ 表示观测值总数。令 $\boldsymbol{Y}$为由 y（s_i, t）个观测值组成的 M 维向量，先按时间点排列，再按位置索引排列。向量$\boldsymbol{\pi}$、 $\boldsymbol{U}$、 $\boldsymbol{V}$、 $\boldsymbol{\epsilon}$、 $\boldsymbol{\epsilon}^*$ 也以相同方式构造。回归变量值的设计矩阵表示为 X。于是，采用向量矩阵记法，上述模型可写为：

$$\boldsymbol{\pi} = \begin{cases} X\boldsymbol{\theta} + \boldsymbol{U} + \boldsymbol{\epsilon}, & \text{for } t \leqslant t_0, \\ X\boldsymbol{\theta}^* + \boldsymbol{U} + \boldsymbol{V} + \boldsymbol{\epsilon}^*, & \text{for } t > t_0, \end{cases} \tag{7}$$

其中， $\boldsymbol{\theta}$ 是将 $(\beta_0, \beta_1, \ldots, \beta_G)'$ 合并后得到的 k 维系数向量，而参数向量 $(\boldsymbol{\gamma}_1', \boldsymbol{\gamma}_2', \ldots, \boldsymbol{\gamma}_H')'$。 $\boldsymbol{\theta}^*$ 的定义方式类似。

此外，根据上述讨论，可写出 $\boldsymbol{U} \sim N_M(0, \sigma_u^2\Sigma_{ut} \otimes \Sigma_{us})$，使得 $(\Sigma_{ut})_{ij} = \exp\{-\phi_{ut}|t_i - t_j|\}$ 且 $(\Sigma_{us})_{ij} = \exp\{-\phi_{vs}||s_i - s_j||\}$。类似地，我们可以定义协方差矩阵 Σ_{vt}、Σ_{vs}，并写出 $\boldsymbol{V} \sim N_M(0, \sigma_v^2(\Sigma_{vt} \otimes \Sigma_{vs}))$。对于白噪声向量，我们知道 $\boldsymbol{\epsilon} \sim N_M(0, \sigma_\epsilon^2\mathbf{I})$ 和$\boldsymbol{\epsilon}^* \sim N_M(0, \sigma_{\epsilon^*}^2\mathbf{I})$，其中$\mathbf{I}$ 表示相应阶数的单位矩阵。

3.2. 贝叶斯估计

我们采用完整的贝叶斯框架来估计模型参数。为此，需要设定合适的先验。需要注意的是，为避免计算中出现可识别性问题，按照Higgs 和 Hoeting (2010年) 的做法，我们将方差参数 σ_u^2、 σ_v^2、 σ_ϵ^2和 $\sigma_{\epsilon^*}^2$ 设为 1。我们采用吉布斯抽样原理，对其他参数进行后验计算。回顾一下，吉布斯抽样是一种马尔可夫链蒙特卡罗（MCMC）技术，其中每个参数都会根据上一轮迭代中其他参数的取值，按照全条件后验分布依次更新。感兴趣的读者可参阅 Gelfand (2000年)，了解吉布斯抽样的起源和实现方法。为说明采样器的步骤，我们使用符号 $\mathcal{F}$ 表示上一轮迭代中除待计算其全条件后验分布的参数以外所有参数所包含的信息。

首先，考虑潜在向量 $\boldsymbol{\pi}$，令 π_i 表示其第 i^{th} 个元素，而 $\boldsymbol{\pi}_{-i}$ 表示除第 i^{th} 个元素外包含所有元素的向量 $\boldsymbol{\pi}$。同理，为向量 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 定义 u_i 和 v_i，而 x_i' 表示式 (7) 中定义的矩阵 X 的第 i^{th} 行。令 $t(i)$ 为对应于同一观测的时间点。然后，通过对 $\epsilon(s_i, t(i))$ 和 $\epsilon^*(s_i, t(i))$ 进行边缘化，可得到 π_i 的全条件分布。沿用前文的记号，若 $\mathcal{F}$表示 $(\boldsymbol{Y}, \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{\theta}, \boldsymbol{\theta}^*, \boldsymbol{\delta}, t_0, \sigma_\epsilon^2, \sigma_{\epsilon^*}^2, \boldsymbol{\pi}_{-i})$，则可以证明

$$\pi_i|\mathcal{F} \sim \begin{cases} TN(x_i'\boldsymbol{\theta} + u_i, \sigma_\epsilon^2; \delta_{j-1}, \delta_j), & \text{if } t(i) \leqslant t_0,\ y_i = j, \\ TN(x_i'\boldsymbol{\theta}^* + u_i + v_i, \sigma_{\epsilon^*}^2; \delta_{j-1}, \delta_j), & \text{if } t(i) > t_0,\ y_i = j, \end{cases} \tag{8}$$

其中， $TN(a, b;\ c, d)$ 表示单变量截断正态分布，它是将均值为 a 、方差为 b 的正态分布截断在区间 (c, d) 后得到的。

对于 $\boldsymbol{\theta}$ 和 $\boldsymbol{\theta}^*$中的每个分量，我们均假设其先验分布为均匀平坦的无限支撑分布。因此，这些参数的后验分布由数据主导，且先验分布中不假设任何信息。于是，这些参数向量的全条件后验分布可写为

$$\begin{aligned} \boldsymbol{\theta}|\mathcal{F} &\sim N_k\left(\Sigma_{\boldsymbol{\theta}}\left[\frac{X^{-'}(\boldsymbol{\pi}^- - \boldsymbol{U}^-)}{\sigma_\epsilon^2}\right], \Sigma_{\boldsymbol{\theta}}\right), \\ \boldsymbol{\theta}^*|\mathcal{F} &\sim N_k\left(\Sigma_{\boldsymbol{\theta}^*}\left[\frac{X^{+'}(\boldsymbol{\pi}^+ - \boldsymbol{U}^+ - \boldsymbol{V}^+)}{\sigma_\epsilon^2}\right], \Sigma_{\boldsymbol{\theta}^*}\right). \end{aligned} \tag{9}$$

在此方程中，$\boldsymbol{\pi}^-$ 和 $\boldsymbol{\pi}^+$ 分别表示变点前后的 $\boldsymbol{\pi}$ 向量。$\boldsymbol{U}^-$、$\boldsymbol{U}^+$、$\boldsymbol{V}^-$、$\boldsymbol{V}^+$、$\boldsymbol{Y}^-$、$\boldsymbol{Y}^+$、T^-、T^+、X^-和 X^+的定义类似。离散矩阵 $\Sigma_{\boldsymbol{\theta}}$和 $\Sigma_{\boldsymbol{\theta}^*}$定义为

$$\Sigma_{\boldsymbol{\theta}} = \left[\frac{X^{-'}X^-}{\sigma_\epsilon^2} + \frac{\Psi^{-1}}{k}\right]^{-1},\ \Sigma_{\boldsymbol{\theta}^*} = \left[\frac{X^{+'}X^+}{\sigma_{\epsilon^*}^2} + \frac{\Psi^{*-1}}{k}\right]^{-1}, \tag{10}$$

其中，Ψ和 Ψ^*是块对角矩阵，其元素依次为（$\mathbf{I}_G$，Ω_{s_1}，Ω_{s_2}，……，Ω_{s_H}）和（$\mathbf{I}_G$，$\Omega^*_{s_1}$，$\Omega^*_{s_2}$，……，$\Omega^*_{s_H}$）。值得注意的是，即使对 $\boldsymbol{\theta}$ 和 $\boldsymbol{\theta}^*$采用高斯先验，而非前述平坦分布，上述条件后验分布仍会是高斯分布（参数值不同）。

接下来，为了计算切点 δ_2、δ_3、……、δ_{m-1}的后验分布，我们采用 Heiner et al. (2022) 用于其他切点参数化的单变量切片抽样算法。该算法通过 FitzJohn（2012）开发的 R 包“diversitree”实现。下面简要介绍单变量切片抽样算法。该算法主要分为三个步骤。首先，从垂直切片 $(0, g(r_0))$ 中抽取一个随机数 z，其中 r_0 是随机变量 R 的起始点，$g(r_0)$ 是其概率密度函数。其次，在 r_0 周围随机选取一个宽度为 w 的区间，并不断扩展，直到区间两端都位于水平切片之外；水平切片上的所有值均等于 z。第三，从上一步得到的宽度为 w的区间中随机抽取一个新点 r_1。更多详情请参阅 Neal（2003）。

对于 δ_j的先验分布，其支持集必须介于 δ_{j-1} 和 δ_{j+1} 之间，因为我们处理的是有序分类数据，如第 3.1 节所述。我们采用这两个界限之间的均匀分布作为 δ_j的先验分布。现在，令 $\boldsymbol{\pi}_j^-$ 和 $\boldsymbol{\pi}_{j+1}^-$ 为向量，它们分别仅包含向量 $\boldsymbol{\pi}^-$ 中对应于向量 $\boldsymbol{Y}^-$ 的类别为 j 和 $j+1$的那些值。类似地，定义 X_j^-、$\boldsymbol{U}_j^-$、X_{j+1}^-和 $\boldsymbol{U}_{j+1}^-$。同样，令 $\boldsymbol{\pi}_j^+$ 和 $\boldsymbol{\pi}_{j+1}^+$ 为向量，它们分别仅包含 $\boldsymbol{\pi}^+$ 中对应于向量 $\boldsymbol{Y}^+$ 的类别为 j 和 $j+1$的那些值。类似地，定义 X_j^+、$\boldsymbol{U}_j^+$、$\boldsymbol{V}_j^+$、X_{j+1}^+、$\boldsymbol{U}_{j+1}^+$和 $\boldsymbol{V}_{j+1}^+$。于是，δ_j 的全条件后验分布可写为

$$\begin{aligned} f(\delta_j|\mathcal{F}) \propto I(\delta_{j-1} < \delta_j < \delta_{j+1}) & \int_{\delta_{j-1}}^{\delta_j} \exp\left\{\frac{-1}{2\sigma_\epsilon^2}\left\|\boldsymbol{\pi}_j^- - X_j^-\boldsymbol{\theta} - \boldsymbol{U}_j^-\right\|^2\right\}\mathrm{d}\boldsymbol{\pi}_j^- \\ & \int_{\delta_j}^{\delta_{j+1}} \exp\left\{\frac{-1}{2\sigma_\epsilon^2}\left\|\boldsymbol{\pi}_{j+1}^- - X_{j+1}^-\boldsymbol{\theta} - \boldsymbol{U}_{j+1}^-\right\|^2\right\}\mathrm{d}\boldsymbol{\pi}_{j+1}^- \\ & \int_{\delta_{j-1}}^{\delta_j} \exp\left\{\frac{-1}{2\sigma_{\epsilon^*}^2}\left\|\boldsymbol{\pi}_j^+ - X_j^+\boldsymbol{\theta}^* - \boldsymbol{U}_j^+ - \boldsymbol{V}_j^+\right\|^2\right\}\mathrm{d}\boldsymbol{\pi}_j^+ \\ & \int_{\delta_j}^{\delta_{j+1}} \exp\left\{\frac{-1}{2\sigma_{\epsilon^*}^2}\left\|\boldsymbol{\pi}_{j+1}^+ - X_{j+1}^+\boldsymbol{\theta}^* - \boldsymbol{U}_{j+1}^+ - \boldsymbol{V}_{j+1}^+\right\|^2\right\}\mathrm{d}\boldsymbol{\pi}_{j+1}^+. \end{aligned} \tag{11}$$

不难看出，δ_j 的后验分布没有闭式解。为了实现切片抽样过程，我们通过适当的变量替换简化上述积分。值得一提的是，我们也在此情境下探索了多变量切片抽样算法，但与单变量切片抽样过程相比，它的计算成本更高，对我们的算法而言效率也更低。

我们将注意力转向时空过程向量 $\boldsymbol{U}$和 $\boldsymbol{V}$。回顾一下，$M \times 1$ 向量 $\boldsymbol{Y}$表示整个数据，而 $\boldsymbol{U}$ 和 $\boldsymbol{V}$则是矩阵按列拼接后得到的形式。

$$U_{T\times n} = \begin{pmatrix} u_{s_1,1} & u_{s_1,2} & \dots & u_{s_1,T} \\ u_{s_2,1} & u_{s_2,2} & \dots & u_{s_2,T} \\ \vdots & \vdots & \vdots & \vdots \\ u_{s_n,1} & u_{s_n,2} & \dots & u_{s_n,T} \end{pmatrix}, V_{T\times n} = \begin{pmatrix} v_{s_1,1} & v_{s_1,2} & \dots & v_{s_1,T} \\ v_{s_2,1} & v_{s_2,2} & \dots & v_{s_2,T} \\ \vdots & \vdots & \vdots & \vdots \\ v_{s_n,1} & v_{s_n,2} & \dots & v_{s_n,T} \end{pmatrix}. \tag{12}$$

需要指出的是，在吉布斯抽样过程中，对于每条马尔可夫链的每次迭代，常规的整个向量后验分布 $\boldsymbol{U}$ 和 $\boldsymbol{V}$都需要对 $M \times M$ 维矩阵求逆。这在计算上开销很大。为解决这一问题，我们利用时空结构的可分离性，采用计算成本较低的步骤。将 $\boldsymbol{V}$划分为 $\boldsymbol{V}_1$，$\boldsymbol{V}_2$，……，$\boldsymbol{V}_T$其中 $\boldsymbol{V}_t$是式（ t^{th}）中矩阵的第 V列，并令12，设 $\boldsymbol{V}_{-t}$ =($\boldsymbol{V}_2$，$\ldots \boldsymbol{V}_{t-1}$，$\boldsymbol{V}_1$，$\boldsymbol{V}_{t+1}$，……，$\boldsymbol{V}_T$)。考虑如下分块结构

$$\Sigma_{vt} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

对于时间相关性矩阵，下标 1 表示与t^{th} 时间点对应的相关性部分，下标 2 表示与其余时间点对应的相关性部分。接着，令 $\boldsymbol{\mu}_{vc}$ =($\Sigma_{12} \otimes \Sigma_{vs}$)($\Sigma_{22}^{-1} \otimes \Sigma_{vs}^{-1}$)$\boldsymbol{V}_{-t}$和 $\Sigma_{vc} = (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}) \otimes \Sigma_{vs}$，根据多元正态分布理论可知

$$\boldsymbol{V}_t | \boldsymbol{V}_{-t}, \sigma_v^2 \sim N_n(\boldsymbol{\mu}_{vc}, \sigma_v^2 \Sigma_{vc}). \tag{13}$$

对 $\boldsymbol{U}$ 也采用相同思路，可得全条件后验分布为 e

$$\boldsymbol{V}_t | \mathcal{F} \sim \begin{cases} N_n(\mu_{vc}, \sigma_v^2 \Sigma_{vc}), & \text{for } t \leqslant t_0, \\ N_n\left(\Sigma_{V_t}\left(\frac{\Sigma_{vc}^{-1}\boldsymbol{\mu}_{vc}}{\sigma_v^2} + \frac{\boldsymbol{\pi}_t - X_t\boldsymbol{\theta}^* - \boldsymbol{U}_t}{\sigma_{\epsilon^*}^2}\right), \Sigma_{V_t}\right), & \text{for } t > t_0, \end{cases} \tag{14}$$

$$\boldsymbol{U}_t | \mathcal{F} \sim \begin{cases} N_n\left(\Sigma_{U_t}\left(\frac{\Sigma_{uc}^{-1}\boldsymbol{\mu}_{uc}}{\sigma_u^2} + \frac{\boldsymbol{\pi}_t - X_t\boldsymbol{\theta}}{\sigma_\epsilon^2}\right), \Sigma_{U_t}\right), & \text{for } t \leqslant t_0, \\ N_n\left(\Sigma_{U_t}\left(\frac{\Sigma_{uc}^{-1}\boldsymbol{\mu}_{uc}}{\sigma_u^2} + \frac{\boldsymbol{\pi}_t - X_t\boldsymbol{\theta}^* - \boldsymbol{V}_t}{\sigma_\epsilon^2}\right), \Sigma_{U_t}\right), & \text{for } t > t_0, \end{cases} \tag{15}$$

其中， Σ_{V_t} 和 Σ_{U_t} 定义为

$$\Sigma_{V_t} = \left[\frac{\Sigma_{vc}^{-1}}{\sigma_v^2} + \frac{\mathbf{I}_n}{\sigma_{\epsilon^*}^2}\right]^{-1}, \; \Sigma_{U_t} = \left[\frac{\Sigma_{uc}^{-1}}{\sigma_u^2} + \frac{\mathbf{I}_n}{\sigma_{\epsilon^*}^2}\right]^{-1}. \tag{16}$$

接下来，我们考察时空协方差函数中的四个衰减参数。可以计算 ϕ_{vt}、 ϕ_{vs}、 ϕ_{ut}和 ϕ_{us}的后验分布，并证明它们不服从任何已知的概率分布。因此，我们像之前对 δ_j 所采用的那样，使用单变量切片抽样算法。这里尤其值得注意的是，我们在连续尺度上估计衰减参数，这与许多其他研究者采用的传统交叉验证方法形成鲜明对比。使用交叉验证方法的问题在于，如果估计的衰减参数多于两个，那么为了处理计算负担，只能考虑参数的一小组可能取值。例如，我们考察 Sahu、Gelfand 和 Holland (2006) 的研究，他们在其模型中使用了两个时空过程来分析空气污染水平。然而，在他们的研究中，必须从空间衰减参数的仅五个取值和时间衰减参数的仅三个取值中找出衰减参数的最优值。显然，这限制了算法探索衰减参数所有可能取值的能力，并带来了无法探索实际样本空间的风险（对于我们的模型而言，这是一个四维立方体）。我们利用上述切片抽样过程来减轻估计衰减参数时的计算负担。对于先验，我们选择 0 到 3 之间的均匀分布，从而得到如下全条件后验分布

$$\begin{aligned} \phi_{vt} | \mathcal{F} &\propto |\Sigma_{vt}|^{-n/2} \exp\left\{\frac{-\boldsymbol{V}'(\Sigma_{vt}^{-1} \otimes \Sigma_{vs}^{-1})\boldsymbol{V}}{2\sigma_v^2}\right\} I(0 < \phi_{vt} < 3), \\ \phi_{vs} | \mathcal{F} &\propto |\Sigma_{vs}|^{-T/2} \exp\left\{\frac{-\boldsymbol{V}'(\Sigma_{vt}^{-1} \otimes \Sigma_{vs}^{-1})\boldsymbol{V}}{2\sigma_v^2}\right\} I(0 < \phi_{vs} < 3), \\ \phi_{ut} | \mathcal{F} &\propto |\Sigma_{ut}|^{-n/2} \exp\left\{\frac{-\boldsymbol{U}'(\Sigma_{ut}^{-1} \otimes \Sigma_{us}^{-1})\boldsymbol{U}}{2\sigma_u^2}\right\} I(0 < \phi_{ut} < 3), \\ \phi_{us} | \mathcal{F} &\propto |\Sigma_{us}|^{-T/2} \exp\left\{\frac{-\boldsymbol{U}'(\Sigma_{ut}^{-1} \otimes \Sigma_{us}^{-1})\boldsymbol{U}}{2\sigma_u^2}\right\} I(0 < \phi_{us} < 3). \end{aligned} \tag{17}$$

以完全相同的方式，我们还可以证明，对于空间变化系数的衰减参数，

$$\begin{aligned}\omega_s^h|\mathcal{F} &\propto |\Omega_s^h|^{-1/2}\exp\left\{\frac{-\boldsymbol{\gamma}_{\boldsymbol{h}}'(\Omega_s^h)^{-1}\boldsymbol{\gamma}_{\boldsymbol{h}}}{2k}\right\}I(0<\omega_s^h<3),\\ \omega_s^{*h}|\mathcal{F} &\propto |\Omega_s^{*h}|^{-1/2}\exp\left\{\frac{-\boldsymbol{\gamma}_{\boldsymbol{h}}^{*'}(\Omega_s^{*h})^{-1}\boldsymbol{\gamma}_{\boldsymbol{h}}^{*}}{2k}\right\}I(0<\omega_s^{*h}<3).\end{aligned} \tag{18}$$

最后，为了计算变点 t_0的全条件后验分布，我们考虑集合 $S_T=\{0,1,2,\ldots,T\}$ 并采用离散均匀先验分布。令 $f_{S_T}(t_0)$ 表示该先验。请注意，端点包含在样本空间中，以便算法能够收敛到任一端点，这实际上意味着不存在变点。然后，与参数 t_0 相对应的吉布斯抽样步骤需要条件后验分布，该分布由以下概率质量函数描述：

$$\begin{aligned}f(t_0|\mathcal{F})\propto I(t_0\in S_T)(2\pi\sigma_\epsilon^2)^{\frac{-nT^-}{2}}\exp\left\{\frac{-1}{2\sigma_\epsilon^2}\left\|\boldsymbol{\pi}^- -(X^-\boldsymbol{\theta}+\boldsymbol{U}^-)\right\|^2\right\}\\ (2\pi\sigma_{\epsilon^*}^2)^{\frac{-nT^+}{2}}\exp\left\{\frac{-1}{2\sigma_{\epsilon^*}^2}\left\|\boldsymbol{\pi}^+ -(X^+\boldsymbol{\theta}^*+\boldsymbol{U}^++\boldsymbol{V}^+)\right\|^2\right\}.\end{aligned} \tag{19}$$

3.3. COVID-19 数据中的多变点检测

在前两个小节中，我们从一般意义上讨论了所提出的变点检测模型的贝叶斯估计。这里，我们解释如何将该模型应用于第 2 节所描述的 COVID-19 数据。对于这一应用，如前所述，我们使用上一周新增死亡人数的对数变换作为非空间变化协变量。另一方面，第一剂 COVID-19疫苗的接种率被用作模型中的空间变异系数。令 log $d(s_i,t-1)$ 表示时间 t和位置 s_i处的第一个协变量，而 $\boldsymbol{x}(t)=(x(s_1,t)\ldots,x(s_n,t))'$表示该位置集合中疫苗接种率的 n 维向量。于是，潜变量的均值结构可以写为

$$\begin{aligned}\mu(s_i,t)&=\beta_0+\beta_1\log d(s_i,t-1)+\boldsymbol{\gamma}(s_i)x(s_i,t),\quad \text{for}\ \ t\leqslant t_0,\\ \mu^*(s_i,t)&=\beta_0^*+\beta_1^*\log d(s_i,t-1)+\gamma^*(s_i)x(s_i,t),\quad \text{for}\ t>t_0.\end{aligned} \tag{20}$$

对于与前一周死亡人数相关的协变量，预计其会对疾病传播产生正向作用。这与以下观点有关：因 COVID-19 导致的死亡对应着更高的病毒载量，从而增加了疾病传播给其他人的可能性（Pujadas 等人，2020年）。然而，随着越来越多的人接种疫苗，我们预计这种影响将变得微不足道，因为疫苗接种后死亡人数会减少。对于第二个协变量，我们预计在疫苗接种的第一阶段，其系数将显著为负，这将表明 COVID-19疫苗在降低大流行传播方面具有有效性。随着疫苗接种的推进，人们很可能会获得群体免疫。例如，MacIntyre、Costantino 和 Trent （2022年）在其研究中证明，澳大利亚在约 66% 的人口接种疫苗后实现了群体免疫。一旦达到这一状态，疫苗在解释疾病传播方面的作用就不应再那么显著。此外，受到 Utazi 等人（2018）关于不同国家麻疹疫苗接种空间变化影响的研究启发，我们假设纽约州不同县的疫苗作用可能并不相同，并且其与疾病传播之间的关系也存在差异。事实上，这正是我们对第一剂疫苗接种采用空间变化系数的主要动机。

我们提出的方法的另一个关键方面是能够评估所估计的变点是否确实具有显著性。按照贝叶斯文献中的惯例，我们采用贝叶斯因子。贝叶斯因子使我们能够比较结构中不存在任何变点的模型与包含变点的模型。令 $\mathcal{M}_1$ 表示我们提出的包含变点的模型（参见式（2）和式（3））

以及 $\mathcal{M}_2$ 通过拟合不含变点的模型（即针对整个数据，使用公式（3）中变点之前的第一行所对应的模型）得到的另一个模型。给定特定模型时，数据的似然为

$$\begin{aligned} f(\boldsymbol{Y}|\mathcal{M}_i) &= \int f(\boldsymbol{Y},\boldsymbol{\theta},\boldsymbol{\theta}^*,\boldsymbol{U},\boldsymbol{V}|\mathcal{M}_i)\mathrm{d}\boldsymbol{\theta}\mathrm{d}\boldsymbol{\theta}^*\mathrm{d}\boldsymbol{U}\mathrm{d}\boldsymbol{V} \\ &= \int f(\boldsymbol{Y}|\boldsymbol{\theta},\boldsymbol{\theta}^*,\boldsymbol{U},\boldsymbol{V},\mathcal{M}_i) f(\boldsymbol{\theta},\boldsymbol{\theta}^*,\boldsymbol{U},\boldsymbol{V}|\mathcal{M}_i)\mathrm{d}\boldsymbol{\theta}\mathrm{d}\boldsymbol{\theta}^*\mathrm{d}\boldsymbol{U}\mathrm{d}\boldsymbol{V} \\ &= \mathbb{E}_{\boldsymbol{\theta},\boldsymbol{\theta}^*,\boldsymbol{U},\boldsymbol{V}}\left[f(\boldsymbol{Y}|\boldsymbol{\theta},\boldsymbol{\theta}^*,\boldsymbol{U},\boldsymbol{V},\mathcal{M}_i)\right] \\ &\approx \frac{1}{m}\sum_{j=1}^{m} f\left(\boldsymbol{Y}|\boldsymbol{\theta}^{(j)},\boldsymbol{\theta}^{*(j)},\boldsymbol{U}^{(j)},\boldsymbol{V}^{(j)},\mathcal{M}_i\right), \end{aligned}$$

其中 $\boldsymbol{\theta}^{(j)} \sim f(\boldsymbol{\theta}|\mathcal{M}_i))$, $\boldsymbol{\theta}^{*(j)} \sim f(\boldsymbol{\theta}^*|\mathcal{M}_i))$, $\boldsymbol{U}^{*(j)} \sim f(\boldsymbol{U}|\mathcal{M}_i))$, $\boldsymbol{V}^{*(j)} \sim f(\boldsymbol{V}|\mathcal{M}_i))$ 是从模型特定分布中得到的实现值。有趣的是，在本例中，如果使用非信息性先验，似然值通常是难以解析求得的。正如 Newton 和 Raftery (1994) 以及 Risser 等人 (2019) 所建议的，在这种情况下，可以借助调和平均数利用参数的后验分布求得似然。具体而言，如果 $\boldsymbol{\beta}_{post}^{(j)}$ 表示从 j^{th}后验样本中得到的参数向量的实现值（其他参数同理），则似然可表示为

$$f(\boldsymbol{Y}|\mathcal{M}_i) \approx \left[\frac{1}{m}\sum_{j=1}^{m}\frac{1}{f(\boldsymbol{Y}|\boldsymbol{\beta}_{post}^{(j)},\boldsymbol{\beta}_{post}^{*(j)},\boldsymbol{U}_{post}^{(j)},\boldsymbol{V}_{post}^{(j)},\mathcal{M}_i)}\right]^{-1}. \tag{21}$$

计算出似然值后，我们使用贝叶斯因子来判断所估计的变点在本问题中是否显著。按照 Kass 和 Raftery (1995) 的方法，我们将贝叶斯因子的阈值设为100，以判断是否存在支持变点模型的决定性证据。回顾一下，模型 $\mathcal{M}_1$ 相对于 $\mathcal{M}_2$ 的贝叶斯因子为

$$BF_{12} = \frac{f(\boldsymbol{Y}|\mathcal{M}_1)}{f(\boldsymbol{Y}|\mathcal{M}_2)}. \tag{22}$$

上述过程不仅有助于判断是否存在单个显著变点，也有助于判断有序分类时空数据中是否存在多个变点。为此，我们提出一种二分分割型算法。文献中已采用该技术检测不同问题中的多个变点（Fryzlewicz, 2014；Cho 和 Fryzlewicz，2015年）。要在我们的研究情境中实现该算法，首先，我们运行前述过程，在整个时间段内查找是否存在变点。在本研究中所有吉布斯采样器的运行过程中，我们通过Gelman-Rubin诊断（Gelman 和 Rubin (1992）监测收敛情况，同时运行三条马尔可夫链。现在，如果贝叶斯因子方法提供了支持变点存在的决定性证据，我们就按照时间域将数据划分为两个区段。随后，在这两个区段上分别运行同一算法，并以递归方式继续，直到数据中不再有显著变点。从实际角度出发，作为一项额外的停止规则，我们规定由变点划分出的任何区段至少必须有三个月（相当于12周）长。因此，如果某个区段的数据不超过24周，我们就不再继续运行该算法。此外，如果发现变点过于接近任一端点，也会根据同一限制将其忽略。

我们在算法 1 中给出了有序分类时空数据多变点检测整个过程的伪代码。请注意，停止准则 C 意味着贝叶斯因子方法不支持变点的存在，或者每个分段的数据量均不超过所要求的数量。我们强调，这是一种主观选择，在其他考量因素下可以轻松调整。

4. 结果

如第2节所述，我们的目标是利用所提出的模型，考虑新冠肺炎传播的有序分类特性，以检测其时空变点。回顾一下，所用数据

算法 1： 时空分类数据中的多变点检测。

输入 ： 数据向量 $\boldsymbol{Y}$、设计矩阵 X 和停止准则 $\mathcal{C}$。
输出 ： 所有参数和变点的后验分布。
初始化： 设置 $\boldsymbol{U}=0$ 和 $\boldsymbol{V}=0$，为 ϕ_{us}、 ϕ_{ut}、 ϕ_{vs}、 ϕ_{vt}、ω_s、 ω_s^*、 $\boldsymbol{\beta}$、 $\boldsymbol{\beta}^*$、 δ_2、 δ_3、 t_0 取可行随机值，三条马尔可夫链。

当 $\mathcal{C}$ *not true* **时**
  (a) 按照上述说明实现模型 $\mathcal{M}_1$ （含变点），求出后验样本，估计参数并计算该模型的似然估计值。；
  (b) 按照上述说明实现模型 $\mathcal{M}_2$ （不含变点），求出后验样本，估计参数并计算该模型的似然估计值。；
  (c) 计算 $\mathcal{M}_1$ 相对于 $\mathcal{M}_2$的贝叶斯因子。；
  (d) 根据 $\mathcal{M}_1$中估计的变点，将数据集划分为两个片段。；
  (e)**如果** *Bayes factor* $>$ *100 & sufficient data are available for one or both segments.* **则**
    使用数据充足的片段中的数据；然后转到步骤(a)。；
  else
    跳出；
  end
结束

本研究的数据包括纽约州 62 个县 122 周的信息（20^{th} 2020年1月至 16^{th} 2022年5月）。本节首先讨论算法识别出的变点及其意义，随后深入分析协变量与大流行传播之间的关系。最后，我们考察疾病在不同时期传播过程中时空依赖的程度。

在此应用中，按照算法1中的步骤，所提出的方法检测到三个变点。表2 展示了分析各阶段估计的变点及相应的贝叶斯因子，包括依据判定标准未达到显著水平的变点。第一个变点估计位于 57^{th} 周，相应的贝叶斯因子高于临界值，表明有证据支持存在变点。在第二阶段，数据集被分为两个时间段，并将所提出的模型分别应用于每个时间段。对于第一个时间段，即数据中的 1^{st} 至 57^{th} 周，由于疫苗接种仅在该时间段末期（约第50周）才开始，因此我们不得不将其从协变量中移除。此处的贝叶斯因子同样支持存在变点，变点位于36^{th} 周。同样，对于第二个时间段，即 58^{th} 至 122^{nd} 周， 96^{th} 周被判定为显著变点。随后，在二元分割算法的第三阶段，当我们再次使用 1^{st} 至 36^{th} 周的数据拟合模型时，检测到的变点位于该时间段的前端，其贝叶斯因子的对数为 -72，表明结构中不存在变点。对 58^{th} 至 96^{th} 周的数据运行模型，也得出了类似的结论。与此同时，在 97^{th} 至 122^{nd} 周的时间段内，变点被发现离该时间段起点过近，因此根据我们的限制予以忽略。最后请注意，由于根据前文所述的假设，剩余时间段的数据不足，因此我们没有对其运行该算法。因此，该数据集中共有三个变点，分别位于 21^{st} 2020年9月、 15^{th} 2021年2月和 15^{th}2021年11月。

TABLE 2
Changepoints and respective Bayes factors for the proposed algorithm. The first column indicates the stages of binary segmentation. Bolded changepoints indicate that it was identified as present/significant in the data.

阶段	时间范围	变点	Date	对数贝叶斯因子
1	1^{st} 至 122^{nd} 周	**57^{th} 周**	15^{th} 2021年2月	44
2	1^{st} 至 57^{th} 周	**36^{th} 周**	21^{st} 2020年9月	64
	58^{th} 至 122^{nd} 周	**96^{th} 周**	15^{th} 2021年11月	1118
3	1^{st} 至 36^{th} 周	8^{th} 周	9^{th} 2020年3月	-72
	37^{th} 至 57^{th} 周	数据不足		
	58^{th} 至 96^{th} 周	79^{th} 周	19^{th} 2021年7月	-102
	97^{th} 至 122^{nd} 周	未找到		

为便于理解，图 4 展示了三个变点（以黑点表示）与该州新增病例总数的关系。图中显示， 15^{th} 2021年2月的变点出现在

美国 COVID-19 冬季阿尔法波次即将结束之际。这也是该国疫苗接种速度迅速加快的时期，纽约州约有 11% 的人口接种了第一剂。相比之下， 21^{st} 2020年9月的变点与该州同一波疫情即将开始的时期相吻合。这无疑是该州最艰难的时期之一，而这两个变点充分反映了这一点。随后我们注意到，德尔塔变异株流行期间，疾病传播方式并未发生显著变化；但奥密克戎变异株疫情波次开始时，病例数在1月底左右达到峰值，这与第三个变点 15^{th} 2021年11月完全吻合。这清楚地表明，所提出的模型有力地支持其能够正确检测纽约不同类型的 COVID-19 疫情波次。

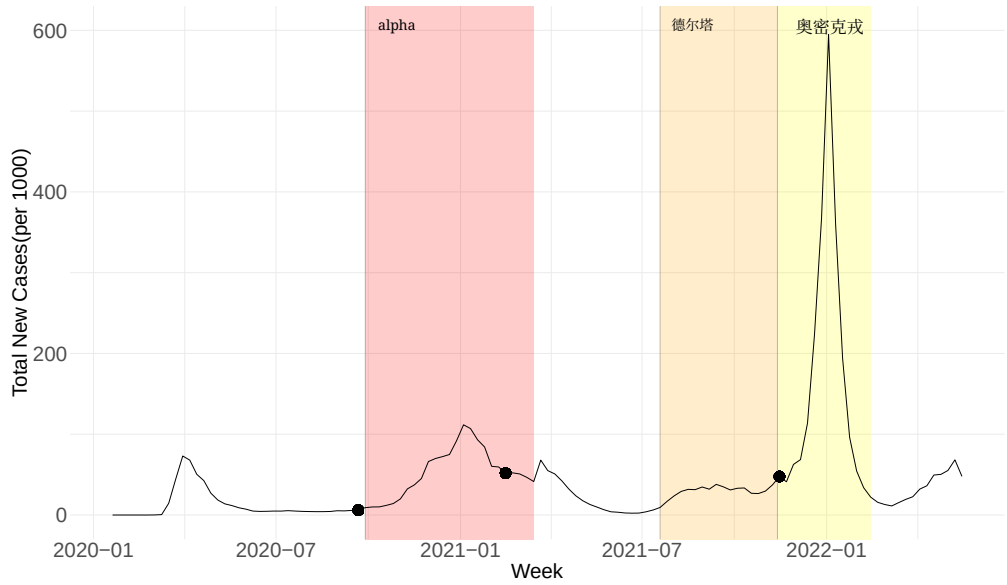


FIG 4. *COVID-19 waves corresponding to new cases for New York State. The graph's black dots denote the proposed model's estimated changepoint.*

接下来，我们讨论将所提出的模型拟合到不同时间段后得到的协变量效应大小。我们在表3中报告后验均值和95%可信区间。疫苗接种的空间变化系数如图3 所示，并将在后续段落中进行讨论。这些数值是根据变点所划分的时间段计算得出的，正如前文所述。值得注意的是，对于截距和 log $d(s_i, t-1)$，四个阶段的估计值存在相当大的差异。在大流行逐渐开始蔓延的第一阶段，截距明显较低。这是合理的，因为这一时期报告的 COVID-19 病例数相对较少。随后，在 37^{th} 至 57^{th} 周的时间段内，阿尔法变异株开始迅速传播。在此之前，疫苗尚未发挥显著的干预作用，因此估计得到的截距值相对较高。在接下来的阶段，随着越来越多的人接种疫苗，新增病例数有所下降，这也解释了为何估计截距值较低。与此相反，在最后一个阶段，潜变量的总体均值呈现出较高的水平。这可以通过 COVID-19 新增病例数的上升得到佐证：奥密克戎变异株逃逸了疫苗接种所诱导的免疫保护，并逐渐席卷全国。感兴趣的读者可以参阅Tian等人(2021年)以及Ao等人(2022年，了解这一方面的相关讨论。

对于 log $d(s_i, t-1)$ 的系数，最初在尚无疫苗可用时，该变量与 COVID-19 感染新增病例数的上升显著相关。事实上，在 1^{st} 至 36^{th}周期间，前一周死亡人数的系数在所有时期中最高，表明它与 COVID-19 的传播水平呈强烈正相关。在这一阶段，纽约州每周报告的死亡人数约为 6000 人。接下来的阶段，该州正经历阿尔法波。此阶段死亡人数有所下降，表明疾病严重程度降低，也自然解释了估计系数值为何较低。此外，在第三个时期，随着相当比例的人口接种了疫苗，我们看到死亡人数的系数估计值

TABLE 3
Parameter estimates and corresponding credible intervals when the proposed model is fitted in different segments.

变量	周范围	后验均值	95%可信区间
截距	1-36	1.079	(0.467, 1.615)
	37-57	7.713	(5.483, 9.525)
	58-96	6.279	(5.124, 7.406)
	97-122	7.978	(6.520, 9.321)
log（前一周死亡人数）	1-36	1.101	(0.679, 1.497)
	37-57	0.296	(0.071, 0.533)
	58-96	0.273	(0.157, 0.389)
	97-122	−0.146	(−0.357, 0.061)
空间衰减参数	1-57	0.0108	(0.0081,0.0129)
	58-122	0.0021	(0.0016,0.0026)
时间衰减参数	1-57	0.264	(0.206,0.315)
	58-122	0.203	(0.165,0.245)

显著下降。这或许可以视为疫苗的有效性。最后，在 97^{th} 至 122^{nd} 周期间，死亡人数的系数被发现并不显著，因为奥密克戎变异株导致重症的可能性较低，但传染性很强（这解释了该阶段较大的截距项）。因此，所提出的模型除了能够捕捉变点外，也恰当地捕捉到了这一效应。

现在，我们来关注疫苗接种率的影响。众所周知，在最初几天，疫苗仅提供给医院工作人员和老年人。因此，在57^{th} 周之前没有可用的相关数据，必须从模型中省略这一协变量。因此，我们在图5中展示了另外两个区段中疫苗接种的空间变异效应。图中绘出了各参数的后验均值，条纹表示可信区间与零充分远离。

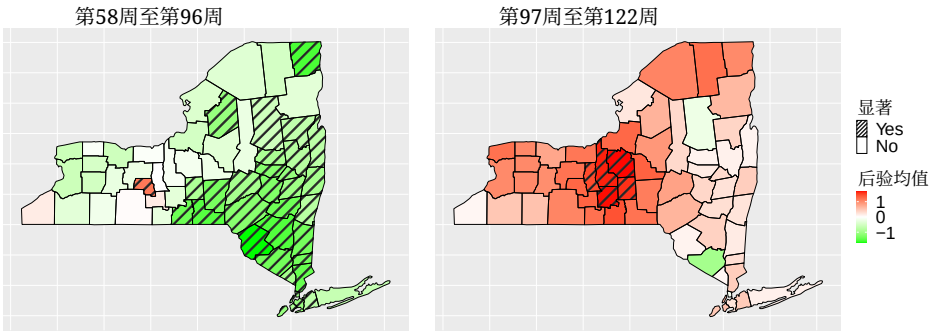


FIG 5. *Spatially varying coefficients (posterior means) for the vaccination prevalence in the last two segments obtained via the changepoints. Significance is equivalent to the credible interval being sufficiently away from zero.*

在 58^{th} 至 96^{th} 周期间，随着疫苗接种率的提高，纽约州所有县的 COVID-19 传播水平均显著下降。62个县中有33个县的这一系数具有统计显著性。这些县几乎涵盖了该州东部地区的所有县，而这些县的人口通常比其他地区更加稠密。因此，可以认为，在疫情初期，疫苗接种在人口较多的县中发挥了更大的作用。相比之下，在 97^{th} 至 122^{nd} 周期间，我们发现疫苗接种通常不具有统计显著性。一种可能的解释是群体免疫的影响。到 22^{nd}2021年11月，纽约州75%的人口已接种第一剂疫苗，而到 16^{th} 2022年5月，这一比例已超过85%。正如第3.3节所讨论的，要实现群体免疫，人口覆盖率需要达到约三分之二。根据世界卫生组织（2021年）的观点，为确保实现针对 COVID-19 的群体免疫，也应使70%的人口接种疫苗。因此，在达到这一人口覆盖率后，疫苗接种的影响便不再显著，直到奥密克戎变异株出现，且人群对其缺乏免疫抵抗力（Ao等人，2022年）。

最后一部分　　在讨论中，我们考察了所划分的四个区段中的时空依赖。

由变点决定。为此，图π（s_i， t）展示了不同时间段潜在过程估计的时空相关性热图6。相关性基于空间和时间衰减参数的后验均值计算，并表示为距离和时间滞后的函数。

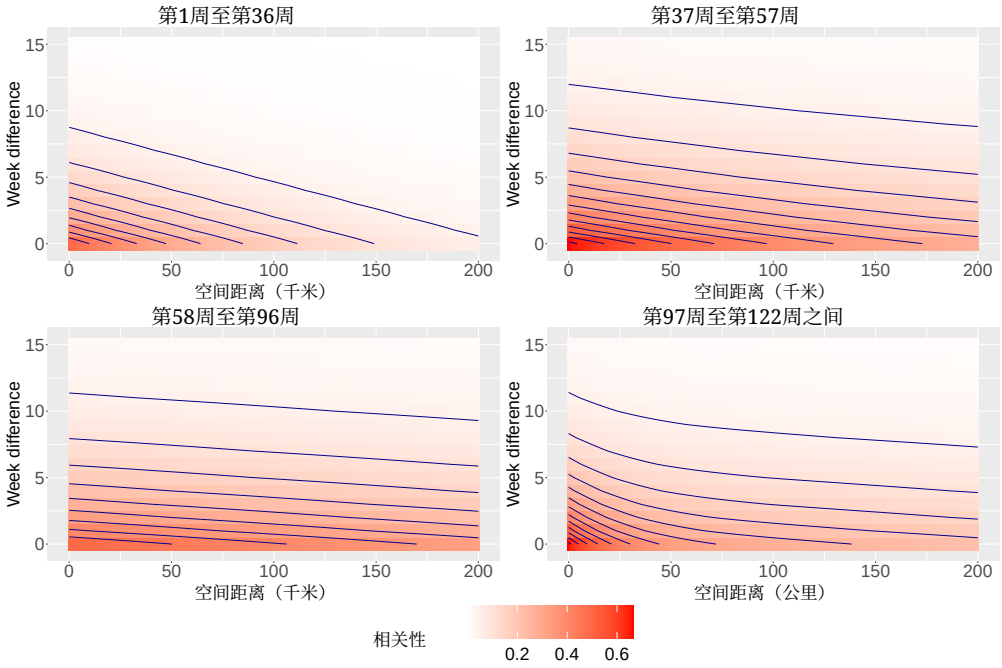


FIG 6. *Estimated correlation for the latent process* $\pi(s_i, t)$ *in four segments. The contour lines represent the points that have the same correlation values.*

与其他所有时期相比，第一个时期的相关性估计值最低。一个可能的原因是，COVID-19感染率仍相当低时，纽约很早就实施了封锁措施。之后，限制措施通过四阶段方案逐步放宽（Husch Blackwell，2021年）。因此，随着大流行的发展，以及阿尔法波期间感染率持续上升（37^{th} 至 57^{th}周），我们观察到较强的时空依赖。事实上，这是所有时期中相关性最高的时期。此后，在 58^{th} 至 96^{th} 周之间，从图中可以看出，空间依赖和时间依赖均有所减弱。这可能是因为在此期间，随着多种疫苗向不同年龄组的所有成年人开放，疫苗接种率开始在各地逐步提高。最后，我们看到最后一个时期的空间依赖和时间依赖模式进一步减弱。如前所述，该阶段流行的奥密克戎变异株比先前的变异株温和，不会导致严重疾病。可以认为，人口也已达到群体免疫。这可能是空间依赖减弱的原因；而时间相关性下降，则可能与奥密克戎变异株导致的疾病持续时间短于德尔塔变异株有关，正如Menni 等人（2022年）和Vitiello 等人（2022年）所阐述的那样。

5. 结论

在这项工作中，我们提出了一种新方法，用于检测有序分类时空数据集中的变点。我们还详细说明了各项模型参数的完整估计过程及相关注意事项。我们将模型应用于纽约州的 COVID-19 数据：依据美国疾病控制与预防中心的指南，将各县每周新增病例转换为有序类别，以表示 COVID-19 的传播水平。模型通过适当构建的潜在过程进行定义，可量化疾病传播中的时空依赖。模型采用完整的贝叶斯框架实现。我们的分析发现了大流行时空模式中有趣的变化，这些变化与特定病毒变异株的传染性密切相关。借助我们

提出的模型，我们还展示了在这段时间内，疫苗接种与不同县的传播水平之间的关系。

一个有趣的发现是，不同时间段的空间依赖程度存在差异，这可能与人员流动和疫苗接种率有关。由于无法获取纽约州各县之间的流动数据，这一问题超出了本文的研究范围。不过，进一步研究这一问题有助于更好地理解感染动态，也是对本研究的有价值拓展。同样，政府实施的限制措施和封锁措施预计会影响新冠肺炎的传播。由于缺乏适当粒度的相关数据，我们的分析尚未纳入与政策实施有关的协变量。这自然也为当前方法的进一步拓展提供了一个思路。

需要指出的是，我们的方法应用于 COVID-19 数据，不仅有助于深入认识这一改变世界的事件，所提出的模型也有望应用于其他领域。例如，该模型可用于流行病学研究（如其他传染病的传播水平）、环境研究（如降水量水平或土壤有机碳浓度）、教育研究（如随时间变化的学校成绩）、社会网络研究（如个人或公司的连通性水平）等。

最后，我们指出，本文提出的方法在数据集较大时会面临计算负担方面的局限，例如，如果希望在单个模型中考察美国所有县。特别值得探究的是，估计全国不同区域的变点，并评估不同区域的新冠肺炎传播在不同时间段有多相似或不同。尽管我们的方法在理论上可行，但在这类情况下，要将模型应用于如此大规模的数据，就需要对贝叶斯算法进行大量近似。

参考文献

Albert, J. H.和 Chib, S. （1993年）。二元和多分类响应数据的贝叶斯分析。《美国统计学会会刊》88 669–679。Altieri, L., Scott, E. M., Cocchi, D. 和 Illian, J. B. （2015年）。时空点过程的变点分析。《空间统计》14 197–207。Altieri, L., Cocchi, D., Greco, F., Illian, J. B. 和 Scott, E. （2016年）。用于时空点过程变点分析的贝叶斯 P 样条和 R 高级计算。《统计计算与模拟杂志》86 2531–2545。Ao, D., Lan, T., He, X., Liu, J., Chen, L., Baptista-Hon, D. T., Zhang, K. 和 Wei, X.（2022年）。SARS-CoV-2 奥密克戎变异株：免疫逃逸与疫苗开发。MedComm 3 e126。Benoit, K. （2012年）。多项逻辑回归与有序逻辑回归。检索日期：2017年12月10日。Berrett, C.和 Calder, C. A. （2012年）。贝叶斯空间 Probit 回归模型的数据增广策略。《计算统计与数据分析》56 478–490。Berrett, C., Gurney, B., Arthur, D., Moon, T. 和 Williams, G. P.（2023年）。用于识别与城市化相关的局部温度变化的贝叶斯变点建模方法。《环境计量学》34 e2794。Bertozzi, A. L., Franco, E., Mohler, G., Short, M. B. 和 Sledge, D. （2020年）。新冠肺炎传播建模与预测面临的挑战。《美国国家科学院院刊》117 16732–16738。Husch Blackwell（2021年）。纽约：各州新冠肺炎指南。https://www. huschblackwell.com/new-york-state-by-state-covid-19-guidance。访问日期：2023年3月。Bradley, J. R., Holan, S. H. 和 Wikle, C. K. （2015年）。用于高维区域数据的多变量时空模型及其在纵向雇主—家庭动态中的应用。《应用统计年鉴》9 1761 – 1791。Brockwell, P. J. . 和 Davis, R. A.（ 1991 年）。时间序列：理论与方法。Springer-Verlag。Campbell, M. K. 和 Donner, A.（1989年）。多项逻辑回归相对于有序逻辑回归的分类效率。《美国统计学会会刊》84 587–591。Chen, M.-H. 和 Dey, D. K. （2000年）。用于分析相关有序响应数据的统一贝叶斯方法。《巴西概率与统计杂志》87–111。

Chernoff, H. 和 Zacks, S. （1964年）。估计随时间发生变化的正态分布的当前均值。《数学统计年刊》35 999–1018。Cho, H. 和 Fryzlewicz, P. （2015年）。通过稀疏化二分分割检测高维时间序列中的多变点。《皇家统计学会期刊：B辑：统计方法论》475–507。Cowles, M. K. （1996年）。加速累积链接广义线性模型的蒙特卡罗马尔可夫链收敛。统计与计算 6 101–111。Daykin, A. R. 和 Moffatt, P. G. （2002年）。分析有序响应：有序概率单位模型综述。《理解统计学：心理学、教育学与社会科学中的统计问题》1 157–166。Dubrow, J. K. （2022年）。本地数据和上游报告作为新冠肺炎行政数据低估误差的来源。国际社会研究方法学杂志 25 471–476。Fiedler, B., Zöller, G., Holschneider, M. 和 Hainzl, S. （2018年）。时空地震活动数据中的多变点检测时空地震活动数据中的多变点检测。《美国地震学会通报》108 1147–1159。FitzJohn, R. G. （2012年）。Diversitree：R中多样化的比较系统发育分析。生态学与进化方法 3 1084–-1092。Centers for Disease Control and Prevention （2022年）。用于监测COVID-19社区级别和COVID-19并实施COVID-19预防策略的指标。https://www.cdc.gov/coronavirus/2019-ncov/downloads/science/Scientific-Rationale- summary_COVID-19-Community-Levels_2022.02.23.pptx。访问日期：2022年2月。Fryzlewicz, P. （2014年）。用于多变点检测的野生二分分割。Gelfand, A. E. （2000年）。吉布斯抽样。《美国统计学会会刊》95 1300–1304。Gelman, A. 和 Rubin, D. B. （1992年）。使用多条序列从迭代模拟中进行推断。统计科学 457–472。Geman, S. 和 Geman, D. （1984年）。随机松弛、吉布斯分布与图像贝叶斯复原。IEEE模式分析与机器智能汇刊 6 721–741。Heiner, M., Heaton, M. J., Abbot, B., White, P., Minaudo, C. 和 Dupas, R. （ 2022 年）。法国各地河流硝酸盐浓度趋势与周期的基于模型的聚类。《农业、生物与环境统计学杂志》。Higgs, M. D. 和 Hoeting, J. A. （2010年）。空间相关有序分类数据的截断潜变量模型。《计算统计与数据分析》54 1999–2011。Imai, K. 和 van Dyk, D. A. （2005年）。使用边际数据增广对多项概率单位模型进行贝叶斯分析。计量经济学杂志 124 311–334。JHU-CSSE, D. （2020年）。约翰斯·霍普金斯大学CSSE的2019新型冠状病毒COVID-19（2019-nCoV）数据仓库。最后访问日期：2022年5月16日。Jin, X., Carlin, B. P. 和 Banerjee, S. （2005年）。用于区域数据的广义分层多变量CAR模型。生物统计学 61 950–961。Kass, R. E. 和 Raftery, A. E. （1995年）。贝叶斯因子。《美国统计学会会刊》90 773–795。Kim, S., Shephard, N. 和 Chib, S. （1998年）。随机波动率：似然推断及与ARCH模型的比较。《经济研究评论》65 361–393。Krishnaiah, P. 和 Miao, B. （1988年）。19 变点评估综述。统计学手册 7 375–402。Liang, J., Bi, G. 和 Zhan, C. （2020年）。使用R对含多分类变量的多项和有序逻辑回归进行分析。《转化医学年鉴》8。Liu, J. S. 和 Wu, Y. N. （1999年）。用于数据增广的参数扩展。《美国统计学会会刊》94 1264–1274。MacIntyre, C. R., Costantino, V. 和 Trent, M. （2022年）。在澳大利亚新南威尔士州疫苗供应有限和充足的情景下，对COVID-19疫苗接种策略和群体免疫进行建模。《疫苗》40 2506–2513。Majumdar, A., Gelfand, A. E. 和 Banerjee, S. （2005年）。时空变点建模。统计规划与推断杂志 130 149–166。Menni, C., Valdes, A. M., Polidori, L., Antonelli, M., Penamakuri, S., Nogal, A., Louca, P.,

May, A., Figueiredo, J. C., Hu, C. 等（2022年）。奥密克戎和德尔塔变异株占主导时期感染SARS-CoV-2的个体的症状患病率、持续时间及住院风险：一项来自 ZOE COVID 研究的前瞻性观察研究。《柳叶刀》399 1618–1624。Mullahy, J.（1997年）。异质性、过多零值与计数数据模型的结构。《应用计量经济学杂志》12 337–350。Nandram, B. 和 Chen, M.-H.（1996年）。重新参数化广义线性模型以加快吉布斯采样器的收敛。《统计计算与模拟杂志》54 129–144。Neal, R. M.（2003年）。切片抽样。《统计年鉴》31 705–767。Newton, M. A. 和 Raftery, A. E.（1994年）。利用加权似然自助法进行近似贝叶斯推断。《英国皇家统计学会杂志：B辑（方法论）》56 3–26。
World Health Organization（2021年）。到2022年年中实现70%的 COVID-19 免疫接种覆盖率。https://www.who.int/news/item/23-12-2021-achieving-70-COVID-19-immunization-coverage-by-mid-2022。访问日期：2023年3月。Pujadas, E., Chaudhry, F., McBride, R., Richter, F., Zhao, S., Wajnberg, A., Nadkarni, G.,Glicksberg, B., Houldsworth, J. 和 Cordon-Cardo, C.（2020年）。SARS-CoV-2 病毒载量可预测 COVID-19 死亡率。medRxiv。Rawat, S. 和 Deb, S.（2021年）。用于分析美国 COVID-19 传播的时空统计模型。《应用统计学杂志》1–20。Ricon-Becker, I., Tarrasch, R., Blinder, P. 和 Ben-Eliyahu, S.（2020年）。COVID-19 感染、住院和死亡率的七日周期：周末社交互动会导致易感人群死亡吗？medRxiv。Risser, M. D., Calder, C. A., Berrocal, V. J. 和 Berrett, C.（2019年）。土壤有机碳的非平稳空间预测。《应用统计年鉴》13 165–188。Roy, V. 和 Hobert, J. P.（2007年）。贝叶斯概率单位回归的马尔可夫链蒙特卡罗算法的收敛速度和渐近标准误。《英国皇家统计学会杂志：B辑（统计方法）》69 607–623。Sahu, S. K., Gelfand, A. E. 和 Holland, D. M.（2006年）。细颗粒物的时空建模。《农业、生物与环境统计学杂志》11 61–86。Tanner, M. A. 和 Wong, W. H.（1987年）。通过数据增广计算后验分布。《美国统计学会会刊》82 528–540。Taylor, L. R.（1961年）。聚合、方差与均值。《自然》189 732–735。Taylor, S. J.（1994年）。随机波动率建模：综述与比较研究。《数学金融》4 183–204。Tian, D., Sun, Y., Zhou, J. 和 Ye, Q.（2021 年）。SARS-CoV-2 德尔塔变异株的全球疫情、关键刺突蛋白突变及免疫逃逸。《免疫学前沿》12。Utazi, C. E., Thorley, J., Alegana, V. A., Ferrari, M. J., Takahashi, S., Metcalf, C. J. E.,Lessler, J. 和 Tatem, A. J.（2018年）。中低收入国家儿童疫苗接种覆盖率的高分辨率年龄分层制图。《疫苗》36 1583–1591。Vitiello, A., Ferrara, F., Auti, A. M., Di Domenico, M. 和 Boccellino, M.（2022年）。奥密克戎变异株的发展进展。《内科学杂志》292 81–90。Xu, Y., Yu, Q., Scribner, R., Theall, K., Scribner, S. 和 Simonsen, N.（2012年）。用于评估酒类销售点管控政策对社区暴力袭击率变化影响的多层时空变点模型。《空间与时空流行病学》3 121–128。Yu, Q., Scribner, R., Carlin, B., Theall, K., Simonsen, N., Ghosh-Dastidar, B., Cohen, D.和 Mason, K.（2008年）。用于关联酒类销售点拆除与社区暴力袭击率变化的多层时空双变点模型。《地理空间健康》2 161。Zacks, S.（1983年）。经典方法与贝叶斯方法在变点问题上的综述：固定样本与序贯检验和估计程序。载于《统计学最新进展》245–269。爱思唯尔。

附录 A：各县探索性分析

在美国，各州之间、县与县之间，甚至城市与城市之间，对 COVID-19 疫情的应对方式和管控政策都存在很大差异。因此，虽然我们的分析旨在考察纽约州整体的传播水平，但我们首先探讨了针对各县单独开展分析是否更为合适。

为此，我们将所提出的模型分别拟合到每个县，并移除空间依赖（或者固定$\Sigma_{us} = 1$）。具体而言，$\mathbf{U} \sim N(\mathbf{0}, \sigma_U^2($ $\Sigma_{ut} \otimes \Sigma_{us}))$ 被简化为 $\mathbf{U} \sim N(\mathbf{0}, \sigma_U^2 \Sigma_{ut})$，$\mathbf{V}$ 也进行了类似处理。

我们首先发现，一些县（约一半）的模型未能在预先设定的时间内收敛。因此，本文提出的允许从邻近地区借用信息的时空模型非常重要。此外，由于单个县的模型难以收敛，我们也仅拟合这些模型来估计一个变点；没有足够的信息来估计多个（如果有的话）变点。

其次，我们发现，对于确实实现收敛的县，变点和系数的估计不确定性区间非常大，并且与其他县的区间有很大重叠，这进一步说明有必要对所有县进行联合分析。我们在此对四个县进行更详细的考察。

我们研究了该州不同地区的四个县，这些县的人口规模各不相同。具体而言，阿勒格尼县（西南部，农村地区，人口46091人）、金斯县（东南部，城市地区，人口2559903人）、尼亚加拉县（西北部，城市/郊区，人口209281人）以及伦斯勒县（东北部，城市/郊区，人口158714人）。这四个县的变点后验均值列于表A.1中。图A.1展示了这四个县经过平滑处理（30周平均值）的传播水平曲线，其中垂直线表示变点均值。值得注意的是，这四个县估计出的变点均早于任何县开始实施疫苗接种的时间（纽约州的疫苗接种始于第50周左右，即2020年12月28日）。模型能够识别出这一时期关系发生变化是合理的；然而，这对于识别和量化更具实际意义的变化并无帮助。具体而言，考虑表 β_2A.1中变点之前疫苗接种的系数估计值。这些系数无论正负，数值都非常大，尤其是相对于Probit链接函数而言。这是因为在变点之前尚未开展疫苗接种，因此后验分布实际上只是返回先验分布[1]。然而，在变点之后，一旦开始接种疫苗，系数β_2^* 就会小得多，但其不确定性区间仍然相互重叠，且包含0。对于前几周死亡人数的系数，我们也观察到了类似的行为，但不存在一个所有县的这一协变量都保持不变的共同时间点。这些挑战表明，我们需要一种能够借鉴邻近地区信息的模型。

TABLE A.1 *Posterior mean (and 95% credible interval) for: the week of the change point, the corresponding date of the change point, and the change point for four counties when the model was fit to the single county data.*

县	变化点	Date	前一周死亡人数 β_1	前一周死亡人数 β_1^*	疫苗接种 β_2	疫苗接种 β_2^*
阿勒格尼县	38 (35, 38)	9月28日 2020	0.743 (-0.11, 1.75)	0.404 (-0.02, 0.89)	7715.6 (-1714, 17610)	-2.97 (-9.65, 2.90)
金斯县	28 (15, 43)	7月27日 2020	37.686 (0.62, 60.63)	0.369 (-0.29, 0.98)	2507.52 (-23402, 45493)	0.55 (-3.73, 4.43)
尼亚加拉	41 (40, 42)	10月26日 2020	0.9 (0.14, 2.07)	0.575 (0.08, 1.13)	-2028.81 (-11313, 7616)	-5.137 (-12.92, 0.23)
伦斯勒 ಜಿಲ್ಲೆ ಯ	37 (11, 44)	9月21日 2020	678.57 (-6053, 15661)	0.266 (-0.11, 0.69)	-1395.54 (-12352, 7351)	-1.833 (-8.52, 4.60)

[1]请注意，为避免变点之前的 X 矩阵出现奇异性，我们在协变量的取值中加入了少量噪声。

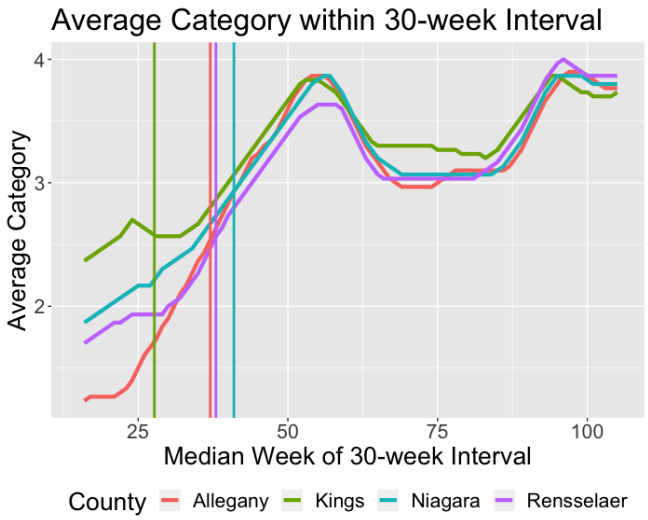


FIG A.1. *30-week average transmission levels for four counties in New York. The vertical line in the corresponding color represents the posterior mean change point estimated when the model was fit to the single county data.*

附录 B：后验计算

本节介绍吉布斯采样器所需后验分布的详细计算过程。完整的联合后验分布可写为

$$\begin{aligned} f(\boldsymbol{\theta},\boldsymbol{\theta}^*,\boldsymbol{U},\boldsymbol{V},\boldsymbol{\pi},\boldsymbol{\phi},\boldsymbol{\delta},t_0|\boldsymbol{Y}) &\propto f(\boldsymbol{Y}|\boldsymbol{\theta},\boldsymbol{\theta}^*,\boldsymbol{U},\boldsymbol{V},\boldsymbol{\pi},\boldsymbol{\delta},\boldsymbol{\phi},t_0)f(\boldsymbol{\pi}|\boldsymbol{U},\boldsymbol{V},\boldsymbol{\theta},\boldsymbol{\theta}^*,\boldsymbol{\delta},\boldsymbol{\phi},t_0)f(\boldsymbol{V}|\boldsymbol{\phi_v}) \\ &\quad f(\boldsymbol{U}|\boldsymbol{\phi}_u)f(\boldsymbol{\theta}|\omega_s)f(\boldsymbol{\theta}^*|\omega_s^*)f(t_0)f(\boldsymbol{\phi})f(\omega_s)f(\omega_s^*)f(\boldsymbol{\delta}) \\ &\propto f(\boldsymbol{Y}|\boldsymbol{\theta},\boldsymbol{\theta}^*,\boldsymbol{U},\boldsymbol{V},\boldsymbol{\pi},\boldsymbol{\phi},t_0)f(\boldsymbol{\pi}^-|\boldsymbol{U}^-,\boldsymbol{\theta},\boldsymbol{\theta}^*,t_0) \\ &\quad f(\boldsymbol{\pi}^+|\boldsymbol{U}^+,\boldsymbol{V}^+,\boldsymbol{\theta}^*,t_0)f(\boldsymbol{V}|\boldsymbol{\phi_v})f(\boldsymbol{U}|\boldsymbol{\phi}_u)f(\boldsymbol{\theta}|\boldsymbol{\omega}_s) \\ &\quad f(\boldsymbol{\theta}^*|\boldsymbol{\omega}_s^*)f(t_0)f(\boldsymbol{\phi})f(\boldsymbol{\omega}_s)f(\boldsymbol{\omega}_s^*)f(\boldsymbol{\delta}), \end{aligned}$$

其中 $\boldsymbol{\phi}=(\boldsymbol{\phi}_u,\ \boldsymbol{\phi}_v)'$、$\boldsymbol{\phi}_u=(\phi_{us},\ \phi_{ut})'$ 和 $\boldsymbol{\phi}_v=(\phi_{vs},\ \phi_{vt})'$。

为了计算 $\boldsymbol{\theta}$ 的条件后验分布，我们只需保留包含 $\boldsymbol{\beta}$ 的项。化简后可得

$$\begin{aligned} f(\boldsymbol{\theta}|\mathcal{F}) &\propto f(\boldsymbol{\pi}^-|\boldsymbol{U}^-,\boldsymbol{\theta},\boldsymbol{\theta}^*,t_0)f(\boldsymbol{\theta}|\omega_s) \\ &\propto \exp\left\{\frac{-1}{2}\left(\boldsymbol{\theta}'\left[\frac{X^{-'}X^-}{\sigma_\epsilon^2}\right]\boldsymbol{\theta}-2\boldsymbol{\theta}'\left[\frac{X^{-'}(\boldsymbol{\pi}^- -\boldsymbol{U}^-)}{\sigma_\epsilon^2}\right]\right)\right\}\exp\left\{\frac{-1}{2}\left(\boldsymbol{\theta}'\left[\frac{\Psi^{-1}}{k}\right]\boldsymbol{\theta}\right)\right\} \\ &\propto \exp\left\{\frac{-1}{2}\left(\boldsymbol{\theta}'\left[\frac{X^{-'}X^-}{\sigma_\epsilon^2}+\frac{\Psi^{-1}}{k}\right]\boldsymbol{\theta}-2\boldsymbol{\theta}'\left[\frac{X^{-'}(\boldsymbol{\pi}^- -\boldsymbol{U}^-)}{\sigma_\epsilon^2}\right]\right)\right\}, \end{aligned}$$

其中 Ψ 是块对角矩阵，其对角元素依次为 $(\mathbf{I}_G,\ \Omega_{s_1},\ \Omega_{s_2},\ \ldots\ldots,\ \Omega_{s_H})$。

与上述多元正态分布进行比较，可得：

$$\boldsymbol{\theta}|\mathcal{F}\sim N_k\left(\Sigma_{\boldsymbol{\theta}}\left[\frac{X^{-'}(\boldsymbol{\pi}^- -\boldsymbol{U}^-)}{\sigma_\epsilon^2}\right],\Sigma_{\boldsymbol{\theta}}\right), \tag{B.1}$$

其中，离散矩阵 $\Sigma_{\boldsymbol{\theta}}$ 定义为

$$\Sigma_{\boldsymbol{\theta}}=\left[\frac{X^{-'}X^-}{\sigma_\epsilon^2}+\frac{\Psi^{-1}}{k}\right]^{-1}.$$

同样，我们得到 $\boldsymbol{\theta}^*$ 的条件后验。

对于向量 $\boldsymbol{\pi}$ 的条件后验，我们考虑其第 i^{th} 个元素（如前所述，记为 π_i），对于向量 $\boldsymbol{U}$、$\boldsymbol{V}$ 和 $\boldsymbol{Y}$，也采用类似的方法。由于向量 $\boldsymbol{\pi}$ 的所有元素在给定 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 向量的条件下彼此独立，因此我们可以写出：

$$f(\pi_i|\mathcal{F}) \propto \begin{cases} f(\pi_i|u_i,\boldsymbol{\theta})\, I(\delta_{j-1}<\pi_i<\delta_j), & \text{for } t(i)\leqslant t_0, y_i=j, \\ f(\pi_i|u_i,v_i,\boldsymbol{\theta}^*)\, I(\delta_{j-1}<\pi_i<\delta_j), & \text{for } t(i)>t_0, y_i=j. \end{cases}$$

因此，我们可以将 π_i 写为条件后验

$$\pi_i|\mathcal{F} \sim \begin{cases} TN(x_i'\boldsymbol{\theta}+u_i,\sigma_\epsilon^2;\delta_{j-1},\delta_j), & \text{for } t(i)\leqslant t_0, y_i=j, \\ TN(x_i'\boldsymbol{\theta}^*+u_i+v_i,\sigma_{\epsilon^*}^2;\delta_{j-1},\delta_j), & \text{for } t(i)>t_0, y_i=j, \end{cases} \tag{B.2}$$

其中，$TN(a,b;\,c,d)$ 表示单变量截断正态分布，该分布由均值为 a、方差为 b 的正态分布在区间 (c,d) 内截断得到。

接下来，由于 δ_j 应位于 δ_{j-1} 和 δ_{j+1}之间，δ_j 的条件后验分布可写为

$$\begin{aligned} f(\delta_j|\mathcal{F}) &\propto f(\boldsymbol{Y}|\boldsymbol{\theta},\boldsymbol{\theta}^*,\boldsymbol{U},\boldsymbol{V},\boldsymbol{\pi},\boldsymbol{\delta},\boldsymbol{\phi},t_0)f(\boldsymbol{\delta}) \\ &\propto I(\delta_{j-1}<\delta_j<\delta_{j+1}) \int\limits_{\mathcal{A}_j} \left(\exp\left\{ \frac{-1}{2\sigma_\epsilon^2} \left\| \boldsymbol{\pi}^- - X^-\boldsymbol{\theta} - \boldsymbol{U}^- \right\|^2 \right\} \mathrm{d}\boldsymbol{\pi}^- \right. \\ &\quad \int\limits_{\mathcal{B}_j} \exp\left\{ \frac{-1}{2\sigma_{\epsilon^*}^2} \left\| \boldsymbol{\pi}^+ - X^+\boldsymbol{\theta}^* - \boldsymbol{U}^+ - \boldsymbol{V}^+ \right\|^2 \right\} \mathrm{d}\boldsymbol{\pi}^+, \end{aligned}$$

其中，$\mathcal{A}_j$ 和 $\mathcal{B}_j$ 分别表示对 j^{th} 类别中所有观测值的积分，其积分上下限为 δ_{j-1} 和 δ_j；而 $(j+1)^{th}$ 类别的积分上下限为 δ_j 和 δ_{j+1}，分别对应变点 t_0之前和之后。

现在，令 $\boldsymbol{\pi}_j^-$ 和 $\boldsymbol{\pi}_{j+1}^-$ 表示向量 $\boldsymbol{\pi}^-$ 中仅包含以下值的向量，其中向量 $\boldsymbol{Y}^-$ 的对应类别分别为 j 和 $j+1$。类似地，定义 X_j^-、$\boldsymbol{U}_j^-$、X_{j+1}^-和 $\boldsymbol{U}_{j+1}^-$。此外，令 $\boldsymbol{\pi}_j^+$ 和 $\boldsymbol{\pi}_{j+1}^+$ 表示向量 $\boldsymbol{\pi}^+$ 中仅包含以下值的向量，其中向量 $\boldsymbol{Y}^+$ 的对应类别分别为 j 和 $j+1$。类似地，定义 X_j^+、$\boldsymbol{U}_j^+$、$\boldsymbol{V}_j^+$、X_{j+1}^+、$\boldsymbol{U}_{j+1}^+$和 $\boldsymbol{V}_{j+1}^+$。

$$\begin{aligned} f(\delta_j|\mathcal{F}) \propto I(\delta_{j-1}<\delta_j<\delta_{j+1}) & \int\limits_{\delta_{j-1}}^{\delta_j} \exp\left\{ \frac{-1}{2\sigma_\epsilon^2} \left\| \boldsymbol{\pi}_j^- - X_j^-\boldsymbol{\theta} - \boldsymbol{U}_j^- \right\|^2 \right\} \mathrm{d}\boldsymbol{\pi}_j^- \\ & \int\limits_{\delta_j}^{\delta_{j+1}} \exp\left\{ \frac{-1}{2\sigma_\epsilon^2} \left\| \boldsymbol{\pi}_{j+1}^- - X_{j+1}^-\boldsymbol{\theta} - \boldsymbol{U}_{j+1}^- \right\|^2 \right\} \mathrm{d}\boldsymbol{\pi}_{j+1}^- \\ & \int\limits_{\delta_{j-1}}^{\delta_j} \exp\left\{ \frac{-1}{2\sigma_{\epsilon^*}^2} \left\| \boldsymbol{\pi}_j^+ - X_j^+\boldsymbol{\theta}^* - \boldsymbol{U}_j^+ - \boldsymbol{V}_j^+ \right\|^2 \right\} \mathrm{d}\boldsymbol{\pi}_j^+ \\ & \int\limits_{\delta_j}^{\delta_{j+1}} \exp\left\{ \frac{-1}{2\sigma_{\epsilon^*}^2} \left\| \boldsymbol{\pi}_{j+1}^+ - X_{j+1}^+\boldsymbol{\theta}^* - \boldsymbol{U}_{j+1}^+ - \boldsymbol{V}_{j+1}^+ \right\|^2 \right\} \mathrm{d}\boldsymbol{\pi}_{j+1}^+, \end{aligned} \tag{B.3}$$

现在，我们计算衰减参数 ϕ_{vt}的条件后验；其他衰减参数的条件后验也可以用类似的方法计算。正如主论文中所述，假设衰减参数的取值介于 0 和 3 之间。这是因为，对于空间或时间上的单位距离，衰减参数达到 3 或更大时，相关性便可忽略不计。使用这一先验，

ϕ_{vt} 的条件后验可写为

$$\begin{aligned} f(\phi_{vt}|\mathcal{F}) &\propto f(\boldsymbol{V}|\phi_{vt})f(\phi_{vt}) \\ &\propto |\Sigma_{vt}|^{-n/2}\exp\left\{\frac{-\boldsymbol{V}'(\Sigma_{vt}^{-1}\otimes\Sigma_{vs}^{-1})\boldsymbol{V}}{2\sigma_v^2}\right\}I(0<\phi_{vt}<3). \end{aligned} \tag{B.4}$$

如主论文第3.2节所述，我们将向量 $\boldsymbol{V}$ 分为 $\boldsymbol{V}_1, \boldsymbol{V}_2, \ldots, \boldsymbol{V}_n$，其中 $\boldsymbol{V}_t$ 是主论文式（3.4）中 V 矩阵的 t^{th} 列。因此，

$$\boldsymbol{V}_t|\boldsymbol{V}_{-t} \sim N_T(\boldsymbol{\mu}_{vc}, \sigma_v^2\Sigma_{vc}),$$

其中 $\boldsymbol{\mu}_{vc} = (\Sigma_{12}\Sigma_{22}^{-1}\otimes \mathbf{I}_n)\boldsymbol{V}_{-t}$ 和 $\Sigma_{vc} = (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}) \otimes \Sigma_{vs}$。然后，对于 $\boldsymbol{V}_t$ 的条件后验分布，如果 t^{th} 时间点大于 t_0，我们注意到

$$f(\boldsymbol{V}_t|\mathcal{F}) \propto f(\boldsymbol{\pi}^+|\boldsymbol{U}^+, \boldsymbol{V}^+, \boldsymbol{\theta}^*, t_0)f(\boldsymbol{V}_t|\boldsymbol{V}_{-t}),$$

从而意味着

$$\begin{aligned} f(\boldsymbol{V}_t|\mathcal{F}) \quad \propto \quad & \exp\left\{-\frac{1}{2\sigma_{\epsilon^*}^2}\left\|\boldsymbol{\pi}^+ - X^+\boldsymbol{\theta}^* - \boldsymbol{U}^+ - \boldsymbol{V}^+\right\|^2\right\} \\ & \exp\left\{-\frac{1}{2\sigma_v^2}(\boldsymbol{V_t} - \boldsymbol{\mu}_{vc})'\Sigma_{vc}^{-1}(\boldsymbol{V}_t - \boldsymbol{\mu}_{vc})\right\}. \end{aligned}$$

仅保留涉及 $\boldsymbol{V}_i$ 的项，上式可以证明与以下表达式成正比。

$$\begin{aligned} f(\boldsymbol{V}_t|\mathcal{F}) \quad &\propto \quad \exp\left\{-\frac{1}{2\sigma_{\epsilon^*}^2}(\boldsymbol{V}_t'\boldsymbol{V}_t - 2\boldsymbol{V}_i'(\boldsymbol{\pi}_t - X_t\boldsymbol{\theta}^* - \boldsymbol{U}_t)) - \frac{1}{2\sigma_v^2}(\boldsymbol{V}_t'\Sigma_{vc}^{-1}\boldsymbol{V}_t - 2\boldsymbol{V}_t'\Sigma_{vc}^{-1}\boldsymbol{\mu}_{vc})\right\} \\ &\propto \quad \exp\left\{-\frac{1}{2}\left(\boldsymbol{V}_t'\left(\frac{\Sigma_{vc}^{-1}}{\sigma_v^2} + \frac{\mathbf{I}_n}{\sigma_{\epsilon^*}^2}\right)\boldsymbol{V}_t - 2\boldsymbol{V}_t'\left(\frac{\Sigma_{vc}^{-1}\boldsymbol{\mu}_{vc}}{\sigma_v^2} + \frac{\boldsymbol{\pi}_t - X_t\boldsymbol{\theta}^* - \boldsymbol{U}_t}{\sigma_{\epsilon^*}^2}\right)\right)\right\} \end{aligned}$$

通过多元正态理论进行直接计算，并考虑另一种情况，即 i^{th}时间点小于或等于 t_0，对 $\boldsymbol{U}$ 进行类似计算可得

$$\boldsymbol{V}_t|\mathcal{F} \sim \begin{cases} N_n(\mu_{vc}, \sigma_v^2\Sigma_{vc}), & \text{for } t \leqslant t_0, \\ N_n\left(\Sigma_{V_t}\left(\frac{\Sigma_{vc}^{-1}\boldsymbol{\mu}_{vc}}{\sigma_v^2} + \frac{\boldsymbol{\pi}_t - X_t\boldsymbol{\theta}^* - \boldsymbol{U}_t}{\sigma_{\epsilon^*}^2}\right), \Sigma_{V_t}\right), & \text{for } t > t_0, \end{cases} \tag{B.5}$$

$$\boldsymbol{U}_t|\mathcal{F} \sim \begin{cases} N_n\left(\Sigma_{U_t}\left(\frac{\Sigma_{uc}^{-1}\boldsymbol{\mu}_{uc}}{\sigma_u^2} + \frac{\boldsymbol{\pi}_t - X_i\boldsymbol{\theta}}{\sigma_\epsilon^2}\right), \Sigma_{U_t}\right), & \text{for } t \leqslant t_0, \\ N_n\left(\Sigma_{U_t}\left(\frac{\Sigma_{uc}^{-1}\boldsymbol{\mu}_{uc}}{\sigma_u^2} + \frac{\boldsymbol{\pi}_t - X_t\boldsymbol{\theta}^* - \boldsymbol{V}_t}{\sigma_\epsilon^2}\right), \Sigma_{U_t}\right), & \text{for } t > t_0, \end{cases} \tag{B.6}$$

其中，Σ_{V_t} 和 Σ_{U_t} 定义为

$$\Sigma_{V_t} = \left[\frac{\Sigma_{vc}^{-1}}{\sigma_v^2} + \frac{\mathbf{I}_n}{\sigma_{\epsilon^*}^2}\right]^{-1}, \ \Sigma_{U_t} = \left[\frac{\Sigma_{uc}^{-1}}{\sigma_u^2} + \frac{\mathbf{I}_n}{\sigma_{\epsilon^*}^2}\right]^{-1}.$$

最后，考虑集合 $S_T = \{0, 1, 2, \ldots, T\}$，按照主文稿中描述的步骤，我们得到 t_0 的条件后验为

$$\begin{aligned} f(t_0|\mathcal{F}) &\propto f(\boldsymbol{\pi}^-|\boldsymbol{U}^-, \boldsymbol{\theta}, \boldsymbol{\theta}^*, t_0)f(\boldsymbol{\pi}^+|\boldsymbol{U}^+, \boldsymbol{V}^+, \boldsymbol{\theta}^*, t_0)f(t_0) \\ &\propto I(t_0 \in S_T)(2\pi\sigma_\epsilon^2)^{\frac{-nT_{t_0}^-}{2}}\exp\left\{\frac{-1}{2\sigma_\epsilon^2}\left\|\boldsymbol{\pi}^- - (X^-\boldsymbol{\theta} + \boldsymbol{U}^-)\right\|^2\right\} \\ &\quad (2\pi\sigma_{\epsilon^*}^2)^{\frac{-nT_{t_0}^+}{2}}\exp\left\{\frac{-1}{2\sigma_{\epsilon^*}^2}\left\|\boldsymbol{\pi}^+ - (X^+\boldsymbol{\theta}^* + \boldsymbol{U}^+ + \boldsymbol{V}^+)\right\|^2\right\}. \end{aligned} \tag{B.7}$$