arXiv:2305.02978v1 [stat.ME] 4 May 2023

使用拉普拉斯近似对具有模式化协方差矩阵的层次广义线性混合模型进行边际推断

Jay M. Ver Hoef[1], Eryn Blagg[2], MichaelDumelle[3], Philip M. Dixon[2],Dale L. Zimmerman[4], Paul Conn[1]

[1]海洋哺乳动物实验室 美国国家海洋和大气管理局国家海洋渔业局 阿拉斯加渔业科学中心 7600 Sand Point Way NE, Seattle, WA 98115 电子邮件：jay.verhoef@noaa.gov

[2] 爱荷华州立大学统计学系，爱荷华州艾姆斯[3] 美国环境保护署，俄勒冈州科瓦利斯市西南第35街200号[4] 爱荷华大学统计学与精算科学系，爱荷华州爱荷华市

2023年5月5日

摘要

通过分层构造，我们对广义线性混合模型采用完全参数化方法，在具有复杂协方差依赖关系的情形下，构建了一类广泛且灵活的模型方法。利用拉普拉斯近似，在将所有固定效应和潜在随机效应积分消去的同时，对协方差参数进行边际估计。拉普拉斯近似依赖牛顿-拉夫森更新，而这也能够得到潜在随机效应的预测。我们在分层广义线性混合模型框架下，针对任意模式化协方差矩阵，建立了完整边际推断的方法，涵盖从协方差参数和固定效应的估计，到对未观测数据进行预测。我们针对六种常用于二元、计数和正连续数据的分布构建了边际似然，并且该框架可以轻松扩展到其他分布。我们通过已知参数的随机过程模拟对这些方法进行了说明，并通过模拟实验展示了其在偏差和区间覆盖率方面的有效性。我们还使用选举结果的二元和比例数据、海洋哺乳动物的计数数据，以及环境中重金属浓度的正连续数据作为示例，说明这六种分布在多种模式化协方差结构下的应用。这些结构包括空间模型（例如地统计模型和区域模型）、时间序列模型（例如一阶自回归模型），以及基于分组的典型随机截距混合模型。

1 引言

经典线性模型依赖于在实数轴上具有连续支撑的正态分布，但许多数据是二元数据、计数数据或正连续数据。这类数据可以通过变换来稳定方差，并生成“近似正态”的经验分布，从而能够使用经典线性模型（例如，Snedecor 和 Cochran，1980，第 288 页）。例如，计数数据可以进行平方根变换。然而，Nelder 和 Wedderburn（1972）提出了广义线性模型（GLM，McCullagh 和 Nelder，1989），将其作为线性模型的一种自然扩展，能够适用于泊松分布（如计数数据）、伯努利分布（如二元数据）等；这类模型已广受欢迎，且通常优于数据变换（例如，Warton 和 Hui，2011）。通过线性混合模型引入潜在随机效应，可以将广义线性模型扩展为一类广义线性混合模型（GLMM，Breslow 和 Clayton，1993）。通常假定这些潜在随机效应是独立同分布的正态变量（Zeger 和 Karim，1991），但潜在随机效应也可能存在时间自相关（例如，Stiratelli 等，1984；Zeger 等，1988）、空间自相关（例如，Clayton 和 Kaldor，1987；Gotway 和 Stroup，1997；Diggle 等，1998），或两者兼有（Cressie 和 Wikle，2011，第 380 页）。层级广义线性混合模型（HGLMM，Lee 和 Nelder，1996）为这一领域的研究提供了统一框架。

1.1 分层广义线性混合模型

大多数广义线性模型（GLMs）都是基于指数分布族提出的（例如，Fisher，1934；Lehmann 和 Casella，2006）。然而，大多数常用软件包采用迭代重加权最小二乘法（Wedderburn，1974），通过广义线性模型的前两个矩（均值函数和方差函数）来拟合模型。这一过程允许构建一类灵活的模型，并可通过单一推断程序（准似然法）进行拟合。这一点十分理想，因为某些

准泊松等广义线性模型（GLMs）没有真正的似然函数。

相比之下，我们将通过分层构造采用完全参数化的方法，为广义线性混合模型建立协方差依赖。我们将使用记号 $[\mathbf{y}|\boldsymbol{\mu}]$ 表示随机变量向量 $\mathbf{y}$ 在参数向量或其他固定变量 $\boldsymbol{\mu}$ 条件下的任意联合概率密度函数。在某些情况下，我们可能希望对多个响应向量进行建模（如时间序列应用），或以多组参数为条件，此时表达式可能更类似于 $[\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_k|\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_\ell, \boldsymbol{\phi}]$。作为一个简单示例，考虑由相互独立的负二项分布的乘积构成的联合概率质量函数，该函数可通过均值向量和一个允许过度离散的公共额外参数进行参数化。我们可以将其写为 $[\mathbf{y}|\boldsymbol{\mu}, \phi]$，其中 $\boldsymbol{\mu}$ 表示均值向量，ϕ 是过度离散参数。

对于本文所考虑的广义线性混合模型的分层构造，我们允许均值随其他随机变量 $\mathbf{w}$ 而变化，并通过多元均值函数 $\mathbf{E(y)} = \mathbf{g}^{-1}\mathbf{(w)}$，以这些随机变量 $[\mathbf{y}|\mathbf{g}^{-1}\mathbf{(w)}\ \boldsymbol{\phi}]$ 为条件。在泊松分布中，$\mathbf{g(\cdot)}$ 的每个元素通常都是对数函数，而一般而言，$\mathbf{g(\cdot)}$ 被称为（多元）连接函数（McCullagh 和 Nelder，1989）。连接函数是单调的，因此 $\mathbf{g}^{-1}\mathbf{(\cdot)}$ 的每个元素都与 $g(\cdot)$ 的相应元素一一对应。回顾一下，泊松分布的均值必须为正；如果$\mathbf{g(\cdot)}$ 的每个元素都是对数函数，那么$\mathbf{g}^{-1}\mathbf{(\cdot)}$ 的每个元素就是指数函数，因此 $\boldsymbol{\mu} = \mathbf{g}^{-1}\mathbf{(w)}$ 始终具有正元素，这使得$\mathbf{w}$ 的每个元素都可以不受约束。

我们仅考虑这样的模型：$\mathbf{w}$ 为 $n \times 1$，且服从通过线性混合模型构建的多元正态分布，

$$\mathbf{w} = \mathbf{X}\boldsymbol{\beta} + \sum_{k=1}^{q} \mathbf{Z}_k \mathbf{r}_k + \boldsymbol{\epsilon}, \tag{1}$$

其中，X 是解释变量的 $n \times p$ 满秩设计矩阵， $\boldsymbol{\beta}$ 是固定效应的 $p \times 1$ 参数向量，Z_k 是第 k 个随机效应向量 r_k 的设计矩阵， $\boldsymbol{\epsilon}$ 表示独立同分布的高斯误差。我们假设，对于所有 k，$\mathrm{E}(\mathrm{r}_k)=0$，$\mathrm{E}(\boldsymbol{\epsilon})=0$，$\mathrm{var}(\mathrm{r}_k)=\mathrm{V}_k$，当 $j \neq k$ 时，$\mathrm{cov}(\mathrm{r}_j, \mathrm{r}_k)=0$，且 $\mathrm{var}(\boldsymbol{\epsilon})=\sigma_0^2 \mathrm{I}$。我们用记号 $[\mathrm{w}|\mathrm{X}\ \boldsymbol{\beta}, \{\mathrm{Z}_k\}, \{\mathrm{V}_k\}, \boldsymbol{\theta}]$ 表示概率密度函数 $\mathrm{w} \sim N(\mathrm{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}})$，其中，

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}}=\sum_k \mathbf{Z}_k \mathbf{V}_k \mathbf{Z}_k'+\sigma_0^2 \mathbf{I}.$$

协方差矩阵 $\{\mathbf{V}_k\}$ 对于 $k=1, \ldots, q$ 可以包含除 σ_0^2 之外的其他协方差参数，这些参数都包含在向量 $\boldsymbol{\theta}$ 中。稍后我们将详细介绍 $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ 。

对于完全参数化的层级模型，可以按层级结构构建一个非常一般的模型，如下所示：

$$[\mathbf{y}, \mathbf{w}|\boldsymbol{\phi}, \mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}]=[\mathbf{y}|\mathbf{g}^{-1}(\mathbf{w}), \boldsymbol{\phi}][\mathbf{w}|\mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}], \tag{2}$$

其中，Berliner（1996年）将 $[\mathrm{y}|\mathrm{g}^{-1}(\mathrm{w})\ \boldsymbol{\phi}]$ 称为数据模型，将 $[\mathrm{w}|\mathrm{X}\ \boldsymbol{\beta}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}]$ 称为过程模型。举一个具体例子，假设 $[\mathrm{y}|\exp(\mathrm{w})]$ 服从泊松分布，而 $[\mathrm{w}|\mathrm{X}\ \boldsymbol{\beta}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}]$ 服从多元正态分布，那么联合似然为

$$[\mathbf{y}, \mathbf{w}|\mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}]=\left(\prod_{i=1}^n \frac{\exp(w_i)^{y_i} \exp(-\exp(w_i))}{y_i!}\right) \frac{\exp\left(-0.5(\mathbf{w}-\mathbf{X}\boldsymbol{\beta})' \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}(\mathbf{w}-\mathbf{X}\boldsymbol{\beta})\right)}{(2\pi)^{n/2}|\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|^{1/2}},$$

并注意，使用 $\exp(w_i)$ 表示 $\mathrm{E}(y_i)$。

1.2 模式化协方差矩阵

要为 (2) 构造似然，我们需要为 (1) 中的 $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ 建立参数模型。这里的限制很少，任何有效的 $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ 协方差模型都可以使用。例如， $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$可以基于典型的混合模型构造，其中 $\mathbf{Z}_k$ 包含用于

随机截距的指示变量，或用于随机斜率的解释变量，并且 $\mathbf{V}_k = \sigma_k^2\mathbf{I}$，然后 $\mathbf{\Sigma}_{\boldsymbol{\theta}} = \sum_k \sigma_k^2 \mathbf{Z}_k\mathbf{Z}_k' + \sigma_0^2\mathbf{I}$。我们也可以考虑时间序列模型（例如 Hamilton，1994）。例如，对于一阶自回归（AR1）模型，若 i 和 j 为整数，令 $q = 1$ 且 $\mathbf{Z}_1 = \mathbf{I}$，则 $\mathbf{V}_1$ 的第 i, j 个元素为 $\sigma_1^2\rho^{|i-j|}/(1-\rho^2)$，其中 $0 < \sigma_1^2$ 和$0 \leq \rho < 1$。同样，我们也可以采用地统计模型（例如 Chiles 和 Delfiner，1999），如指数自协方差模型，其中 $q = 1$，$\mathbf{Z}_1 = \mathbf{I}$，且 $\mathbf{V}_1$ 的第 i, j 个元素为 $\sigma_1^2\exp(-\delta_{i,j}/\rho)$，其中 $\delta_{i,j}$ 是第 i个位置与第 j个位置之间的欧氏距离，$0 < \sigma_1^2$，且 $\rho > 0$。其他空间协方差类型包括条件自回归模型（CAR，Besag，1974；Cressie，1993）和同时自回归模型（SAR，Whittle，1954；Ver Hoef 等，2018）、时间序列中的移动平均模型（例如 Hamilton，1994）以及空间统计学中的移动平均模型（Haining，1978）、时空模型（Cressie 和 Wikle，2011，第 380 页），以及定义在非欧氏拓扑空间上的模型，例如球面（如地球；Huang 等，2011；Gneiting，2013）和道路（Ver Hoef，2018）、溪流（Ver Hoef 和 Peterson，2010）等网络。由于协方差矩阵可以作为方差分量通过对协方差矩阵求和来构造，因此，上述所有模型的混合都能产生丰富多样的模式化协方差矩阵，用于建模依赖结构。下文将基于任意有效的协方差矩阵开展推断。

1.3 HGLMM 的推断

数据模型 $[\mathbf{y}|\mathbf{g}^{-1}(\mathbf{w})\ \boldsymbol{\phi}]$（可采用与数据类型相符的任意分布），与过程模型 $[\mathbf{w}|\mathbf{X}\ \boldsymbol{\beta}, \mathbf{\Sigma}_{\boldsymbol{\theta}}]$（可采用任意模式化协方差矩阵）相结合，构成了一种非常丰富且灵活的分层构造（2）。这类模型并非新近提出。

分析方法大致分为两类。最直观的方法是贝叶斯方法，它计算所有潜在变量和参数的后验分布。由于

棘手积分通常通过马尔可夫链蒙特卡洛（MCMC）方法来处理（Gelfand 和 Smith，1990年；Gilks 等，1996年），如今这类方法已发展出许多变体。在我们的研究背景下，贝叶斯层级模型极为流行，最初应用于空间统计学（例如 Clayton 和 Kaldor，1987年）、聚类数据（例如 Zeger 和 Karim，1991）、时间序列（例如 Berliner，1996年）以及纵向数据（Kleinman 和 Ibrahim，1998年）等领域； `WinBUGS` 软件的推出也为其应用提供了助力（Lunn 等，2000）。

第二种方法试图在边缘层面估计协方差参数和固定效应，同时对所有潜在随机效应进行积分消去。也可以将马尔可夫链蒙特卡洛方法用作数值积分器来实现这一点（例如，Zhang，2002；Christensen，2004），但这种方法可能相当缓慢，因此更常用且具有确定性的方法是采用拉普拉斯近似（Tierney 和 Kadane，1986）。具体而言，Rue 等人（2009）提出了积分嵌套拉普拉斯近似（INLA），用于基于高斯马尔可夫随机场进行近似贝叶斯推断。他们利用条件自回归模型的边缘规格来提高计算效率，并使用若干一阶拉普拉斯近似来估计固定效应。Wood（2020）借助稠密协方差矩阵进一步提高了计算效率。软件 `glmmTMB` （Brooks 等人，2017）采用自动微分作为拟合这些模型的一般计算方法。我们的研究主要基于 Evangelou 等人（2011）以及 Bonat 和 Ribeiro Jr（2016）的工作；他们针对二元数据和计数数据，结合地统计协方差结构使用了二阶拉普拉斯近似。更一般的表述见 Bonat 和 Ribeiro Jr（2016），我们将在方法部分指出我们的研究与 Bonat 和 Ribeiro Jr（2016）之间的差异。

尽管这一主题已有相当长的历史，但对于协方差矩阵具有空间或时间相关性结构，或包含其他常用随机效应的情形，目前尚无统一的框架。我们的总体目标是提供完整的边际推断，涵盖协方差参数和固定效应的估计，以及预测。

对于未观测数据，在 HGLMM 框架中适用于任意模式化协方差矩阵。具体而言，我们的目标是：1）求得协方差参数 $\boldsymbol{\theta}$ 和 $\boldsymbol{\phi}$ 的边际最大似然估计与受限最大似然估计，2）预测 $\mathbf{w}$ 的潜在值，3）估计固定效应 $\boldsymbol{\beta}$，以及 4）预测在未采样时间或位置处生成$\mathbf{w}$ 的过程的新值。

本文其余部分安排如下。第 2 节中，我们使用拉普拉斯近似，通过牛顿–拉夫森迭代更新，推导 $\boldsymbol{\theta}$ 和 $\boldsymbol{\phi}$ 的边际最大似然估计，同时得到 $\mathbf{w}$ 的预测。基于 $\mathbf{w}$ 的预测，我们进一步构建 $\boldsymbol{\beta}$ 的估计量，并给出适当的置信区间，以及对生成 $\mathbf{w}$ 的过程中新取值的预测，并提供相应的预测区间。第 3 节中，我们通过模拟展示所有方法，并验证前述推导。第 4 节通过三个独立的实例，利用真实数据集进一步说明这些方法。最后，我们在第 5 节中进行讨论并总结全文。

2 方法

在考虑 HGLMMs 的层次模型表述时，我们希望对分布 $[\mathbf{w},\mathbf{y}|\boldsymbol{\phi},\mathbf{X}\ \boldsymbol{\beta},\boldsymbol{\Sigma}_{\boldsymbol{\theta}}] = [\mathbf{y}|\mathbf{g}^{-1}(\mathbf{w})\ \boldsymbol{\phi}][\mathbf{w}|\mathbf{X}\ \boldsymbol{\beta},\boldsymbol{\Sigma}_{\boldsymbol{\theta}}]$ 进行边际化，得到仅包含数据和方差/协方差参数的分布，并同时消除 $\boldsymbol{\beta}$。拉普拉斯方法有助于实现这一目标。

2.1 HGLMM 的拉普拉斯方法

首先，考虑对 $\boldsymbol{\beta}$ 以及 $\mathbf{w}$ 进行积分，

$$[\mathbf{y}|\boldsymbol{\phi},\boldsymbol{\theta}] = \int_{\mathbf{w}\in\mathbb{R}^n}\int_{\boldsymbol{\beta}\in\mathbb{R}^p}[\mathbf{w},\mathbf{y}|\boldsymbol{\phi},\mathbf{X},\boldsymbol{\beta},\boldsymbol{\Sigma}_{\boldsymbol{\theta}}]d\boldsymbol{\beta}d\mathbf{w} = \int_{\mathbf{w}}[\mathbf{y}|\mathbf{g}^{-1}(\mathbf{w}),\boldsymbol{\phi}]\int_{\boldsymbol{\beta}}[\mathbf{w}|\mathbf{X},\boldsymbol{\beta},\boldsymbol{\Sigma}_{\boldsymbol{\theta}}]d\boldsymbol{\beta}d\mathbf{w}.$$

当 $[\mathbf{w}|\mathbf{X}\ \boldsymbol{\beta}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}]$ 服从高斯分布时，$\int_{\boldsymbol{\beta}}[\mathbf{w}|\mathbf{X}\ \boldsymbol{\beta}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}]d\boldsymbol{\beta}$ 是限制性（也称为残差）最大似然估计（REML，见附录）的似然。注意，REML 最初被推导为一组 $n-p$ 观测值的独立线性组合（称为误差对比）的似然（Patterson and Thompson，1971，1974），而关于通过积分推导 REML 的文献很少。或者，考虑 $[\mathbf{w}|\mathbf{X}\ \boldsymbol{\beta}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}]$其中 $\boldsymbol{\beta}$ 已替换为其条件（给定 $\mathbf{w}$）最大似然（ML）估计量，$\hat{}\boldsymbol{\beta} = (\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\ \mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\ \mathbf{w}$。那么，这两种情况都不含 $\boldsymbol{\beta}$,

$$[\mathbf{w}|\mathbf{X}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}] = \frac{1}{C_n}\exp[(\mathbf{w} - \mathbf{X}\hat{\boldsymbol{\beta}})'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}(\mathbf{w} - \mathbf{X}\hat{\boldsymbol{\beta}})],$$

其中，对于最大似然估计，$C_n = \overline{\sqrt{2\pi^{n/2}|\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|}}$；对于限制性最大似然估计，$C_n = \sqrt{2\pi^{(n-p)/2}|\boldsymbol{\Sigma}_{\boldsymbol{\theta}}||\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\ \mathbf{X}|}$。请注意，Bonat 和 Ribeiro Jr（2016年）仅考虑了对 $\mathbf{w}$ 积分后的边际似然，而未考虑将 $\boldsymbol{\beta}$ 也积分消去（如限制性最大似然估计中所做的那样）或回代代入（如最大似然估计中所做的那样）时的似然。

为了获得数据和协方差参数的边缘分布，我们需要计算以下积分，

$$[\mathbf{y}|\boldsymbol{\phi}, \mathbf{X}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}] = \int_{\mathbf{w}}[\mathbf{y}|\mathbf{g}^{-1}(\mathbf{w}), \boldsymbol{\phi}][\mathbf{w}|\mathbf{X}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}]d\mathbf{w}.$$

令 $\ell(\mathbf{w}\ \cdot) = \text{对数}([\mathbf{y}|\mathbf{g}^{-1}(\mathbf{w})\ \boldsymbol{\phi}][\mathbf{w}|\mathbf{X}\ \boldsymbol{\Sigma}_{\boldsymbol{\theta}}])$，并考虑 $\int e^{\ell(\mathbf{w},\cdot)}d\mathbf{w}$。令 $\mathbf{v}$ 为梯度向量，其第 i 个元素

$$v_i(\boldsymbol{\phi}, \boldsymbol{\theta}) = \frac{\partial\ell(\mathbf{w}, \cdot)}{\partial w_i},$$

并令$\mathbf{H}$ 为海森矩阵，其第 $i,\ j$ 个元素为，

$$H_{i,j}(\boldsymbol{\phi}, \boldsymbol{\theta}) = \frac{\partial^2\ell(\mathbf{w}, \cdot)}{\partial w_i \partial w_j},$$

其中，在 $v_i(\boldsymbol{\phi},\boldsymbol{\theta})$ 和 $H_{i,j}(\boldsymbol{\phi},\boldsymbol{\theta})$ 中，我们均展示了对参数 $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ 的依赖关系。此后，我们假设对数似然具有足够良好的性质，使得 $\mathbf{H}$ 为负定。利用 $\ell(\mathbf{w}\ \cdot)$ 在某一点 $\mathbf{a}$ 附近的多元泰勒级数展开，

$$\int_{\mathbf{w}} e^{\ell(\mathbf{w},\cdot)}d\mathbf{w} \approx \int_{\mathbf{w}} e^{\ell(\mathbf{a},\cdot)+\mathbf{v}'(\mathbf{w}-\mathbf{a})+1/2(\mathbf{w}-\mathbf{a})'\mathbf{H}(\mathbf{w}-\mathbf{a})}d\mathbf{w}.$$

现在，若$\mathbf{a}$ 是$\mathbf{v}\ =\mathbf{0}$的一个取值，则

$$\int_{\mathbf{w}} e^{\ell(\mathbf{w},\cdot)}d\mathbf{w} \approx e^{\ell(\mathbf{a},\cdot)}\int_{\mathbf{w}} e^{-1/2(\mathbf{w}-\mathbf{a})'(-\mathbf{H})(\mathbf{w}-\mathbf{a})}d\mathbf{w} = e^{\ell(\mathbf{a},\cdot)}(2\pi)^{n/2}|-\mathbf{H}_{\mathbf{a}}(\boldsymbol{\phi},\boldsymbol{\theta})|^{-1/2},$$

其中，$\mathbf{H}_{\mathbf{a}}(\boldsymbol{\phi},\ \boldsymbol{\theta})$表示在$\mathbf{a}$ 处取值的$\mathbf{H}$ ，我们再次显示它对 $\boldsymbol{\phi}$ 和$\boldsymbol{\theta}$的依赖关系。最右侧的结果是多元高斯分布归一化常数中的熟知结果。因此，

$$[\mathbf{y}|\boldsymbol{\phi},\mathbf{X},\boldsymbol{\Sigma}_{\boldsymbol{\theta}}] = \int_{\mathbf{w}} e^{\ell(\mathbf{w},\cdot)}d\mathbf{w} \approx [\mathbf{y}|\mathbf{g}^{-1}(\mathbf{a}),\boldsymbol{\phi}][\mathbf{a}|\mathbf{X},\boldsymbol{\Sigma}_{\boldsymbol{\theta}}](2\pi)^{n/2}|-\mathbf{H}_{\mathbf{a}}(\boldsymbol{\phi},\boldsymbol{\theta})|^{-1/2}. \tag{3}$$

2.2 协方差参数的边际最大似然

由式（3）可得，依赖于求得$\mathbf{a}$ 的 $\boldsymbol{\phi}$ 和 $\boldsymbol{\theta}$ 的近似边际最大似然估计量为

$$\{\hat{\boldsymbol{\phi}},\hat{\boldsymbol{\theta}}\} = \underset{\boldsymbol{\phi},\boldsymbol{\theta}}{\arg\max}\left(\log[\mathbf{y}|\mathbf{g}^{-1}(\mathbf{a}),\boldsymbol{\phi}] + \log[\mathbf{a}|\mathbf{X},\boldsymbol{\Sigma}_{\boldsymbol{\theta}}] - \frac{1}{2}\log\left(|-\mathbf{H}_{\mathbf{a}}(\boldsymbol{\phi},\boldsymbol{\theta})|\right)\right), \tag{4}$$

其中我们省略了不包含 $\boldsymbol{\phi}$ 或 $\boldsymbol{\theta}$的项。请注意，log[a|X $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$] 的形式与标准高斯模型中的 ML 或 REML 对数似然相同，但这里是在

在 a处，其中对于 ML

$$\log[\mathbf{a}|\mathbf{X},\boldsymbol{\Sigma_\theta}] = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma_\theta}| - \frac{1}{2}(\mathbf{a}-\mathbf{X}\hat{\boldsymbol{\beta}}_\mathbf{a})'\boldsymbol{\Sigma_\theta}^{-1}(\mathbf{a}-\mathbf{X}\hat{\boldsymbol{\beta}}_\mathbf{a}), \tag{5}$$

以及对于限制性最大
似然

$$\log[\mathbf{a}|\mathbf{X},\boldsymbol{\Sigma_\theta}] = -\frac{n-p}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma_\theta}| - \frac{1}{2}\log|\mathbf{X}'\boldsymbol{\Sigma_\theta}^{-1}\mathbf{X}| - \frac{1}{2}(\mathbf{a}-\mathbf{X}\hat{\boldsymbol{\beta}}_\mathbf{a})'\boldsymbol{\Sigma_\theta}^{-1}(\mathbf{a}-\mathbf{X}\hat{\boldsymbol{\beta}}_\mathbf{a}), \tag{6}$$

其中，在两种情况下，$\hat{}\ \boldsymbol{\beta}_\mathbf{a} = (\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\ \mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\ \mathbf{a}$。结果（4）取决于找到满足$\mathbf{a}$ 的值，使得$\mathbf{v}\ = \mathbf{0}$。为此，我们在给定 $\boldsymbol{\phi}$ 和 $\boldsymbol{\theta}$的条件下使用牛顿-拉夫森法，接下来将对此进行描述。

假设在给定$\mathbf{g}^{-1}(\mathbf{w})$的条件下，$\mathbf{y}$ 的各元素条件独立，则有

$$\log([\mathbf{y}|\mathbf{g}^{-1}(\mathbf{w}),\boldsymbol{\phi}][\mathbf{w}|\mathbf{X},\boldsymbol{\Sigma_\theta}]) = \sum_{i=1}^{n}\log[y_i|g_i^{-1}(w_i),\boldsymbol{\phi}] - \frac{1}{2}(\mathbf{w}-\mathbf{X}\hat{\boldsymbol{\beta}})'\boldsymbol{\Sigma_\theta}^{-1}(\mathbf{w}-\mathbf{X}\hat{\boldsymbol{\beta}}) + C, \tag{7}$$

其中，C 由不含$\mathbf{w}$的项组成。令$\mathbf{d}\phi$ 为其第 i 个分量，

$$d_i \equiv \frac{\partial \log[y_i|g_i^{-1}(w_i),\boldsymbol{\phi}]}{\partial w_i},$$

并注意到

$$\frac{\partial[-\frac{1}{2}(\mathbf{w}-\mathbf{X}\hat{\boldsymbol{\beta}})'\boldsymbol{\Sigma_\theta}^{-1}(\mathbf{w}-\mathbf{X}\hat{\boldsymbol{\beta}})]}{\partial \mathbf{w}} = -\boldsymbol{\Sigma_\theta}^{-1}\mathbf{w} + \boldsymbol{\Sigma_\theta}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}},$$

因此，（7）的梯度为

$$\mathbf{v} = \mathbf{d}_\phi - \boldsymbol{\Sigma_\theta}^{-1}\mathbf{w} + \boldsymbol{\Sigma_\theta}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{d}_\phi - \mathbf{P}_{\boldsymbol{\theta}}\mathbf{w}, \tag{8}$$

其中 $\mathbf{P}_{\boldsymbol{\theta}} = \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} - \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\ \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\ \mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}$ 。对于海森矩阵，令 $\mathbf{D}\phi$ 为一个对角矩阵

其对角线上的第 i 个元素为，

$$D_{i,i} \equiv \frac{\partial^2 \log[y_i|g_i^{-1}(w_i), \boldsymbol{\phi}]}{\partial w_i^2}, \tag{9}$$

由于条件独立性，$i \neq j$ 处的所有二阶偏导数均为 0，因此所有非对角线元素都为零。于是，式（7）的海森矩阵为

$$\mathbf{H} = \mathbf{D}_{\boldsymbol{\phi}} - \mathbf{P}_{\boldsymbol{\theta}}. \tag{10}$$

请注意，式（8）和式（10）与 Bonat 和 Ribeiro Jr（2016）中的梯度和海森矩阵不同，这是因为我们在式（7）中使用了 $\boldsymbol{\beta}$，其中包含$\mathbf{w}$，而他们使用的是 $\boldsymbol{\beta}$。事实上，将式（8）和式（10）中的$\mathbf{P}_{\boldsymbol{\theta}}$替换为 $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}$，即可得到 Bonat 和 Ribeiro Jr（2016）中的梯度和海森矩阵。这是一个重要区别，因为这样我们就可以只针对协方差参数进行优化，而无需同时针对 $\boldsymbol{\beta}$ 进行优化。

表1列出了几种常见分布和链接函数的 d_i and$D_{i,i}$ 。在表1中，我们对负二项分布、伽马分布和贝塔分布采用了另一种参数化方式，使得 E(y) $= \mu$。我们还对逆高斯分布重新进行了参数化，所有分布的详细信息见附录。

表1：HGLMM 的灵活性，展示了如何将不同分布与不同的结构化协方差矩阵相匹配。我们还展示了分布、逆链接函数，以及对数似然各部分关于 w_i 的一阶和二阶偏导数。

对数$[\mathbf{y}\|\mathbf{g}^{-1}(\mathbf{w})\ \boldsymbol{\phi}]$		$-(1/2)$对数$\|-\mathbf{H}_{\mathbf{a}}(\boldsymbol{\phi},\boldsymbol{\theta})\|$		$+$对数$[\mathbf{a}\|\mathbf{X}\ \boldsymbol{\Sigma}_{\boldsymbol{\theta}}]$
分布	$\boldsymbol{\mu}=\mathbf{g}^{-1}(\mathbf{w})$	d_i	$D_{i,i}$	$\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$类型
二项分布	$\boldsymbol{\mu}=\frac{\exp(\mathbf{w})}{1+\exp(\mathbf{w})}$	$y_i-\frac{n_i\exp(w_i)}{1+\exp(w_i)}$	$\frac{n_i\exp(w_i)}{(1+\exp(w_i))^2}$	随机效应
泊松	$\boldsymbol{\mu}=\exp(\mathbf{w})$	$y_i-\exp(w_i)$	$-\exp(w_i)$	地统计
负二项分布	$\boldsymbol{\mu}=\exp(\mathbf{w})$	$\frac{\phi(y_i-e^{w_i})}{\phi+e^{w_i}}$	$\frac{\phi e^{w_i}(\phi+y_i)}{(\phi+e^{w_i})^2}$	空间区域
伽马分布	$\boldsymbol{\mu}=\exp(\mathbf{w})$	$-\phi+y_i\phi e^{-w_i}$	$-y_i\phi e^{-w_i}$	时间序列
逆高斯分布	$\boldsymbol{\mu}=\exp(\mathbf{w})$	$\phi\left(\frac{y}{2e^{w_i}}-\frac{e^{w_i}}{2y}\right)+\frac{1}{2}$	$\frac{\phi(e^{2w_i}+y_i^2)}{2ye^{w_i}}$	时空
Beta	$\boldsymbol{\mu}=\frac{\exp(\mathbf{w})}{1+\exp(\mathbf{w})}$	$\frac{-\phi e^{w_i}k_0(w_i\|\phi,\ y_i)}{(e^{w_i}+1)^2}$	$\frac{-\phi e^{2w_i}k_1(w_i\|\phi,\ y_i)}{(e^{w_i}+1)^4}$	

$k_0(w_i|\phi,y_i)=\psi^{(0)}\left(\frac{\phi e^{w_i}}{1+e^{w_i}}\right)-\psi^{(0)}\left(\frac{\phi}{1+e^{w_i}}\right)+$对数$\left(\frac{1}{y_i}-1\right)k_1(w_i|\phi,y_i)$

$=\phi\left(\psi^{(1)}\left(\frac{\phi e^{w_i}}{1+e^{w_i}}\right)+\psi^{(1)}\left(\frac{\phi}{1+e^{w_i}}\right)\right)-2$双曲正弦$(w_i)\left(k_0(w_i|\phi,y_i)\ +2$双曲正切$^{-1}(1-2y_i)\right)$

$\psi^{(n)}(\cdot)$ 是双伽马函数的第 n 阶导数。双曲正弦和双曲正切分别是双曲正弦函数和双曲正切函数。

在给定 $\boldsymbol{\phi}$ 和 $\boldsymbol{\theta}$的条件下，牛顿-拉夫森更新为：

$$\mathbf{w}^{[k+1]}=\mathbf{w}^{[k]}-\mathbf{H}^{-1}\mathbf{v},$$

并且在收敛时，我们将$\mathbf{a}\ =\mathbf{w}$ 代入 (4)，以便在给定$\boldsymbol{\phi}$ 和 $\boldsymbol{\theta}$时计算似然。请注意，这使得边际最大似然需要进行双重迭代，因为我们在优化 $\boldsymbol{\phi}$ 和 $\boldsymbol{\theta}$的同时求解$\mathbf{a}$ 。也可以使用其他最大化方法，例如期望最大化算法，但通常牛顿-拉夫森算法收敛很快（根据我们的经验，通常约需 10 次迭代），Bonat 和 Ribeiro Jr（2016）也倾向于使用该算法。不过，有时需要调整步长，以免$\mathbf{v}$ 发散。例如，检查$\mathbf{v}^{[k+1]}=\mathbf{d}\phi\ -\mathbf{P}_{\boldsymbol{\theta}}\mathbf{w}^{[k+1]}$很简单且快速；如果 $\mathbf{v}^{[k+1]}$ 是

按照某种准则（例如 $\mathbf{v}$ 的最大元素或平均元素），比 $\mathbf{v}$“更大”，然后取

$$\mathbf{w}^{[k+1]} = \mathbf{w}^{[k]} - \alpha \mathbf{H}^{-1}\mathbf{v},$$

其中 $0 < \alpha < 1$。在下面的模拟中，我们按上述方法检查 $\mathbf{v}^{[k+1]}$，并在 $\mathbf{v}^{[k+1]}$ 的最大元素大于 $\mathbf{v}$ 的最大元素时，设定 $\alpha = 0.1$。使用牛顿-拉夫森法的优点是它能提供 $\mathbf{H}$，这有助于在估计固定效应和进行预测时调整方差，下一节将对此进行说明。

总之，协方差参数和 $\mathbf{w}$ 的估计可按以下步骤进行：

1. 为协方差参数 $\boldsymbol{\phi}$ 和 $\boldsymbol{\theta}$ 取初始值。例如，对于方差分量（如 σ_0^2 和 σ_1^2），将 var(g(y)) 平均分配给各个方差分量。如果解释变量较多，可以对 g(y)) 拟合线性模型，并使用残差方差。

2. 为 $\mathbf{w}$ 选取初始值。例如，设 $\mathbf{w} = g$(y)，或使用对 g(y) 拟合线性模型所得的残差。

3. 使用牛顿-拉夫森法，根据（4）中的给定 $\boldsymbol{\phi}$ 和 $\boldsymbol{\theta}$ 估计 $\mathbf{w}$ =**a**。

4. 对 $\mathbf{w}$、$\boldsymbol{\phi}$ 和 $\boldsymbol{\theta}$，计算（4）中的对数似然。

5. 对不同的 $\boldsymbol{\phi}$ 和 $\boldsymbol{\theta}$ 重复步骤 3 和 4，同时在步骤 4 中优化对数似然，直到收敛。

2.3 固定效应的推断

为了估计 ϕ 和 θ ，每次计算似然时，都需要使用牛顿-拉夫森法优化$\mathbf{w}$的似然，我们将其称为$\mathbf{a}$。在估计 ϕ 和 θ时达到收敛后，我们也就优化了$\mathbf{w}$。我们将最优值记为 $\hat{\mathbf{w}} = \mathbf{a}$。

Bonat 和 Ribeiro Jr（2016）提出了用于估计 β 并获得置信区间的剖面似然方法，但该方法计算量很大，且不适用于 β中包含许多系数的情况。另一种 β 估计量的构造方法是：在假设$\mathbf{w}$ was 已观测到的情况下，将广义最小二乘估计量中的未观测$\mathbf{w}$ 替换为其预测值$\hat{\mathbf{w}} = \mathbf{a}$，该预测值按上一小节所述方法得到。由此得到估计量 $\beta = \mathbf{B}\hat{\mathbf{w}}$，其中 $\mathbf{B} = (\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}$ 。为了估计$\hat{}$ β的方差，方便起见，可将$\mathbf{w}$ 视为已观测并以此为条件，然后使用以下众所周知的结果，通常称为全方差公式，

$$\operatorname{var}(\hat{\mathbf{w}}) = \mathrm{E}_{\mathbf{w}}[\operatorname{var}(\hat{\mathbf{w}}|\mathbf{w})] + \operatorname{var}_{\mathbf{w}}[\mathrm{E}(\hat{\mathbf{w}}|\mathbf{w})]. \tag{11}$$

由于通过似然对 $\hat{}$w 进行优化，我们假设 $\hat{}$w|w 近似服从 N(w, $\mathrm{F}_{\mathbf{w}}^{-1}$)，其中 $\mathrm{F}_{\mathbf{w}}$ 是观测费希尔信息；或者较宽松地假设 E($\hat{}$w|w) = w 且 var($\hat{}$w|w) = $\mathrm{F}_{\mathbf{w}}^{-1}$近似成立。因此，对于 (11) 中的第二项，$\operatorname{var}_{\mathbf{w}}[\mathrm{E}(\hat{}\mathrm{w}|\mathrm{w})] = \boldsymbol{\Sigma}_{\boldsymbol{\theta}}$，我们用 $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}$ 近似该项，其中以估计的参数 $\hat{}$ ters θ 替代 θ。对于 (11) 中的第一项，观测费希尔信息等价于$-\mathrm{H}_{\mathrm{w}}(\phi, \theta)^{-1}$，这里我们展示了它对参数 ϕ 和 θ的依赖关系，以及对来自 (10) 中 Dϕ 的 w 的依赖关系（参见表1中的 D_i、$_i$ 示例）。要得到费舍尔信息，需要取期望，$\mathrm{E}_{\mathbf{w}}[\operatorname{var}(\hat{}\mathrm{w}|\mathrm{w})]$, ，但这很复杂；Efron 和 Hinkley (1978) 主张使用观测费希尔信息，因此我们只需将 $-\mathrm{H}_{\mathrm{w}}(\phi, \theta)^{-1}$ 中的 w 替换为 $\hat{}$w = a，并将 ϕ 和 θ 替换为它们的估计值 $\hat{}$ ϕ 和 $\hat{}\theta$,

并将其记为 $-\mathbf{H}_{\hat{\mathbf{w}}}(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})^{-1}$。那么，估计的固定效应的协方差矩阵的一个估计量为

$$\widehat{\mathrm{var}}(\hat{\boldsymbol{\beta}}) = \mathbf{B}[\mathrm{var}(\hat{\mathbf{w}})]\mathbf{B}' = \mathbf{B}[-\mathbf{H}_{\hat{\mathbf{w}}}(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}})^{-1}]\mathbf{B}' + \mathbf{C}_{\hat{\boldsymbol{\beta}}}, \tag{12}$$

其中$\mathbf{C}_{\hat{\boldsymbol{\beta}}} = \mathbf{B}\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}\mathbf{B}'$，可简化为$\mathbf{C}_{\hat{\boldsymbol{\beta}}} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}_{\hat{\boldsymbol{\theta}}}\mathbf{X})^{-1}$，这是在$\mathbf{w}$ 可观测时使用广义最小二乘所得的固定效应通常估计协方差矩阵。

2.4 预测推断

到目前为止，我们已经估计了 $\boldsymbol{\theta}$、 $\boldsymbol{\phi}$和 $\boldsymbol{\beta}$，预测了$\mathbf{w}$，并获得了 $\boldsymbol{\beta}$ 和$\mathbf{w}$的估计协方差矩阵$\hat{}$。现在来考虑对未采样数据进行预测的任务，这些数据可能位于空间或时间上，也可能由设计确定。我们用向量$\mathbf{u}$表示未采样的 $\{w_i\}$ 。我们可以将线性模型(1)扩展为

$$\begin{pmatrix} \mathbf{w} \\ \mathbf{u} \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{X}\boldsymbol{\beta} \\ \mathbf{X}_{\mathbf{u}}\boldsymbol{\beta} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\theta}} & \boldsymbol{\Sigma}_{\mathbf{wu}} \\ \boldsymbol{\Sigma}'_{\mathbf{wu}} & \boldsymbol{\Sigma}_{\mathbf{uu}} \end{pmatrix}\right). \tag{13}$$

我们的目标是预测$\mathbf{u}$。如果$\mathbf{w}$ 可观测，那么$\mathbf{u}$ 的最佳线性无偏预测量（BLUP）为 $\boldsymbol{\Lambda}\mathbf{w}$，其中 $\boldsymbol{\Lambda} = \mathbf{X}_{\mathbf{u}}\mathbf{B} + \boldsymbol{\Sigma}'_{\mathbf{wu}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} - \boldsymbol{\Sigma}'_{\mathbf{wu}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{XB}$ （Goldberger，1962年）。然而，由于$\mathbf{w}$ 不可观测，可以在该表达式中用$\hat{}\mathbf{w}$替代$\mathbf{w}$ ，从而得到$\mathbf{u}$ 的替代预测量$\hat{} = \boldsymbol{\Lambda}\hat{}\mathbf{w}$。

为了确定$\hat{}\mathbf{u}$的抽样性质，我们还需要再次进行一些调整，以说明在最佳线性无偏预测中用$\hat{}\mathbf{w}$ 替代$\mathbf{w}$ 的情况（目前，在创建协方差矩阵时，我们将协方差参数视为已知条件）。同样假设$\hat{}\mathbf{w}$ 是$\mathbf{w}$的无偏估计量，可以很容易地看出，这个替代预测量是无偏的，即 $\mathrm{E}(\boldsymbol{\Lambda}\hat{}\mathbf{w}) = \mathrm{E}(\mathbf{u})$。现在，我们希望得到与该预测量相关的预测误差方差的估计量，即 $\mathrm{var}(\hat{}\mathbf{u} - \mathbf{u}) = \mathrm{var}(\boldsymbol{\Lambda}\hat{}\mathbf{w} - \mathbf{u})$。注意，预测误差

方差的最佳线性无偏预测为

$$\operatorname{var}(\boldsymbol{\Lambda}\mathbf{w}-\mathbf{u})=\boldsymbol{\Sigma}_{\mathbf{uu}}-\boldsymbol{\Sigma}'_{\mathbf{wu}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\boldsymbol{\Sigma}_{\mathbf{wu}}+\mathbf{K}\mathbf{C}_{\boldsymbol{\beta}}\mathbf{K}', \tag{14}$$

其中，$\mathbf{K}=\mathbf{X}_{\mathbf{u}}-\boldsymbol{\Sigma}'_{\mathbf{wu}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{X}$（Goldberger，1962年）。为了获得我们所提出的替代预测量的预测误差方差，像式（11）中那样，以$\mathbf{w}$和$\mathbf{u}$为条件进行计算会更加方便，即，

$$\operatorname{var}(\boldsymbol{\Lambda}\hat{\mathbf{w}}-\mathbf{u})=\mathrm{E}_{\mathbf{w},\mathbf{u}}[\operatorname{var}(\boldsymbol{\Lambda}\hat{\mathbf{w}}-\mathbf{u}|\mathbf{w},\mathbf{u})]+\operatorname{var}_{\mathbf{w},\mathbf{u}}[\mathrm{E}(\boldsymbol{\Lambda}\hat{\mathbf{w}}-\mathbf{u}|\mathbf{w},\mathbf{u})].$$

由于假设 ˆw 对 w 具有无偏性，我们有 E(Λ ˆw −u|w, u) = Λw −u，其方差由式 (14) 给出。在条件意义下，var$_{\hat{\mathbf{w}}}$(Λ ˆw −u) 不依赖于 u，因此 E$_{\mathbf{w},\mathbf{u}}$[var(Λ ˆw −u|w, u)] = E$_{\mathbf{w}}$(Λ[−H$_{\mathbf{w}}$($\boldsymbol{\phi}$, $\boldsymbol{\theta}$)$^{-1}$]Λ′)。与上一节中的做法一样，我们不再求期望，而是通过在 H 中用其估计量 ˆw =a 替换 w ˆ ˆ，并用 $\boldsymbol{\phi}$ 和 $\boldsymbol{\theta}$ 替换 $\boldsymbol{\phi}$ 和 $\boldsymbol{\theta}$，直接使用观测费希尔信息。将这些结果结合起来，我们得到

$$\widehat{\operatorname{var}}(\boldsymbol{\Lambda}\hat{\mathbf{w}}-\mathbf{u})=\boldsymbol{\Lambda}[-\mathbf{H}_{\hat{\mathbf{w}}}(\hat{\boldsymbol{\phi}},\hat{\boldsymbol{\theta}})^{-1}]\boldsymbol{\Lambda}'+\boldsymbol{\Sigma}_{\mathbf{uu}}-\boldsymbol{\Sigma}'_{\mathbf{wu}}\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}^{-1}\boldsymbol{\Sigma}_{\mathbf{wu}}+\mathbf{K}\mathbf{C}_{\hat{\boldsymbol{\beta}}}\mathbf{K}'. \tag{15}$$

所有协方差矩阵都取决于 $\boldsymbol{\theta}$，尽管只有$\mathbf{w}$的ˆ协方差矩阵在记号中明确体现了这一点。我们将这些矩阵中的 $\boldsymbol{\theta}$ 替换为其估计值 $\hat{\boldsymbol{\theta}}$，其中，用于估计 $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ 的拟合协方差函数也用于估计 $\boldsymbol{\Sigma}_{\mathbf{wu}}$ 和 $\boldsymbol{\Sigma}_{\mathbf{uu}}$。

3 模拟

我们首先通过一个模拟的空间数据集来展示我们的方法，这样我们就知道所有参数和$\mathbf{w}$的真实值。我们在一个(0，1)×(0，1)（单位正方形）区域上创建了一个包含 20×20 个数据位置的正方形网格。在这个网格上，我们从一个多元

正态分布中生成了一个$\mathbf{w}$，其均值向量为 $\mathbf{1}\beta_0$，其中 $\beta_0 = 2$；协方差矩阵则通过计算指数自协方差模型确定：$\mathrm{cov}(w(\mathbf{s}_i)\ w(\mathbf{s}_j)) = \sigma_1^2 \exp(-\delta_{i,\ j}/\rho) + \sigma_0^2 \mathcal{I}(\delta_{i,\ j} = 0)$，其中$\mathbf{s}_i (i = 1, \ldots,\ 400)$ 是位置 i 处的空间坐标向量，$\delta_{i,\ j}$ 表示第 i 个位置与第 j 个位置之间的欧氏距离，$\mathcal{I}(\cdot)$是指示函数，当其参数为真时取1，否则取0。我们设定 $\sigma_1^2 = 1$、$\rho = 1$和 $\sigma_0^2 = 0.0001$。$\mathbf{w}$ 中的400个模拟值如图1B所示。给定$\mathbf{w}$，我们在每个数据位置独立模拟一个均值为$\exp(w_i)$的泊松随机变量，结果如图1A所示。

使用图1A中的模拟数据，并假设均值和指数协方差函数的参数未知，我们使用限制性最大似然对 $\boldsymbol{\theta} = (\sigma_1^2,\ \rho,\ \sigma_0^2)$ 的式（4）中的似然进行了优化。由此得到的参数估计值为 $\hat{}\sigma_1^2 = 1.247$、$\hat{}\rho = 1.341$ 和 $\hat{}\sigma_0^2 = 1.392 \times 10^{-11}$。$\sigma_1^2$和 ρ的似然曲面如图1C所示。明显的脊线表明，似然中 σ_1^2 与 ρ呈正相关，这在地统计模型中很常见。对 $\boldsymbol{\theta} = (\sigma_1^2,\ \rho,\ \sigma_0^2)$ 的估计还得到了 $\hat{}\mathbf{w}$，如图1D所示；结果表明，我们很好地恢复了真实模拟$\mathbf{w}$ 中的空间模式。总体截距 β_0 的估计值为 0.852，估计标准误为 0.865。

图1见此处~~：模拟泊松计数数据的~~估计。A. 使用文中所述模型模拟的计数数据。B. 模拟数据的真实$\mathbf{w}$ 值。C. 模拟数据的边际似然曲面。白色圆圈表示估计值。D. 预测的 $\hat{}\mathbf{w}$值。

当然，这只是一次模拟。为了评估估计量的偏差和置信区间的覆盖率，我们进行了一项规模更大的模拟实验。我们在单位正方形内随机选取的 200 个数据位置上模拟数据。我们从多元正态分布中进行模拟，其自协方差模型与所使用的模型完全相同

之前的做法，但采用均值结构

$$\mathrm{E}(w_i) = \beta_0 + \beta_1 x_i + \beta_2 \tau_i + \beta_3 x_i \tau_i$$

其中， x_i 是从 N(0，1) 中随机独立模拟得到的， τ_i 是以概率 $p = 0.5$ 为参数随机独立模拟得到的 Bernoulli 变量。我们设定 $\boldsymbol{\beta}$ =(0.5，0.5，-0.5，0.5)′。我们还在单位正方形内的 10×10 正方形网格上创建了 100 个预测位置，并在这些预测位置模拟了解释变量。总共从式（13）所示的模型中模拟了 $300\ w_i$ 个值（200 个观测值，100 个用于预测）。随后，我们以在w条件下来自泊松分布的计数作为观测数据；在 200 个数据位置上，我们分别独立模拟一个泊松随机变量，其均值等于 exp(w_i)。通过这种方式，我们模拟了 2000 个大小为 300 的数据集，以评估偏差以及置信区间/预测区间覆盖率。

对于每个模拟数据集，我们首先使用式（4）估计协方差参数，其中 log[a|**X** $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$] 是式（6）给出的限制性最大似然对数似然，拟合时我们仍使用指数自协方差模型。我们分别将 σ_1^2 和ρ的参数空间上界设为 var($g(y)$) 的 10 倍，以及所有空间位置之间的最大距离的 10 倍。这样做是因为有时 σ_1^2 或 ρ的估计值（也可能两者）似乎会无限增大，这对应于似然中一条非常平坦的脊线，与图1C中的情况相似；而估计和预测只取决于它们的比率（Zhang，2004年）。这使得如此多次模拟中的估计过程更加稳定。

我们将估计得到的协方差参数作为自协方差模型的代入值，以获得 $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}$，并结合估计得到的w，通过 $\boldsymbol{\beta} = \mathbf{B}\hat{}\mathbf{w}$ 估计固定效应。为估计偏差，我们对所有 2000 个模拟数据集中的 $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}$ （逐元素）取平均值；为估计均方误差（MSE），我们对所有 2000 个模拟数据集中的 $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^2$ （逐元素）取平均值。此外，我们还构建了名义置信水平为 90% 的置信区间。

ˆ ˆ ˆ 如 $\beta_j \pm 1.645$ ^se(β_j)，其中 ^se(β_j) 是（$j+1$）在（12）（$j = 0,1,2,3$）的对角线元素的平方根。我们计算了在 2000 次模拟中置信区间包含真实值的比例。如果我们能够很好地估计 β_j及其方差，经验覆盖率应接近 90%。我们还计算了基于使用$\mathbf{C}\hat{\beta}$的对角线元素的置信区间覆盖率，即未经调整的广义最小二乘方差估计量。

结果如表2所示，我们可以看到，在估计 $\boldsymbol{\beta}$ 中的任何参数时，偏差都非常小。鉴于均方误差是偏差平方与方差之和，很明显，在这些模拟中，偏差只是均方误差中非常小的一部分。β_0 的置信区间覆盖率较低，但对于正态空间模型而言，估计总体截距同样很困难。使用（12）时，β_1、β_2 和 β_3 的置信区间覆盖率都非常接近90%；但仅使用 $\mathbf{C}\hat{\beta}$ 时，它们则显著低于90%。

我们还使用 $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}$ 中估计得到的协方差参数以及预测得到的 ˆ$\mathbf{w}$，利用 ˆ$\mathbf{u} = \boldsymbol{\Lambda}$ˆ$\mathbf{w}$，针对每个模拟数据集的全部 100 个预测位置进行预测。为了估计预测偏差，我们计算了每个模拟数据集中 ˆ$\mathbf{u} - \mathbf{u}$ 各元素的平均值，其中 $\mathbf{u}$ 包含这 100 个预测位置处的模拟值，然后再对 2000 个模拟数据集的结果取平均。我们还构造了 90% 预测区间，形式为 ˆ$u_k \pm 1.645$ ^se(ˆu_k) ($k = 1, \ldots, 100$)，其中 ^se(ˆu_k) 是式（15）中第 k 个对角线元素的平方根。我们计算了在 100 次预测和 2000 次模拟中，预测区间包含真实值的比例，该比例理论上应接近 90%。表2给出了预测结果，结果表明几乎没有偏差的迹象。使用式（15）时，覆盖率非常接近 90%；而使用朴素方法（14）时，覆盖率则过低。

表2：固定效应估计的偏差和覆盖率 $\boldsymbol{\beta}$ ，以及未观测位置处$\mathbf{u}$ 的预测。MSE为均方（预测）误差，ratio为偏差^2/MSE。覆盖率针对90%置信区间和预测区间；CI90_c 使用了（12）和（15）中的校正版本，而CI90_u 显示的是采用基于C$\boldsymbol{\beta}$ 的未校正标准误估计量，以及根据（14）计算的未校正预测标准误时的覆盖率。粗体数值超出了真实值为0.9的2000次独立伯努利试验所得的99%置信区间。

效应	bias	MSE	比率	CI90_u	CI90_c
$\hat{\beta}_0$	0.042	0.530	0.003	**0.750**	**0.773**
$\hat{\beta}_1$	-0.005	0.008	0.003	**0.367**	0.899
$\hat{\beta}_2$	0.004	0.024	0.001	**0.310**	0.910
$\hat{\beta}_3$	-0.005	0.018	0.001	**0.357**	0.911
$\hat{\mathbf{u}}$	0.039	0.138	0.011	**0.700**	0.898

除了泊松分布外，我们还对表1中的其他五种分布进行了类似的模拟。二项分布（Bernoulli）不需要 No ϕ，但对于贝塔分布，我们设定了 $\phi = 10$；对于负二项分布、伽马分布和逆高斯分布，我们设定了 $\phi = 1$。我们的替代估计量和预测量在所有五种情况下似乎都没有偏差，因此我们只在表3中展示校正后的90%置信区间覆盖率。除一种情况外，这些区间的置信区间覆盖率和预测区间覆盖率都接近90%；例外是贝塔分布，其覆盖率略低，尤其是在预测方面。还需要进行更多模拟，才能全面确定这些分布对应的区间何时会存在不足。

表3：表1中未纳入表2的五种分布下固定效应 $\boldsymbol{\beta}$ 的估计区间覆盖率，以及对未观测位置的 $\mathbf{u}$ 进行预测的区间覆盖率；这五种分布分别为二项分布（bino）、贝塔分布、负二项分布（nbin）、伽马分布（gamm）和逆高斯分布（iGau）。覆盖率针对90%置信区间和预测区间，使用（12）和（15）中的校正版本。粗体数值落在真实值为0.9的2000次独立伯努利试验的99%置信区间之外。

效应	bino	beta	nbin	gamm	iGua
$\hat{\beta}_0$	**0.712**	**0.923**	**0.730**	**0.839**	**0.708**
$\hat{\beta}_1$	0.901	**0.828**	0.885	0.898	0.908
$\hat{\beta}_2$	0.914	**0.837**	0.898	0.894	0.888
$\hat{\beta}_3$	0.915	**0.816**	0.901	0.892	0.900
$\hat{\mathbf{u}}$	0.887	**0.744**	0.883	0.892	**0.875**

4 示例

我们通过三个示例数据集展示这些方法。这些数据集涵盖表1中的所有分布，并结合了通过空间自回归模型、时间序列模型、地统计模型以及包含随机效应的方差分量模型构建的协方差矩阵。

4.1 得克萨斯州1980年总统选举投票率

该数据集包含美国1980年总统选举中，19岁以上人口的投票比例，统计范围为得克萨斯州的254个县。美国全国范围的数据由Pace和Barry（1997年）收集并报告，现可在 R 程序包 `spData` （Bivand等，2023年）中获取。我们仅选取了得克萨斯州的数据子集。响应变量以比例形式报告，但我们

还创建了一个二元变量：比例大于0.5时赋值为1，否则赋值为0。这里，我们将对二元响应变量拟合二项分布（由于所有样本量均为1，实际上是伯努利分布），并对比例响应变量拟合贝塔分布。

该数据集中包含三个解释变量：1）拥有大学学历的人口比例；2）住房拥有率；3）人均收入。其中，对于这三个变量，其取值均以符合投票资格的19岁以上总人口为基准。图2展示了三个解释变量分别与比例响应变量的 logit 变换值之间的散点图。为了实现关系的线性化，我们对住房拥有率这一解释变量进行了三次方变换，并对人均收入取自然对数。因此，我们所考虑的线性模型为，

$$\mathbf{w} = \mathbf{X}\boldsymbol{\beta} + \mathbf{r}_1,$$

其中 $\mathbf{X}$ 包含一列全为 1 的列，用于表示总体均值，另外三列用于表示（变换后的）解释变量。

此处为图2：~~投票率响应变量的~~ logit变换与三个解释变量之间的散点图。请注意，部分解释变量进行了变换，其中住房拥有率的比例进行了立方处理，人均收入则取了自然对数。

对于空间随机效应 $\mathbf{r}_1$，，我们针对 254 个县的数据拟合了两种空间自回归模型：条件自回归（CAR）模型和同时自回归（SAR）模型。这些模型依赖于邻居定义，而非直接依赖距离。我们将某个县的邻居定义为其质心与该县质心之间的距离在 150 km 以内的任何其他县。根据这一定义，一些县只有一个邻居，而另一些县则有多个邻居；最多的有 38 个。令 $\mathbf{W} = (W_{i,j})$ 表示邻居关联矩阵，其中 $W_{i,j} = 1$ 如果县 j 是县 i 的邻居，则 $W_{i,j} = 0$

否则，对角线上的元素均为零（即一个位置不是其自身的邻居）。令 $\mathbf{W}_{rs}$ 为 $\mathbf{W}$ 的“行标准化”版本，其中 $\mathbf{W}$ 的任意一行中的元素均除以该行的元素之和，$W_{i,+} = \sum_j W_{i,j}$。那么，CAR模型的协方差矩阵为

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = \sigma^2(\mathbf{I} - \rho\mathbf{W}_{rs})^{-1}\mathbf{M}_{rs},$$

其中，$\mathbf{M}_{rs}$ 是一个对角矩阵，其第 i个对角元素为$1/W_{i,+}$。SAR协方差矩阵为

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = \sigma^2[(\mathbf{I} - \rho\mathbf{W}_{rs})(\mathbf{I} - \rho\mathbf{W}'_{rs})]^{-1}.$$

在这两个协方差矩阵中，ρ 是一个空间依赖参数。

对于采用二元响应变量、三个解释变量和 CAR协方差矩阵的模型，使用 REML 和公式 (4)，我们估计得到 $\sigma^2 = 0.625$ 和 $\rho = 0.999$。将模型的协方差矩阵改为 SAR协方差矩阵后，我们估计得到 σ^2 =0.265 和 $\rho = 0.964$。公式 (4) 中负二倍对数似然的最小值在 CAR模型下为 298.14，而在 SAR模型下为 279.15，这表明 SAR模型是更好的选择。表4 给出了固定效应估计值。请注意，基于$\mathbf{C}\hat{\beta}$ 的朴素方法与公式 (12) 给出的校正版本，其标准误差存在很大差异。

表4：使用二元响应变量对得克萨斯州投票率数据进行分析所得的固定效应估计值表。估计值列记为 Est.；s.e.$_u$ 是仅使用第 2.3 节中的C_β 计算的朴素标准误；s.e.$_c$ 是使用公式 (12) 计算的校正标准误。t值为估计值除以校正标准误；p值是在效应确实为零的情况下，超出该 t 值的概率。

	SAR模型					CAR模型	
效应	Est.	标准误$_u$	标准误$_c$	t值	p值	Est.	标准误$_c$
截距	-5.725	0.502	2.397	2.388	0.0177	-2.898	1.975
大学	4.439	0.701	3.694	1.202	0.2307	4.137	2.979
房主	69.768	2.221	12.771	5.463	<0.0001	56.635	10.318
收入	-0.171	0.260	1.414	0.121	0.9040	-0.860	1.151

对于比例投票率响应变量，使用了表1中的贝塔分布；对于CAR协方差矩阵，我们估计了 $\sigma^2 = 0.312$、 $\rho = 0.999$和 $\phi = 48.7$；对于SAR协方差矩阵，我们估计了 $\sigma^2 = 0.0126$、 $\rho = 0.941$和 $\phi = 46.9$。(4)式中负两倍对数似然的最小值，CAR模型为-554.6，SAR模型为-557.8，这表明SAR模型是更好的选择，与二元数据的情况相同。表5给出了固定效应估计值。与二元数据的情况一样，基于C_β 的朴素方法与(12)式给出的校正版本，在标准误方面存在较大差异。表4和表5中，SAR模型与CAR模型的系数估计值及其精度总体模式相似。比较表4和表5可见，表5中的估计值似乎更精确，尤其是人均收入变量的显著性方面。这并不意外，因为将比例投票率数据转换为二元数据必然会导致信息损失。

表5：使用比例响应变量的德克萨斯州投票率数据估计的固定效应表。各标题与表4相同。

	SAR模型					CAR模型	
效应	Est.	s.e.$_u$	标准误$_c$	t值	p值	Est.	标准误$_c$
截距	-1.614	0.105	0.253	6.379	<0.0001	-1.427	0.340
大学	0.407	0.154	0.407	1.000	0.3184	0.481	0.394
房主	8.711	0.486	1.300	6.703	<0.0001	8.552	1.273
收入	0.470	0.057	0.142	3.317	0.0010	0.390	0.143

预测的$\hat{\mathbf{w}}$ 值见图3。二元数据的空间可视化显示，1和0似乎存在一定的聚集现象（图3A）。使用SAR模型得到的二元数据预测$\hat{\mathbf{w}}$ 值，在得克萨斯州北部“锅柄区”和得克萨斯州中部最高（图3B）；使用CAR模型得到的二元数据预测$\hat{\mathbf{w}}$值也呈现类似模式（图3C）。原始比例投票率数据经过逻辑变换后的结果见图3D，而SAR模型的预测$\hat{\mathbf{w}}$ 值（图3E）似乎对原始数据进行了平滑处理，其空间模式与二元数据（图3B）以及使用CAR模型得到的预测$\hat{\mathbf{w}}$ 值（图3F）相似。请注意，对于空间模型，$\mathbf{w}$ 与总体均值 β_0存在一定的交互作用。因此，图3中的色标没有采用相同的断点，因为各子图之间的相对模式最为重要。

图3见此处~~：得克萨斯州投票率数据~~的原始数据和预测的空间随机效应（$\mathbf{w}$）。A）原始二元数据，空心圆表示0，实心圆表示1；B）使用SAR模型对二元数据预测的$\hat{\mathbf{w}}$ ； C）使用CAR模型对二元数据预测的$\hat{\mathbf{w}}$ ； D）经logit变换的比例投票率数据；E）使用SAR模型对比例投票率数据预测的$\hat{\mathbf{w}}$ ； F）使用CAR模型对比例投票率数据预测的$\hat{\mathbf{w}}$ 。

对于贝　塔分布，使用了用于比例　　l

投票率数据中，ϕ 估计为 46.9。$[y|\mu,\phi]$ 的概率密度取决于 μ，而 μ受解释变量影响。图 4 展示了原始数据的直方图，以及拟合的概率密度函数 $[y|\mu,46.9]$；该函数服从贝塔分布，对应 $\mu=0$分别为 0.3、0.5 和 0.8 的情况。在均值给定的条件下，拟合分布与原始直方图可能看起来大不相同。

~~图 4 见此处~~：~~比例投票率的直方图~~，以及在 μ 取值为 0.3（虚线）、0.5（实线）和 0.8（点线）时，$\phi=46.9$ 服从贝塔分布的拟合概率密度函数。

4.2 阿拉斯加港海豹数量

30多年来，美国政府下属的美国国家海洋和大气管理局渔业局阿拉斯加渔业科学中心海洋哺乳动物实验室，一直在阿拉斯加全境开展港海豹航空调查。这些调查主要在夏末海豹换毛期间进行，是监测和估算港海豹数量的主要方法（Muto et al., 2022）。根据遗传采样结果，阿拉斯加的所有海豹被划分为13个不同的“种群”，即遗传群体。每个种群都有相应的数量估算；本文使用的种群是锡特卡／查塔姆海峡种群。该数据集包含716个观测值，采集年份为1998年、2003年、2008年至2011年以及2015年。为了调节体温，海豹常会爬出水面；与在水中相比，这样更容易清点。该种群所有已知的港海豹上岸地点被划分为74个样本多边形。有些样本多边形每年会进行多次普查，另一些则在某些年份被跳过。每次航空计数普查的解释变量包括距低潮时间和一天中的时间。

我们在表1中使用泊松模型和负二项模型，为海豹计数数据构建分层广义线性混合模型（GLMMs）。我们考虑模型（1）中的一个特定情形:

$$\mathbf{w} = \mathbf{X}\boldsymbol{\beta} + \mathbf{r}_1 + \mathbf{Z}_2\mathbf{r}_2 + \boldsymbol{\epsilon},$$

其中，$\mathbf{X}$ 包含一列总体均值；73列用于表示各样本多边形的均值效应指标（相对于第一个多边形的均值偏差，第一个多边形的效应已并入总体均值）；一列表示一天中的时间，即从太阳正午（太阳位于天顶时）起所经过的日内时间比例；一列表示距低潮时间，即距低潮的小时数（该区域的潮汐周期约为12小时）；以及表示一天中的时间平方和距低潮时间平方的列。因此，$\mathbf{X}$ 共有78列。假定 $n \times 1$ 随机效应$\mathbf{r}_1$ 具有平稳一阶自回归（AR1）时间序列模型的协方差结构，因此 $\text{cov}(r_{1,i}, r_{1,j}) = \sigma_1^2 \rho^{|t_i - t_j|}/(1-\rho^2)$，其中 t_i和 t_j 分别为第 i 个和第 j个观测值计数所在的年份。此外，我们假定不同多边形的计数彼此独立，因此自协方差仅存在于同一多边形各年份之间，从而得到具有分块对角结构的协方差矩阵。然而，有些多边形在同一年内进行了重复测量，因此$\mathbf{Z}_2$ 是由多边形与年份交互作用构成的设计矩阵，而 $\sigma_2^2 \mathbf{Z}_2 \mathbf{Z}_2'$ 允许同一多边形内每年重复采样之间存在额外相关性。并非所有年份都进行了重复测量，且$\mathbf{Z}_2$ 有518列，其中σ_2^2 为$\mathbf{r}_2$的方差。使用泊松分布模型时，我们采用式（1）中的 $\boldsymbol{\epsilon}$ 来允许存在额外的不相关过度离散，其中 $\text{var}(\boldsymbol{\epsilon}) = \sigma_0^2$；而对于可直接处理过度离散的负二项分布，我们令 $\boldsymbol{\epsilon} = \mathbf{0}$。

表6：港海豹计数数据在负二项分布和泊松模型下的估计的固定效应及相关量。表头与表4相同。

效应	负二项分布 Est.	标准误$_u$	标准误$_c$	t值	p值	泊松 标准误$_u$	标准误$_c$
距低潮时间	-0.081	0.00005	0.042	1.938	0.0530	0.041	0.043
(距低潮时间)^2	-0.064	0.00003	0.026	2.495	0.0128	0.025	0.026
一天中的小时	-0.273	0.00013	0.107	2.542	0.0113	0.104	0.107
(一天中的小时)^2	-0.747	0.00017	0.146	5.120	$<$0.0001	0.139	0.144

对于具有AR1协方差结构的泊松模型，使用限制性最大似然法并依据式（4），我们估计了 $\sigma_0^2 = 0.859$、 $\sigma_1^2 = 0.660$、 $\sigma_2^2 = 0.0003$ 和 $\rho = 0.940$；而对于具有AR1协方差模型的负二项分布，我们估计了 $\sigma_1^2 = 3.637$、 $\sigma_2^2 = 0.0012$、$\rho = 0.997$和 $\phi = 1.529$。式（4）中负二倍对数似然的最小值，泊松模型为8416.11，负二项模型为8242.07，表明负二项模型更为合适。表6列出了负二项模型的固定效应估计值，但也一并列出泊松模型的朴素方差和校正方差。尤为值得注意的是，当将 $\boldsymbol{\epsilon}$ 纳入泊松模型的$\mathbf{w}$ 值时，$\mathbf{C}\beta$ 的对角线元素较大；按照式（12）通过$\mathbf{B}(-\mathbf{H}_\mathbf{a}^{-1})\mathbf{B}'$ 进行调整后，这些元素仅略有增加。这与负二项模型形成对比：其$\mathbf{C}\beta$的对角线元素非常小，这一点从表6中的 s.e.$_u$ 可见。然而，负二项模型和泊松模型的 s.e.$_c$ 几乎完全相同；这并不令人意外，因为负二项分布与对数正态泊松分布具有相同的方差—均值关系，仅参数化方式不同（例如，Nakashima, 1997年）。

拟合的　　解释变量有助于深入了解海豹的行为，表明海豹　　s

更喜欢在正午前后和低潮时离开水中上岸休息。模型证实，这些时段的计数最高，因为表6中的系数绘制在图5中。图中显示的是对数尺度上的变化，其他解释变量均固定为零，因此图5可解读为：解释变量每变化一个单位，预期计数的比例变化的对数。

图5见此处：A）一天中的小时和 B）距低潮时间对港海豹数量的拟合效应。拟合效应显示在其他所有协变量均保持为零时，预期比例变化的对数。

此处为图6：74个地点中4个地点的预测 exp(w) 值。空心圆表示原始计数，实心圆表示经过反变换、恢复至数据原始尺度后的预测 exp(w) 值，并由实线连接。虚线表示经过反变换后的预测区间。

我们的最终目标是利用第2.4节中的结果，预测每个地点的海豹丰度。对74个不同地点中的4个地点取指数后的预测 w值如图6所示，这4个地点标记为AC10、AC11、BC01和BC02。对于每个年份，我们使用式（15）并结合预测区间，对式（13）中的$\hat{\mathbf{u}}$ 进行预测，然后对预测值和预测区间都取指数。可以看出，误差较大，但考虑到每个地点的观测值相对较少，这并不令人意外。请注意，在估计自相关参数 ρ时，我们借助了不同地点之间的信息，假设所有地点的$\mathbf{w}$具有相同程度的自相关。还请注意，总体而言，预测值往往会向该地点的总体均值收缩；但尤其对于地点BC02，预测值高于观测值。这是因为预测值已针对解释变量（一天中的时间和距低潮的时间）的最佳条件进行了标准化。BC02地点几乎每次都是在次优条件下计数的，这完全有可能是因为太阳正午时可能正值高潮。如果不愿意等待并仅在低潮时进行采样，就不可能通过抽样设计来优化解释变量

发生在接近太阳正午时。

4.3 苔藓中的重金属浓度

克鲁森施特恩角国家公园位于美国阿拉斯加西北部，附近是红狗矿，这里开采锌、铅、镉和其他重金属。卡车沿一条穿越克鲁森施特恩角国家公园的道路，将矿石从红狗矿运输到海岸。有人推测，这些卡车会使粉尘逸散到环境中。苔藓从空气中获取了大量营养物质，因此是监测附着于空气粉尘上的重金属的理想生物监测器。2001年（Hasselbach等，2005年）以及2006年（Neitlich等，2017年），研究人员两次对苔藓进行重金属采样，且靠近道路的区域采样更加密集。采集了苔藓组织当年的年生长部分，随后将其研磨、均质化，并送往实验室分析。在此，我们仅考虑铅浓度，尽管还分析了许多其他元素。纳入的潜在重要解释变量包括距运输道路的距离、道路一侧（北侧或南侧）以及采样年份。数据集共有365条记录，其中244条来自2001年，121条来自2006年。

铅浓度本身具有正值特性，且通常呈偏态分布，因此 Hasselbach 等人（2005年）和 Neitlich 等人（2017年）将响应变量转换为对数尺度。不同的是，本文采用表1中给出的伽马分布和逆高斯模型。对于协方差结构，我们考虑线性模型（1）的一种特殊情形，

$$\mathbf{w} = \mathbf{X}\boldsymbol{\beta} + \mathbf{r}_1 + \mathbf{Z}_2\mathbf{r}_2 + \mathbf{Z}_3\mathbf{r}_3 + \boldsymbol{\epsilon},$$

其中，$\mathbf{X}$ 包含一列总体均值、一列表示年份2006年的指示列（2001年已纳入总体均值）、距道路距离（以米为单位）的对数，以及一列表示道路南侧的指示列（北侧已纳入总体均值）。我们还考虑了距道路距离的对数与道路所在侧之间的交互作用。

随机效应 $\mathbf{r}_1$ 被假定具有地统计自协方差结构，具体由指数模型给出，即 $\operatorname{cov}(r_1(\mathbf{s}_i)\ r_1(\mathbf{s}_j)) = \sigma_1^2 \exp(-\delta_{i,j}/\rho)$，其中 $\mathbf{s}_i$ 是一个包含第 i 个地点空间坐标的向量，而 $\delta_{i,j}$ 是地点 $\mathbf{s}_i$ 和 $\mathbf{s}_j$ 之间的欧氏距离。参数 σ_1^2 通常称为部分基台，而 ρ 是变程参数，用于控制自协方差随距离增加而衰减的速率。 $\boldsymbol{\epsilon}$ 的方差 σ_0^2 通常称为块金效应。我们假设不同年份的响应相互独立。在同一年份和同一地点内，部分地点采集了重复样本，以考虑微尺度变异；也就是说，先抓取一把苔藓，然后再抓取另一把（假设两次抓取之间的距离为零）。因此，$\mathbf{Z}_2$ 是一个包含地点指示变量的设计矩阵；这会使来自同一地点的任何样本具有更高的自相关性。部分样本被研磨成两个重复样本用于实验室分析，因为测量浓度的仪器或样本均质化方式可能存在一定变异。因此，$\mathbf{Z}_3$ 是一个包含嵌套于地点内的重复样本指示变量的设计矩阵；由于这些重复样本来自同一个重复样本组，因此它们具有更高的自相关性。

采用伽马分布和指数协方差模型，并使用最大似然（而非限制性最大似然）时，负两倍对数似然等于 2318.253；结果显示，道路侧的主效应并不显著。当固定效应发生变化时，限制性最大似然无法提供用于似然比较的嵌套模型（Verbeke 和 Molenberghs, 2000，第 75 页），因此我们使用最大似然而非限制性最大似然。我们去掉该主效应后重新拟合模型，式（4）的负两倍对数似然等于 2319.391。无论使用赤池信息准则还是似然比检验，我们都有证据从模型中剔除道路侧的主效应。这个包含交互作用、但不包含道路侧主效应的模型假定，在距运输道路 1 m 处，北侧和南侧的$\mathbf{X}\beta$ 相同。完成这一步后，协方差参数的边际估计为 $\sigma_1^2 = 0.1703$、$\sigma_2^2 = 0.0634$、$\sigma_3^2 = 0.0266$、$\sigma_0^2 = 0.0023$、$\rho = 9.033$ 和 $\phi = 2218$。我们拟合了一个逆高斯模型，其

均值和协方差结构相同；该模型的负两倍对数似然为 2319.354，与伽马模型的数值几乎相同。逆高斯模型的协方差参数估计值为 $\sigma_1^2 = 0.1699$、 $\sigma_2^2 = 0.0650$、 $\sigma_3^2 = 0.0265$、$\sigma_0^2 = 0.0028$、 $\rho = 9.133$ 和 $\phi = 2382$，与伽马模型中相应的参数几乎相同。有趣的是，如果对数浓度采用正态空间模型拟合，我们得到的估计值为 $\sigma_1^2 = 0.1702$、 $\sigma_2^2 = 0.0633$、$\sigma_3^2 = 0.0267$、 $\sigma_0^2 = 0.0028$、 $\rho = 9.055$，与伽马模型和逆高斯模型的估计值都几乎相同。

表4：使用伽马分布时苔藓铅数据的估计固定效应表。标题与表4相同；这里计算p值时，我们使用自由度为365 - $4 = 361$的t分布。

效应	Est.	标准误$_u$	标准误$_c$	t值	p值
截距	8.074	0.20148	0.20178	40.0169	<0.0001
Year	-0.442	0.22149	0.22157	1.9932	0.0470
到道路的距离	-0.578	0.01835	0.01839	31.4377	<0.0001
距道路距离：南侧	-0.112	0.01165	0.01166	9.6119	<0.0001

三个模型的固定效应表估计值几乎完全相同，因此这里只给出伽马模型的结果；见表7。请注意，与前两个示例不同，在本例中，基于 s.e.$_u$ 和 s.e.$_c$ 得出的估计固定效应标准误差异很小。当几乎所有变异都由 $\mathbf{w}$ 捕获，且 $\log[\mathbf{y}|\mathbf{g}^{-1}(\mathbf{w})\ \phi]$ 的贡献相较之下很小时，就可能出现这种情况。对于伽马分布和逆高斯分布，这一贡献部分受 ϕ 控制，而其估计值非常大。为直观说明这一点，请看图7。原始数据的直方图（图7A）显示，这些数据具有极端偏态。然而，当均值为40时（以 w 值为条件），拟合的伽马分布相当窄，且

呈对称分布（图7B）。当均值增至150（图7C）和850（图7D）时，伽马分布变得更加分散，这是因为伽马分布的方差随 μ^2 增大。

图7见此处：A）苔藓中铅浓度的直方图，B）在$\mu = 40$ 处、使用 $\phi = 2218$ 拟合的伽马分布概率密度（实线），C） $\mu = 150$，以及D） $\mu = 850$。

对 w-值的预测对于伽马分布模型和逆高斯分布模型而言相似，而且二者与对数浓度的正态模型也非常相似。图8展示了伽马模型下2001年和2006年按年份对**w** 的预测。预测位置被分为三组，其中一组在运输道路附近间距较小，随着远离道路，各后续组的位置间距逐渐增大。显然，道路附近的预测值最大（图8）；同时，预测值总体上从2001年到2006年有所下降，这可能是由于从2001年到2006年间发生了变化：运输道路上运载矿石的卡车开始使用覆盖物。盛行风来自南侧，显然道路北侧的预测值更高。预测标准误呈现出地统计学中典型的模式，即在样本附近或样本密集区域（图8）比远离样本的区域更小。

图8见此处：阿拉斯加克鲁森施特恩角国家公园内运输道路附近各地点在2001年和2006年的预测值及其标准误。绿色 $\times$ 符号表示样本位置。

---

5 讨论与结论

我们开发了一套非常灵活的框架，可在分层广义线性混合模型框架中对二元、计数、正连续型及其他类型的数据进行建模。几乎

统计学中已知的众多分布几乎可以涵盖任何数据类型，而这些分布几乎都能与任意模式化协方差矩阵相匹配，表1列出了一份简要清单。我们的示例展示了表1中的所有分布；对于协方差矩阵，我们使用了 CAR 和 SAR 空间自回归模型、AR1 时间序列模型以及指数地质统计模型。我们还展示了如何将随机效应与其他协方差结构混合，以构建复杂的协方差矩阵。在 HGLMM 框架中，任何协方差矩阵都是可能的，包括时空协方差、球面数据的协方差，以及针对溪流、道路等线性网络推导出的协方差等。使用拉普拉斯近似，得到的对数似然由数据分布的对数似然、针对正态分布数据的最大似然或限制性最大似然对数似然，以及海森矩阵的行列式（4）组成。

我们针对线性模型中最常见的三类目标之一，推导了边际推断。首先，为了估计固定效应并进行预测，我们必须估计所有协方差参数，这可通过式（4）完成。随后，由于 $\mathbf{w}$ 是模型中的潜在变量而非观测值，因此有必要对方差进行调整；估计固定效应时通过式（12）进行调整，在未采样位置进行预测时则通过式（15）进行调整。

这些模型的计算量可能很大，因为它们需要计算海森矩阵的行列式；在我们的实现中，还需要计算其逆矩阵。似然优化需要进行双重迭代，因为在优化协方差参数的似然过程中会使用牛顿-拉夫森更新，每次更新都需要$\mathbf{H}^{-1}$。尽管这可能会限制我们 HGLMM框架所能处理的数据集规模，但我们想指出一些节省时间的方法。首先，从式（10）可以看出

$$\mathbf{H} = [\mathbf{D}_{\boldsymbol{\phi}} - \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}] + [\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{X}](\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{X})^{-1}[\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}],$$

我们在此添加括号，以表明该式具有 Sherman-Morrison-Woodbury 形式$\mathbf{A}+\mathbf{BCB}'$（Sherman 和 Morrison，1949；Woodbury，1950）。如果$\mathbf{A}$ 是 $n \times n$，但其逆矩阵可以快速求得，且$\mathbf{C}$ 的维度较小，那么逆矩阵（$\mathbf{A}+\mathbf{BCB}'$）$^{-1}$ 的计算速度可以远快于完整的$n \times n$ 逆矩阵。例如，考虑我们关于港海豹的第二个示例，其中有 716 条记录，分布在 74 个采样地点。我们假设地点内部遵循时间序列模型，而不同地点之间相互独立，因此协方差矩阵具有分块对角结构。因此，$\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}$ 具有分块逆矩阵，而$\mathbf{D}_{\boldsymbol{\phi}}$ 是对角矩阵，所以 $[\mathbf{D}_{\boldsymbol{\phi}}-\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}]^{-1}$ 可以按块求逆，这比对整个 $n \times n$ 矩阵求一次逆快得多。同样，$\mathbf{Z}_k\mathbf{Z}_k'$通常是分块对角的，而且可以递归地对多个方差分量应用 Sherman-Morrison-Woodbury 定理（Dumelle 等，2021）。

对于这些模型而言，一个重要的考量是式 (1) 中的独立成分 $\boldsymbol{\epsilon}$ 与 $\boldsymbol{\phi}$ 在 $[\mathbf{y}|\mathbf{g}^{-1}(\mathbf{w})\ \boldsymbol{\phi}]$ 中的相互作用。$\boldsymbol{\phi}$ 中的参数通常控制方差，并且可能与 $\boldsymbol{\epsilon}$ 混淆。作为一个极端示例，假设 $[\mathbf{y}|\mathbf{g}^{-1}(\mathbf{w})\ \boldsymbol{\phi}]$ 是一个正态分布，其中 $\mathbf{g}^{-1}(\mathbf{w})$ 对每个元素都采用恒等函数，而 ϕ 只有一个元素——方差参数。那么，ϕ 和 σ_0^2 将无法识别。更常见的情况是，$\boldsymbol{\phi}$ 控制方差与均值的关系，但我们预计仍可能存在一定程度的混淆。对于任何特定的数据集，都可以通过绘制 σ_0^2 和 ϕ 的对数似然图（类似于图1C）来进行研究，或者凭借对特定模型中这些参数如何相互作用的更多经验来判断。

本文中的 HGLMM框架可与广义线性模型的混合模型扩展进行对比。广义线性模型框架以正则指数分布族为基础，由此得到所谓的“典范”链接函数。例如，伽马分布的典范链接函数是 $-1/\mu$，但通常会将其改为 $w = g(\mu) = 1/\mu$。然而，这意味着 $\mu = g^{-1}(w) = 1/w$，但由于 w 可能为负值，因此 μ 也可能取负值。在采用伪似然和迭代加权最小二乘法的矩估计建模框架中，只要这些值

与参数空间足够接近，这种情况就可以接受，同时也允许使用多种链接函数，从而提供很大的灵活性。然而，在完全参数化的 HGLMM框架中，如果 $\mathbf{y}|\mathbf{g}^{-1}(\mathbf{w})\ \phi]$ 中的$\mathbf{g}^{-1}(\mathbf{w})$ 超出均值的参数空间，就无法计算对数似然。对于广义线性混合模型，必须选择符合参数空间要求的链接函数和均值函数。

我们概述了 HGLMM 下的边际推断。还有许多未提及的主题值得探讨。例如，我们可能希望对预测进行推断，其中 $\mathbf{w}$ 经过反变换成为 $\mathbf{g}^{-1}(\mathbf{w})$，并且将 $\mathbf{y}|\mathbf{w}$ 的变异性纳入其中。我们也可能希望对 $\mathbf{g}^{-1}(\mathbf{w})$的其他函数进行推断，例如区块平均值。同样，我们可能希望对式 (1) 中 $\mathbf{r}_i$ 的随机效应（最佳线性无偏预测）进行推断。与大多数线性模型一样，在对拟合模型、处理效应等进行推断时，我们可以考虑β的线性组合，或 β 参数的对比。本文仅介绍了基本框架，还有许多有待开展的后续研究课题。

致谢

本项目获得了美国国家海洋渔业局、美国国家海洋和大气管理局以及美国环境保护署（美国环境保护署）的资金支持。论文中的研究结果和结论仅代表作者的观点，不一定代表审稿人、美国环境保护署或美国国家海洋渔业局、美国国家海洋和大气管理局的观点。提及任何商品、产品或公司名称，并不意味着美国政府对其予以认可。

数据和软件可用性

所有数据和代码均位于 GitHub代码仓库中,
`https://github.com/jayverhoef/hglmm.git`

参考文献

Berliner, L. M.（1996年），《层次贝叶斯时间序列模型》，载于《最大熵与贝叶斯方法：*1995*年圣菲（美国新墨西哥州）会议录》，第十五届最大熵与贝叶斯方法国际研讨会，Springer，第15–22页。

Besag, J.（1974），《格点系统的空间相互作用与统计分析（附讨论）》，《皇家统计学会会刊》*B*辑，36，192–236。

Bivand, R.、Nowosad, J. 和 Lovelace, R.（2023年），*spData*：空间分析数据集，R软件包，版本 2.2.2。

Bonat, W. H. 和 RibeiroJr, P. J.（2016年），《空间广义线性混合模型的实用似然分析》，*Environmetrics*，27，83–89。

Breslow, N. E. 和 Clayton, D. G.（1993），《广义线性混合模型中的近似推断》，《美国统计学会会刊》，88，9–25。

Brooks, M. E., Kristensen, K., Van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H.J., Machler, M. 和 Bolker, B.M.（2017年），《glmmTMB 在零膨胀广义线性混合建模软件包中兼顾速度与灵活性》，《*R*杂志》，9，378–400。

Chiles, J.-P. 和 Delfiner, P.（1999），地统计学：空间不确定性建模，纽约：John Wiley & Sons。

Christensen, O.F.（2004），「基于模型的地统计学中的蒙特卡洛最大似然」，计算与图形统计学杂志，13，702–718。

Clayton, D. 和 Kaldor, J.（1987年），“用于疾病制图的年龄标准化相对风险经验贝叶斯估计”，《生物统计学》，43，671–681。

Cressie, N. 和 Wikle, C. K.（2011），时空数据统计学，新泽西州霍博肯：John Wiley & Sons。

Cressie, N. A. C.（1993），空间数据统计学（修订版），纽约：John Wiley & Sons。

Diggle, P. J.、Tawn, J. A. 和 Moyeed, R. A.（1998年），“基于模型的地统计学（附讨论）”，皇家统计学会会刊，*C*辑：应用统计学，47，299–326。

Dumelle, M.、Ver Hoef, J. M.、Fuentes, C. 和 Gitelman, A.（2021），“时空随机过程的线性混合模型表述及乘积、求和和乘积–求和协方差函数的计算改进”，空间统计学，43，100510。

Efron, B. 和 Hinkley, D. V.（1978），“评估最大似然估计量的准确性：观测费舍尔信息与期望费舍尔信息”，《生物统计学》，65，457–483。

Evangelou, E.、Zhu, Z. 和 Smith, R. L.（2011），“使用高阶拉普拉斯近似对空间广义线性混合模型进行估计与预测”，《统计规划与推断杂志》，141，3564–3577。

Fisher, R. A.（1934），《数学似然的两个新性质》，《伦敦皇家学会会刊：数学与物理学论文集》，144，285–307。

Gelfand, A. E. 和 Smith, A. F. M.（1990年），《计算边际密度的基于抽样的方法》，《美国统计学会会刊》，85，398–409。

Gilks, W. R.、Richardson, S. 和 Spiegelhalter, D. J.（1996年），《介绍马尔可夫链蒙特卡罗》，载于 马尔可夫链蒙特卡罗实践，Chapman and Hall，第1–19页。

Gneiting, T.（2013），“球面上的严格正定和非严格正定函数”，*Bernoulli*，19，1327–1349。

Goldberger, A. S.（1962年），“广义线性回归模型中的最佳线性无偏预测”，《美国统计学会会刊》，57，369–375。

Gotway, C. A. 和 Stroup, W. W.（1997年），“一种用于空间数据分析和预测的广义线性模型方法”，农业、生物与环境统计学杂志，2，157–178。

Haining, R. P.（1978），“空间相互作用的移动平均模型”，英国地理学家协会会刊，3，202–225。

Hamilton, J. D. (1994), 时间序列分析, 第 2 卷，美国新泽西州普林斯顿：普林斯顿大学出版社。

Hasselbach, L., VerHoef, J.M., Ford, J., Neitlich, P., Crecelius, E., Berryman, S., Wolk, B., 和 Bohle, T.（2005年），“阿拉斯加红狗矿附近国家公园管理局土地及其邻近区域镉和铅沉积的空间模式”，整体环境科学，348，211–230。

Huang, C., Zhang, H. 和 Robeson, S. M.（2011），“关于常用协方差函数和球面上的变异函数的有效性”，数学地球科学，43，721–733。

Kleinman, K. P. 和 Ibrahim, J. G.（1998），“广义线性混合模型的半参数贝叶斯方法”，医学统计学，17，2579–2596。

Lee, Y. 和 Nelder, J. A.（1996年），《分层广义线性模型》，皇家统计学会会刊：*B*辑（方法论），58，619–656。

Lehmann, E. L. 和 Casella, G.（2006年），点估计理论，Springer Science & Business Media。

Lunn, D. J., Thomas, A., Best, N. 和 Spiegelhalter, D.（2000），《WinBUGS——一种贝叶斯建模框架：概念、结构与可扩展性》，统计学与计算，10，325–337。

McCullagh, P. 和 Nelder, J. A.（1989），《广义线性模型（第*2*版）》，Chapman & Hall Ltd.

Nakashima, E.（1997年），《负二项模型中估计的若干方法》，统计数学研究所年刊，49，101–115。

Neitlich, P. N., Hoef, J. M. V., Berryman, S. D., Mines, A., Geiser, L.H., Hasselbach, L. M., 和 Shiel, A. E.（2017年），“阿拉斯加红狗矿运输道路沿线国家公园管理局土地上重金属沉积空间格局的变化趋势，2001年–2006年”，*PLOS ONE*，12，e0177936。

Nelder, J. A. 和 Wedderburn, R. W. M.（1972），“广义线性模型”，皇家统计学会会刊，*A*辑：综合，135，370–384。

Pace, R. K. 和 Barry, R.（1997年），《空间自回归估计量的快速计算》，《地理分析》，29，232–247。

Patterson, H. 和 Thompson, R.（1974），《方差分量的最大似然估计》，载于第八届国际生物统计学会议论文集，生物统计学会，华盛顿特区，第197–207页。

Patterson, H. D. 和 Thompson, R.（1971），“区组大小不等时的区组间信息恢复”，《生物统计学》，58，545–554。

Rue, H.、Martino, S. 和 Chopin, N.（2009），“使用积分嵌套拉普拉斯近似对潜在高斯模型进行近似贝叶斯推断”，《皇家统计学会会刊：*B*辑（统计方法学）》，71，319–392。

Sherman, J. 和 Morrison, W. J.（1949），“当原矩阵某一列或某一行的元素发生变化时，对应逆矩阵的调整”，《数理统计年刊》，20，621。

Snedecor, G. W. 和 Cochran, W. G.（1980），《统计方法》，第七版，爱荷华州立大学。

Stiratelli, R.、Laird, N. 和 Ware, J. H.（1984），“二元响应的序列观测随机效应模型”，《生物统计学》，40，961–971。

Tierney, L. 和 Kadane, J. B.（1986年），《后验矩和边缘密度的精确近似》，《美国统计学会会刊》，81，82–86。

Ver Hoef, J. M.（2018），《线性网络和非欧几里得距离的克里金模型：注意事项与解决方案》，《生态学与进化方法》，9，1600–1613。

Ver Hoef, J. M. 和 Peterson, E.（2010），《河流网络空间统计模型的移动平均方法（附讨论）》，《美国统计学会会刊》，105，6–18。

Ver Hoef, J. M.、Peterson, E. E.、Hooten, M. B.、Hanks, E. M. 和 Fortin, M.-J.（2018），《基于生态数据进行统计推断的空间自回归模型》，《生态学专论》，88，36–59。

Verbeke, G. 和 Molenberghs, G.（2000），纵向数据的线性混合模型，Springer-Verlag Inc.

Warton, D. I. 和 Hui, F. K.（2011），“反正弦变换荒谬至极：生态学中的比例分析”，《生态学》，92，3–10。

Wedderburn, R. W. M.（1974），“准似然函数、广义线性模型和高斯-牛顿法”，《生物统计学》，61，439–447。

Whittle, P.（1954），“关于平面上的平稳过程”，《生物统计学》，41，434–449。

Wood, S. N.（2020），“简化的积分嵌套拉普拉斯近似”，《生物统计学》，107，223–230。

Woodbury, M. A.（1950），修正矩阵的求逆，统计研究组，普林斯顿，新泽西州，出版物：备忘录报告42。

Zeger, S. L. 和 Karim, M. R.（1991），“带随机效应的广义线性模型：一种吉布斯抽样方法”，《美国统计学会会刊》，86，79–86，出版商：[AmericanStatisticalAssociation, Taylor& Francis,Ltd.]。

Zeger, S. L.、Liang, K.-Y. 和 Albert, P. S.（1988），“纵向数据模型：一种广义估计方程方法”，《生物统计学》，44，1049–1060，出版商：[Wiley, International Biometric Society]。

Zhang, H.（2002），“空间广义线性混合模型的估计与预测”，《生物统计学》，58，129–136。

—（2004年），“基于模型的地统计学中的不一致估计与渐近相等插值”，《美国统计学会会刊》，99，250–261。

图

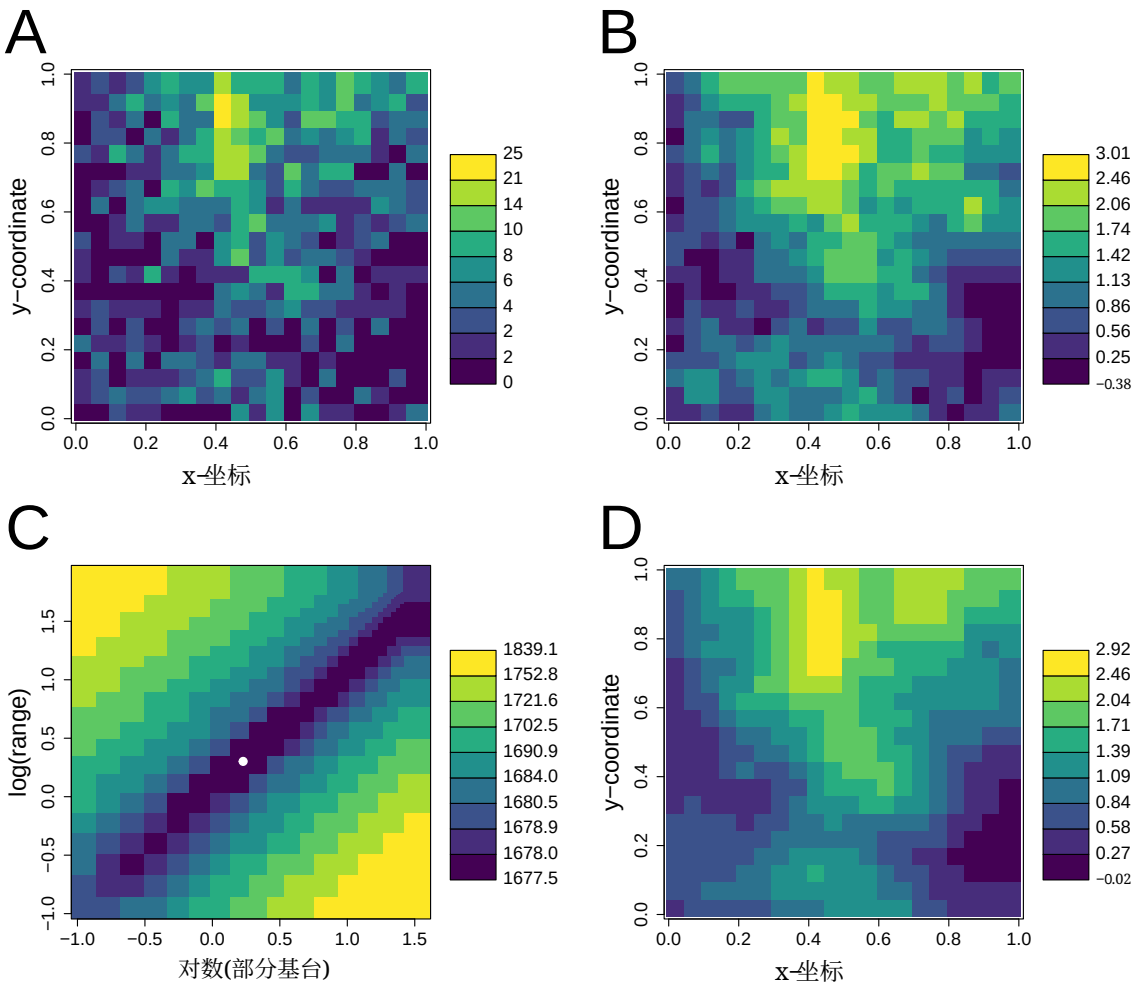

图1：模拟泊松计数数据的估计。A. 使用文中所述模型模拟的计数数据。B. 模拟数据的真实$\mathbf{w}$ 值。C. 模拟数据的边际似然曲面。白色圆圈表示估计值。D. 估计的$\hat{\mathbf{w}}$ 值。

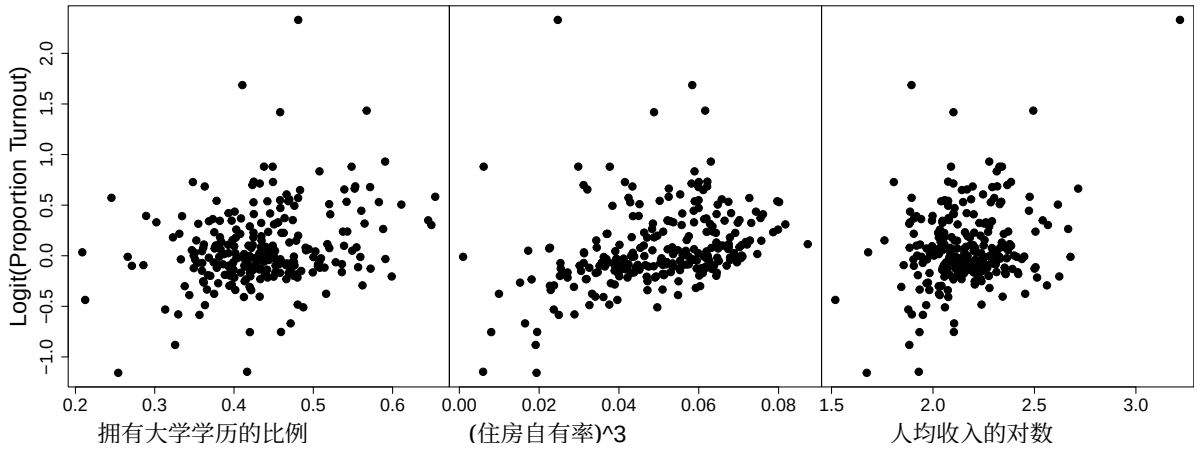

图2：投票率响应变量的 logit变换与三个解释变量的散点图。请注意，部分解释变量进行了变换：住房拥有率取立方，人均收入取自然对数。

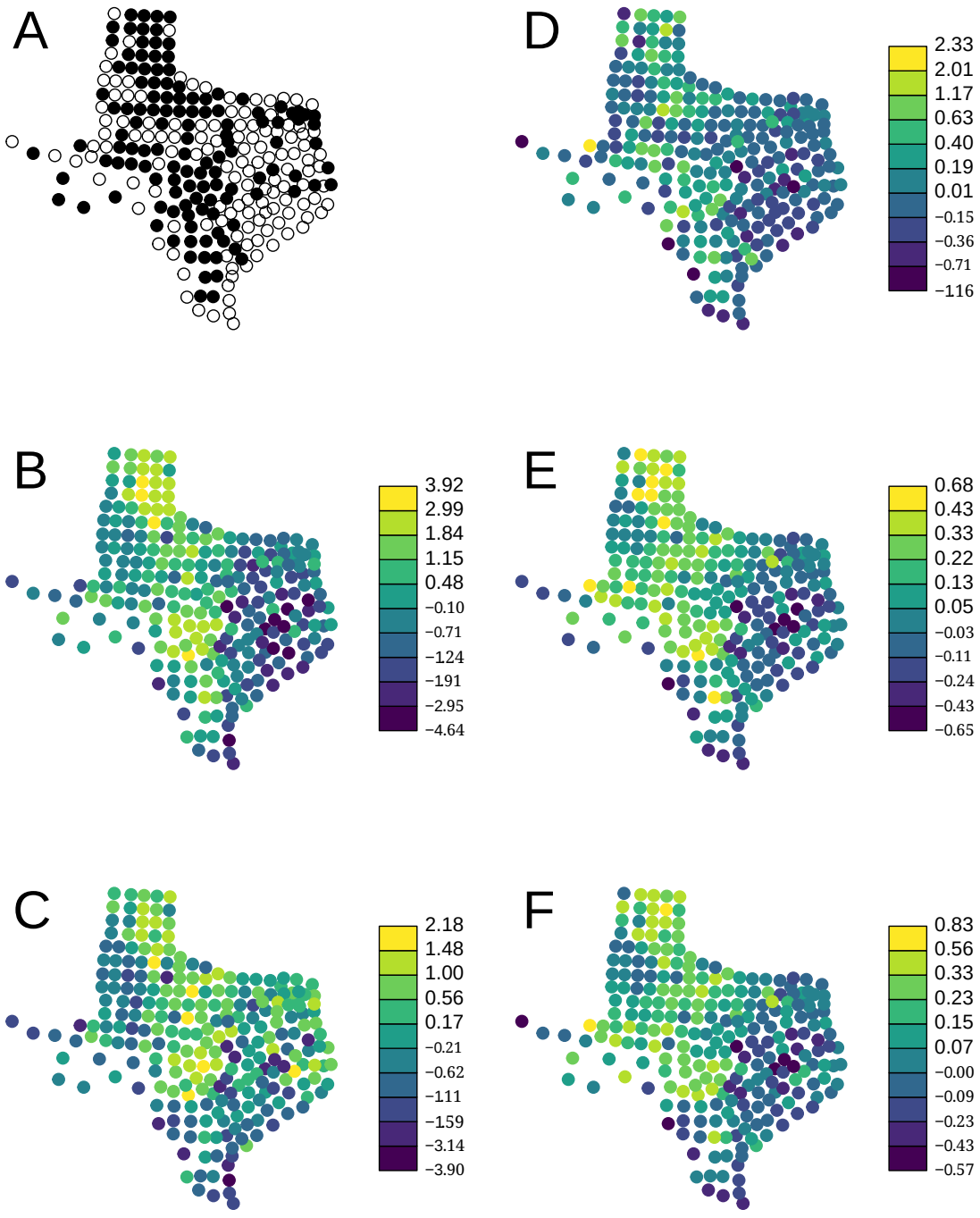

图3：得克萨斯州投票率数据的原始数据和预测的空间随机效应（w）。A）原始二元数据，空心圆表示0，实心圆表示1；B）使用SAR模型对二元数据预测的 $\hat{w}$ ； C）使用CAR模型对二元数据预测的 $\hat{w}$ ； D）经过logit变换的比例投票率数据；E）使用SAR模型对比例投票率数据预测的 $\hat{w}$ ； F）使用CAR模型对比例投票率数据预测的 $\hat{w}$ 。

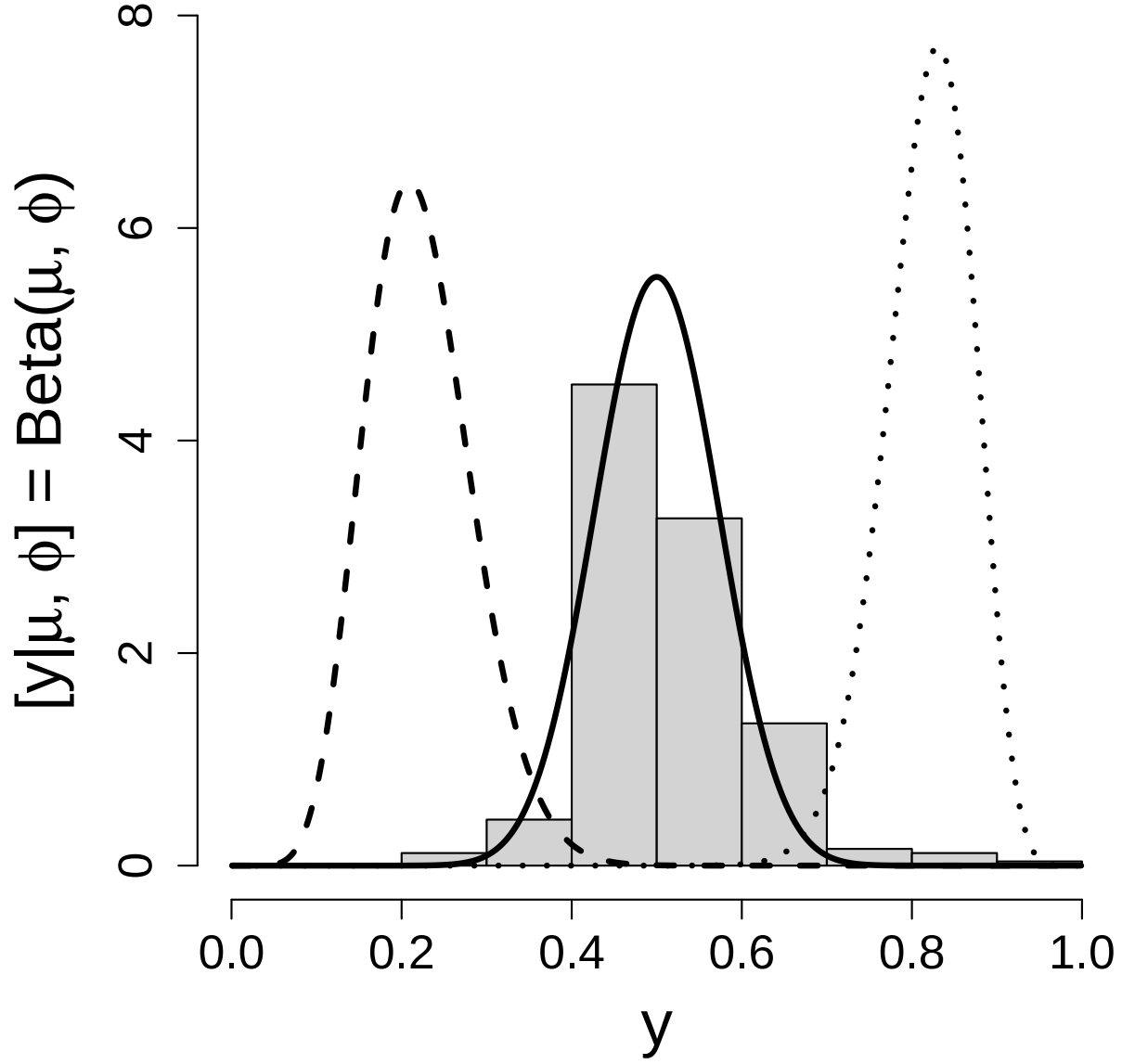

图 4：比例投票率的直方图，以及在 μ 值分别为 0.3（虚线）、0.5（实线）和 0.8（点线）时贝塔分布的拟合概率密度函数，其中 $\phi = 46.9$ 。

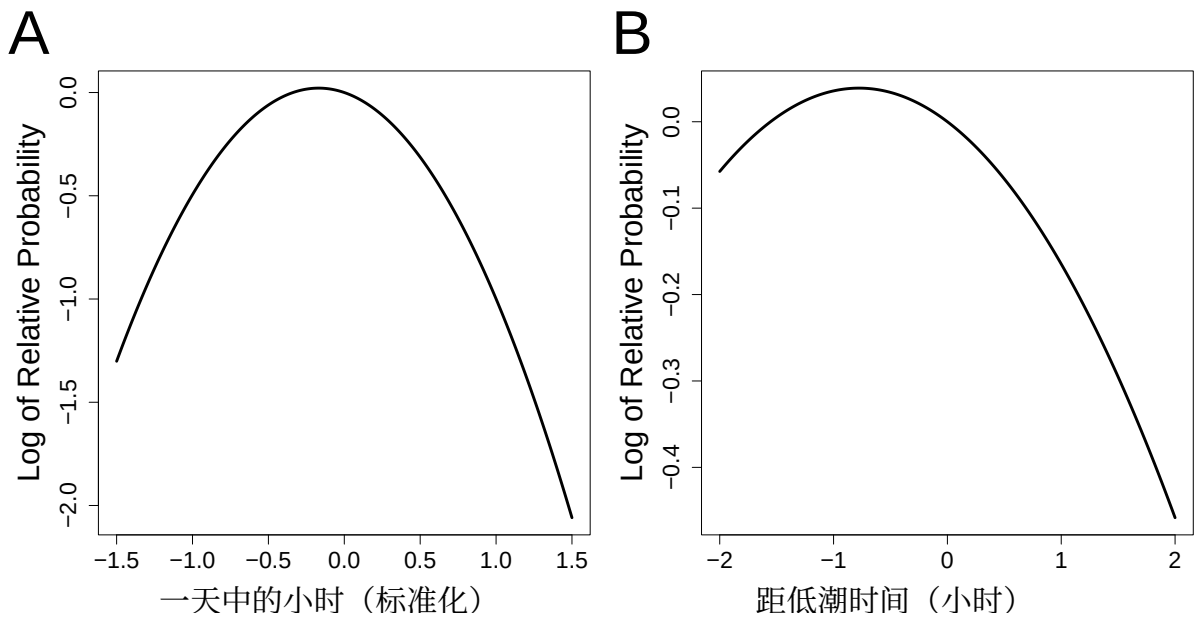

图5：A）一天中的小时和 B）距低潮时间对港海豹数量的拟合效应。拟合效应表示在其他所有协变量均为零时，预期比例变化的对数。

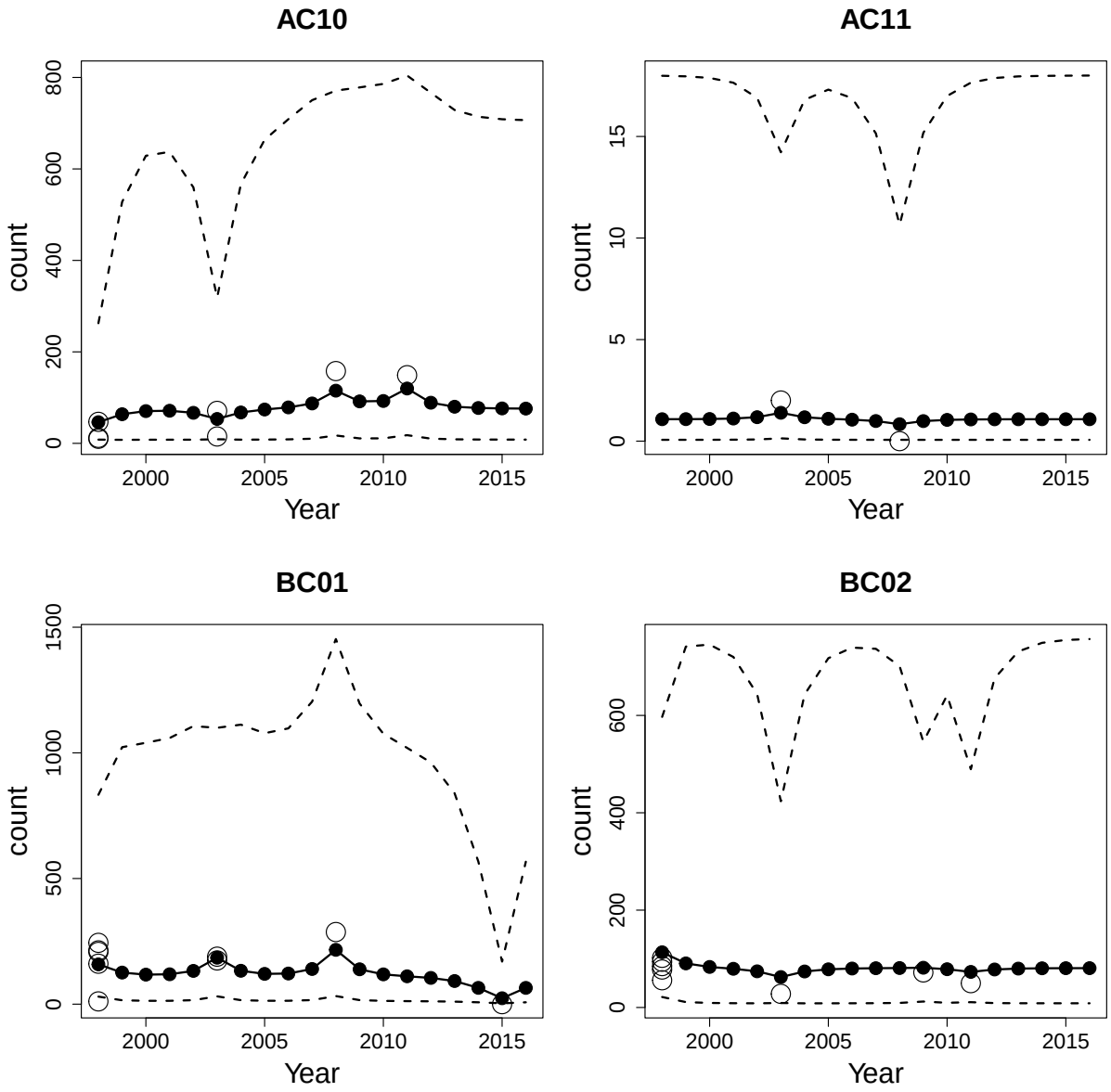

图6：74个地点中4个地点的预测 exp(w) 值。空心圆表示原始计数，实心圆表示将数据反变换至原始尺度后得到的预测 exp(w) 值，并由实线连接。虚线表示反变换后的预测区间。

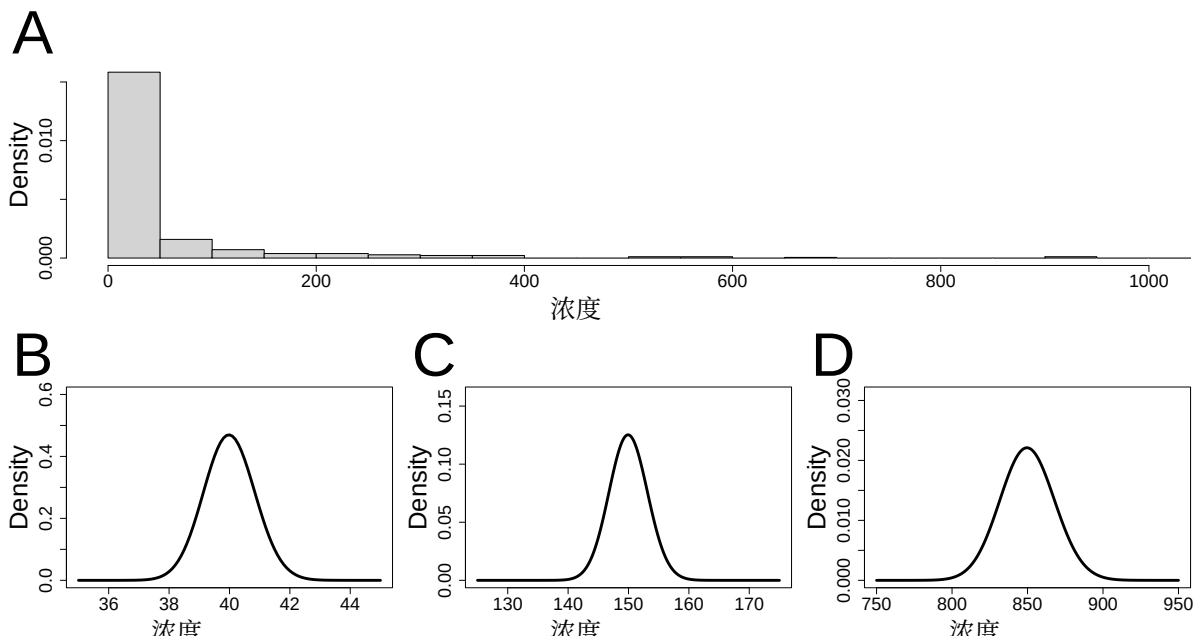

图7：A）苔藓中铅浓度的直方图，B）伽马分布的拟合概率密度（实线），其中 $\mu = 40$为 $\phi = 2218$，C） $\mu = 150$，D） $\mu = 850$。

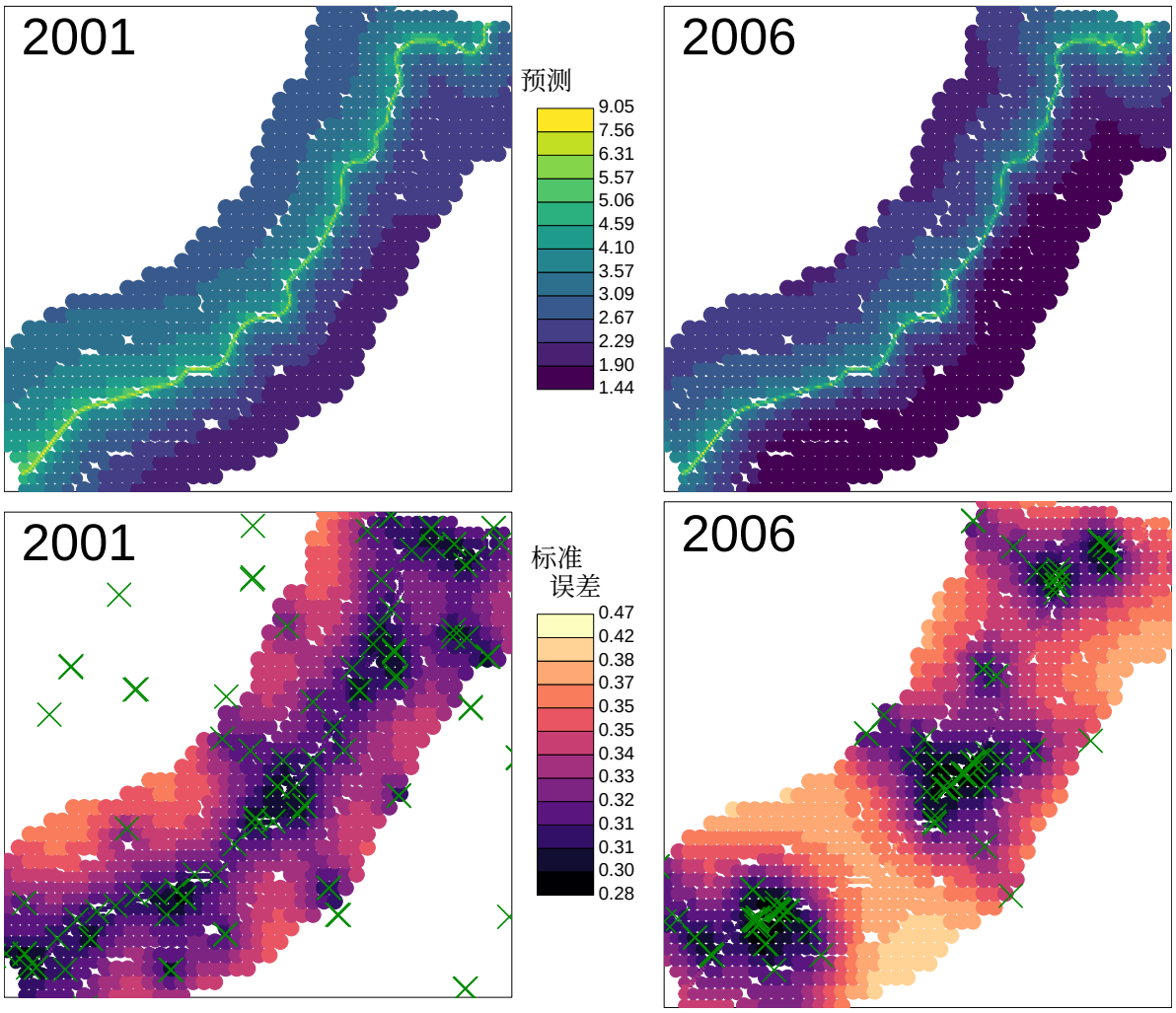

图 8：2001年和2006年阿拉斯加克鲁森施特恩角国家公园内运输道路附近各地点的预测值及其标准误。绿色 × 符号表示样本位置。

6 附录

6.1 通过积分推导REML

考虑一般线性模型的多元正态分布，

$$[\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta}] = \frac{\exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right)}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}}, \tag{A.1}$$

其中 $\mathbf{y}$是响应变量的 $n\times 1$向量，$\mathbf{X}$是解释变量的 $n\times p$设计矩阵， $\boldsymbol{\beta}$ 是固定效应的 $p\times 1$ 向量， $\boldsymbol{\theta}$ 包含协方差参数，这些参数包含在 $n\times n$ 协方差矩阵 $\boldsymbol{\Sigma}$中。通过对固定效应 $\boldsymbol{\beta}$积分，可以得到REML方程，

$$\int_{\mathbb{R}^p} f(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})d\boldsymbol{\beta},$$

从而得到一个仅由协方差参数 $\boldsymbol{\theta}$ 和数据$\mathbf{y}$构成的似然函数。具体而言，

$$-2\ln\left(\int_{\mathbb{R}^p} f(\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta})d\boldsymbol{\beta}\right) = (n-p)\ln(2\pi) + \ln|\boldsymbol{\Sigma}| + \ln|\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}| + (\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}),$$

其中 $\hat{}\ \boldsymbol{\beta}$ =$(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}$。

证明

将 (A.1) 写为

$$\begin{aligned}[\mathbf{y};\boldsymbol{\beta},\boldsymbol{\theta}] &= \frac{\exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}+\mathbf{X}\hat{\boldsymbol{\beta}}-\mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}+\mathbf{X}\hat{\boldsymbol{\beta}}-\mathbf{X}\boldsymbol{\beta})\right)}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}},\\ &= \frac{\exp\left(-\frac{1}{2}[(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}) + (\mathbf{X}\hat{\boldsymbol{\beta}}-\mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}\hat{\boldsymbol{\beta}}-\mathbf{X}\boldsymbol{\beta}) + C\right)}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}},\end{aligned}$$

其中 $C = 2(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{X}\boldsymbol{\beta}) = 0$。提取不含 $\boldsymbol{\beta}$ 的项，

$$\int_{\mathbb{R}^p} [\mathbf{y}; \boldsymbol{\beta}, \boldsymbol{\theta}] d\boldsymbol{\beta} = M \int_{\mathbb{R}^p} \exp\left(-\frac{1}{2}(\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{X}\boldsymbol{\beta})\right) d\boldsymbol{\beta},$$

其中 $M = \exp[-\frac{1}{2}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})]/[(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}]$。注意，

$$\begin{aligned}
&\int_{\mathbb{R}^p} \exp\left(-\frac{1}{2}(\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{X}\boldsymbol{\beta})\right) d\boldsymbol{\beta}, \\
=&\int_{\mathbb{R}^p} \exp\left(-\frac{1}{2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\right) d\boldsymbol{\beta}, \\
=&2\pi^{p/2}|(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})|^{-1/2},
\end{aligned}$$

回顾正定$\mathbf{A}_{m\times m}$ 以及任意维数相容的$\mathbf{x} \neq \mathbf{0}$,

$$\int_{\mathbb{R}^m} \exp(-\mathbf{x}'\mathbf{A}\mathbf{x}/2) d\mathbf{x} = (2\pi)^{m/2}|\mathbf{A}|^{-1/2}.$$

因此，我们得到

$$[\mathbf{y}; \boldsymbol{\theta}] = \int_{\mathbb{R}^p} [\mathbf{y}; \boldsymbol{\beta}, \boldsymbol{\theta}] d\boldsymbol{\beta} = \frac{\exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})\right)}{(2\pi)^{(n-p)/2}|\boldsymbol{\Sigma}|^{1/2}|\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}|^{1/2}},$$

并取 $-2\ln[\mathbf{y};\ \boldsymbol{\theta}]$ 即可得到所需结果。

6.2 分布参数化

6.2.1 负二项分布

对于负二项分布， y_i 是一个具有概率密度函数的非负整数

(PDF)

$$[y|\mu,\phi] = \frac{\Gamma(y+\phi)}{\Gamma(\phi)y!}\left(\frac{\mu}{\mu+\phi}\right)^{y}\left(\frac{\phi}{\mu+\phi}\right)^{\phi},$$

其中 $0<\mu<1, 0<\phi$， $\mathsf{E}(Y) = \mu$， $\mathsf{var}(Y) = \mu+\mu^2/\phi$， 且 $\Gamma(\cdot)$ 是伽马函数。

6.2.2 伽马分布

对于伽马分布， y_i 为正，其概率密度函数为

$$[y|\mu,\phi] = \frac{1}{\Gamma(\phi)}\left(\frac{\phi}{\mu}\right)^{\phi} y^{\phi-1}\exp\left(\frac{-y\phi}{\mu}\right),$$

其中 $0<\mu, 0<\phi$， $\mathsf{E}(Y) = \mu$， 且 $\mathsf{var}(Y) = \mu^2/\phi$。

6.2.3 贝塔分布

对于贝塔分布， $0<y_i<1$ 其概率密度函数为

$$[y|\mu,\phi] = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)}y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1},$$

其中 $0<\mu<1, 0<\phi$, $\mathsf{E}(Y) = \mu$， 且 $\mathsf{var}(Y) = \mu(1-\mu)/(1+\phi)$。

6.2.4 逆高斯分布

逆高斯分布通常表示为:

$$[y;\mu,\lambda] = \sqrt{\frac{\lambda}{2\pi y^3}}\exp\left(-\frac{\lambda(y-\mu)^2}{2\mu^2 y}\right), \tag{A.2}$$

其中 $y > 0$、 $\mu > 0$ 和 $\lambda > 0$。在这种参数化中， λ 是形状参数，且 $\mathrm{E}(Y) = \mu$ ，$\mathrm{var}(Y) = \mu^3/\lambda$。为了使 (1) 中的 μ 保持为正且 w 不受约束，我们令 $\boldsymbol{\mu} = \exp(\mathbf{w})$。然而，根据这种构造，由 (9) 可得

$$D_{i,i} = \frac{(e^{w_i} - 2y_i)}{\phi e^{2w_i}},$$

而某些 $D_{i,\ i}$ 在 $e^{w_i} > 2y_i$ 时可以为正，这可能导致 (10) 中的 $\mathbf{H}$ 奇异。我们提出另一种参数化方式。对于逆高斯模型， λ 通常会进行尺度变换；此处我们也采用这种做法，令 $\phi = \lambda/\mu = \lambda/\exp(w)$，得到一个 μ-尺度-λ逆高斯模型，

$$[y; \mu, \lambda] = \sqrt{\frac{\phi \exp(w)}{2\pi y^3}} \exp\left(-\frac{\phi(y - \exp(w))^2}{2\exp(w)y}\right), \tag{A.3}$$

其中 $\phi > 0$ ，且现在 $\mathrm{var}(Y) = \mu^2/\phi$。在这种参数化方式下，我们有

$$D_{i,i} = -\frac{\phi(e^{2w_i} + y_i^2)}{2ye^{w_i}},$$

该式始终为负，因此 (10) 始终表现良好。在这种构造下，我们还有

$$d_i = \phi\left(\frac{y}{2e^{w_i}} - \frac{e^{w_i}}{2y}\right) + \frac{1}{2}.$$