arXiv:2305.04318v2 [stat.ME] 3 Jul 2023

变换高斯地统计模型中参数的剖面似然

预印本

Ruoyong Xu
统计科学系，多伦多大学，700 University Ave., 加拿大安大略省多伦多市，邮编 M5G 1Z5,
ruoyong.xu@mail.utoronto.ca

Patrick Brown
统计科学系 多伦多大学 700 University Ave., 加拿大安大略省多伦多市 M5G 1Z5
patrick.brown@utoronto.ca

摘要

由于反复分解大型方差矩阵会带来沉重的计算负担，剖面似然很少用于地统计模型。在地统计模型中，考虑协方差参数的不确定性可能会产生重大影响，因为某些协方差参数难以识别；问题严重到通常将 Matérn 相关函数的可微性参数视为固定值。对于各向异性空间模型，问题更加复杂，因为还需考虑两个额外参数。本文作出以下贡献：1，针对采用Matérn族相关函数的高斯空间模型（包括各向异性模型），提出了一种用于计算剖面似然的方法论。该方法论采用一种新颖的重新参数化方式来生成代表点，并在软件实现中使用图形处理器进行并行剖面似然计算。

2，结果表明，Matérn形状参数的剖面似然通常相当平坦，但仍可识别，且通常能够排除非常小的取值。

3，模拟研究以及对真实数据实例的应用表明，基于剖面似然的协方差参数和回归参数置信区间，其覆盖率优于传统的标准 Wald型置信区间。

关键词 剖面似然 · 线性地统计模型 · 并行计算

1 引言

采用 Matérn族相关函数的各向异性空间模型十分重要，但由于涉及许多模型参数，因此难以实现。除了五个相关参数——尺度参数（ϕ_X）、形状（κ）、块金效应（ν^2）、各向异性比率（ϕ_R）和各向异性角度（ϕ_A）——还需要估计方差参数（σ）、Box-Cox变换参数（λ）以及 p 回归参数 $\beta = (\beta_1, \ldots\ldots, \beta_p)$。剖面似然可用于降低完整似然的维数；对于难以处理的参数，通常建议绘制其剖面似然函数。在空间模型中，相关参数的似然值可能与正态分布相差甚远（例如，κ 的剖面似然曲面往往可能十分平坦），而形状、块金效应或方差参数的真实值有时可能接近边界（例如，κ 的真实值可能为无穷大）。在这些情况下，最大似然估计（MLE）处似然的曲率（与Fisher信息矩阵相关）无法有效近似似然面的其余部分，因此标准Wald置信区间（置信区间）并不准确。

大多数空间软件包由于反复进行似然评估所带来的计算负担，不使用剖面似然。令 $\omega = (\phi_X, \kappa, \nu^2, \phi_Y, \phi_A)$ 为相关参数向量，令 $\psi = (\omega, \sigma, \lambda)$ 和$\beta_i, i \in \{1, p\}$ 表示 β 的一个元素。地统计学 R 软件包 ***geostatsp*** 将参数 ψ 固定为其最大似然估计，记作 ψ，并使用伪似然 $\ell(\beta_i; \psi)$ 来评估每个 β_i。对于

ˆ 协方差参数，剖面似然 $\ell_p(\psi, \beta_\psi; y)$ 的二阶导数是相对于 ψ 通过数值方法求得的，以获得在 ψ 处评估的信息矩阵。在地统计学背景下，基于似然的置信区间会很有优势，因为“边界问题”和“非正则似然”时常出现。基于似然的置信区间通过对 ψ 进行最大化计算，即计算剖面似然 $\ell_p(\beta_i) = \ell(\beta_i, \psi_{\beta_i})$，然后根据每个 β_i 的剖面似然曲线推导置信区间。***geostatsp*** 中有一个函数，能够为部分模型参数（仅限 λ 和 ω ）计算基于似然的置信区间。对于 ϕ_X 的固定值，例如，***geostatsp*** 会对其余参数（$\kappa, \nu^2, \phi_R, \phi_A$）的联合似然进行数值优化。对大量不同的 ϕ_X 值重复这一过程，然后通过插值获得插值剖面似然，并据此计算ϕ_X的基于似然的置信区间。此过程需要多次运行优化函数，因此可能非常耗时；如果空间数据包含大量观测值，还可能失败。获得代入式 Wald 置信区间是相对“简单的方法”，因为计算最大似然估计和信息矩阵所需的量都可以从用于求取相关参数最大似然估计的优化器中获得。然而，$\ell(\beta_i; \psi)$ 未能将相关参数估计值的不确定性纳入 β_i 的置信区间，因此从理论上讲，***geostatsp*** 的回归参数置信区间会窄于 β_i 的基于似然的置信区间。在地统计模型中，考虑相关参数的不确定性影响重大，因为部分相关参数难以识别；这一问题严重到 Matérn 相关的可微性参数 κ 通常被视为固定在某个合理值上。

本文提出一种方法，可在不近似协方差矩阵的情况下，计算具有稠密协方差矩阵（Matérn）的高斯线性地统计模型（LGM）（各向异性或各向同性）中所有参数的剖面似然和基于似然的置信区间。该方法利用图形处理器（GPU）的并行计算能力，大幅降低了计算剖面似然的负担，使得可以在图形处理器上并行进行大量不同参数值下的重复模型拟合，并借助 GPU 的本地内存加快计算。该方法已在 R 软件包 ***gpuLik*** 中实现并提供。

本文其余部分安排如下。第 2 节介绍本方法所需的背景知识。第 3 节详细说明本方法的核心思想、步骤和计算工作。第 4 节简要讨论图形处理器计算中的注意事项，并说明如何在图形处理器上并行化计算工作。第 5 节介绍两项模拟研究。第 6 节通过将本方法应用于小型和大型真实数据集，展示其适用性。第 7 节总结全文，并讨论未来研究方向。关于使用限制最大似然对线性地统计模型进行似然推断的说明见附录。

2 预备知识

应用变换处理观测值Y_i，$i = 1$，.....，n，通常可以改善线性地统计模型的拟合效果。对于偏离正态分布的偏态随机场，可以通过 Box-Cox 变换（BCT） [1]进行建模。对于取值为正的观测值 Y_i，Box-Cox变换后的响应 Y_i'形式如下

$$Y_i' = b(Y_i; \lambda) = \begin{cases} (Y_i^\lambda - 1)/\lambda : & \lambda \in \mathbb{R} \quad \text{and} \quad \lambda \neq 0, \\ \log(Y_i) : & \lambda = 0. \end{cases}$$

BCT 的目的是使数据尽可能接近正态分布，从而满足线性模型通常采用的假设。BCT 无法保证任何概率分布都能转化为正态分布。不过，即使无法精确映射到正态分布，BCT 仍可能提供有用的近似[2]。考虑一个零均值高斯随机场 $U = [U(s_1)$.....，$U(s_n)]^\intercal$，在位置 s_i 处获得的 Box-Cox 变换后的响应Y_i'，$i = 1$，.....，n，其解释变量为 $X(s_i)$，建模如下

$$Y_i'|U(s_i) \overset{ind}{\sim} \mathrm{N}(\mu(s_i), \tau^2),$$
$$\mu(s_i) = X(s_i)\beta + U(s_i).$$

给定 $U(s_i)$，Y_i' 相互独立地服从观测方差为 τ^2 的正态分布。$\beta = (\beta_1$，.....，$\beta_p)^\intercal$ 是回归参数向量。位置 $s + h$ 与 s 之间的协方差形式如下，其中 $\rho(\cdot)$ 是相关函数，而 σ^2 是空间方差或残差变异中的变异。

$$\mathrm{COV}\left[U(s+h), U(s)\right] = \sigma^2 \rho(h; \phi, \kappa)$$

令 $Y' = (Y'_1, \ldots\ldots, Y'_n)^\top$ 为变换后观测值的向量，$X = [X(s_1)\ldots\ldots, X(s_n)]^\top$ 为一个 $n \times p$ 设计矩阵，则我们将模型改写为紧凑形式

$$Y' \sim \text{MVN}\left(X\beta, \sigma^2 R(\phi, \kappa) + \tau^2 I\right),$$
$$R_{ij} = \rho(s_i - s_j; \phi, \kappa),$$

这里 $R(\cdot)$ 是相关矩阵。记 $V = R(\phi, \kappa) + \nu^2 I$，且 $\nu^2 = \tau^2/\sigma^2$，则

$$Y' \sim \text{MVN}\left(X\beta, \sigma^2 V\right). \tag{1}$$

2.1 Matérn相关与几何各向异性

Stein [3] 强烈建议使用 Matérn 空间相关类，因为该类包含一个控制过程可微性的参数。我们采用以下 Matérn 参数化形式，其中包含一个用于模拟方向效应的旋转参数。

$$\rho(x; \phi, \kappa) = \frac{2^{\kappa-1}}{\Gamma(\kappa)} \left[\sqrt{8\kappa} d(x)\right]^\kappa K_\kappa \left[\sqrt{8\kappa} d(x)\right] \tag{2}$$

$$d(x) = \left\| \begin{pmatrix} 1/\phi_X & 0 \\ 0 & 1/\phi_Y \end{pmatrix} \cdot \begin{pmatrix} \cos\phi_A & -\sin\phi_A \\ \sin\phi_A & \cos\phi_A \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right\| \tag{3}$$

这里 $K_\kappa(\cdot)$ 表示阶数为 κ 的第二类修正贝塞尔函数，κ 是决定 $U(x)$ 可微性的形状参数。当 $\kappa = 0.5$ 时，Matérn 相关函数与指数相关 $\exp(-d)$ 一致。当 $\kappa \to \infty$ 时，它趋近于高斯相关 $\exp(-2d^2)$。标准各向同性相关具有一个（标量）范围参数 ϕ，该参数控制两个位置之间的相关性随距离增加而衰减的速度。各向异性模型包含三个用于确定点间距离的参数。旋转参数 ϕ_A 表示相关性最强的方向，两个范围参数 ϕ_X 和 ϕ_Y 则控制相关性在最强和最弱方向上的衰减速度。当 $\phi_X = \phi_Y$ 时，模型为各向同性模型，此时角度 ϕ_A 不起作用。

2.2 基于剖面似然的推断

设 $y' = (y'_1, \ldots\ldots, y'_n)^\top$ 为观测值 $y = (y_1, \ldots\ldots, y_n)^\top$ 的 Box-Cox 变换。此外，将 $\omega = (\phi_X, \kappa, \nu^2, \phi_Y, \phi_A)$ 记为协方差参数向量。给定 y 时，LGM 参数的对数似然函数为

$$\begin{aligned} -2\ell(\beta, \sigma^2, \lambda, \omega; y) =& (y' - X\beta)^\top (\sigma^2 V)^{-1} (y' - X\beta) + \log|\sigma^2 V| \\ & - 2 * (\lambda - 1) \sum_{i=1}^n \log y_i + n \log(2\pi), \end{aligned} \tag{4}$$

(4)式右边的第三项是 Box-Cox 变换的雅可比行列式。

给定 ω 和 λ，对数似然函数 (4) 在以下位置取得最大值。

$$\hat{\beta}_{\omega,\lambda} = (X^\top V^{-1} X)^{-1} X^\top V^{-1} y', \quad \text{and} \tag{5}$$

$$\hat{\sigma}^2_{\omega,\lambda} = \frac{1}{n} (y' - X\hat{\beta}_{\omega,\lambda})^\top V^{-1} (y' - X\hat{\beta}_{\omega,\lambda}). \tag{6}$$

将 (5) 和 (6) 代入 (4)，即可得到 ω 和 λ的剖面对数似然（PLL）

$$\begin{aligned} -2\ell_p(\omega, \lambda; y) =& n \log \frac{(y' - X\hat{\beta}_{\omega,\lambda})^\top V^{-1} (y' - X\hat{\beta}_{\omega,\lambda})}{n} + \log|V| \\ & - 2(\lambda - 1) \sum_{i=1}^n \log y_i + n \log(2\pi) + n. \end{aligned} \tag{7}$$

模型参数的最大似然估计（MLE）没有封闭形式的公式。对 (7) 进行数值优化，可得到最大似然估计 $\hat{\omega}$ 和 λ；再将它们代回 (5) 和 (6)，即可得到最大似然估计 β 和 $\hat{\sigma}$。附录中简要介绍了使用限制最大剖面似然对模型参数进行推断的方法。

2.3 基于信息和基于似然的置信区间

我们用 θ 表示任意参数向量 $\theta = (\theta_1 \dots \theta_q)$ 的最大似然估计。单个参数 θ_i 的渐近 100$(1-\alpha)$% 置信区间可以通过两种方式构造。

1. Wald 方法，或基于信息的置信区间

$$\{\theta_i : |\theta_i - \hat{\theta}_i| \leq z_{1-\alpha/2}[I(\hat{\theta})^{-1}]_{i,i}^{1/2}\} \tag{8}$$

此处 $I(\theta)$ 是在最大似然估计处计算的观测费舍尔信息，而 $z_{1-\alpha/2}$ 是标准正态分布的 $1-\alpha/2$ 百分位数。

2. 基于似然的置信区间

$$\{\theta_i : \ell[\theta_i, \hat{\theta}_{-i}(\theta_i)] - \ell(\hat{\theta}) \geq c_{1,\alpha}/2\}, \qquad \hat{\theta}_{-i}(\theta_i) = \text{argmin}_{\theta_{-i}} \ell(\theta_i, \theta_{-i}) \tag{9}$$

此处 θ_{-i} 表示去掉第 i 个元素后的向量 θ，而 (θ_i, θ_{-i}) 称为约束最大似然估计。$c_{1,\alpha}$ 是自由度为 $\mathbb{P}(\chi_1^2 \leq c_{1,\alpha}) = \alpha$ 的卡方分布的分位数。

请注意，这两个结果都是关于 Y' 维数 n 的渐近结果。P. J. Diggle 和 Giorgi [4] 以及 Haskard [5] 等人主张在空间模型中使用基于似然的置信区间，不过两者都指出，如果 θ 的真实值位于参数空间的边界，或者 θ 中一个或多个元素定义了 Y' 的尺度参数，那么这种近似就不成立。Baey、Courn'ede 和 Kuhn [6] 以及 Visscher [7] 表明，对于非空间模型，当任意一组随机效应的方差等于零时，检验统计量的渐近分布是由具有不同自由度的卡方分布混合而成。当 θ 的真实值远离边界时，地统计模型中这种近似的准确性尚未得到研究。

3 方法论

我们方法背后的主要思想是，通过并行多次评估似然来构建基于似然的置信区间。我们首先构建似然的二次近似，然后选取大量具有代表性的 ω 值来评估似然。我们根据这些点构建插值函数，并用它来计算剖面似然。该算法中耗时的部分是数千次评估似然，而我们的 GPU 实现能够同时对数百个参数值进行计算；每次单独的似然评估还会进一步并行化，并使用数千个 GPU 工作项。

LGM 有 $7+p$ 个模型参数：五个相关参数 $\omega = (\phi_X, \kappa, \nu^2, \phi_Y, \phi_A)$、方差参数$\sigma^2$、Box-Cox 参数 λ，以及 p 个 β 元素。所有这些参数的剖面似然均按以下步骤计算。

步骤1，通过对剖面似然进行数值最大化，获得 MLE 的 $\hat{\omega}$ 和 λ。

步骤2，使用对数似然的二次近似，创建一组代表性参数$\omega_1, \dots, \omega_K$和 $\lambda_1, \dots, \lambda_M$。

步骤3，针对参数 ω_k 和 λ_k的 $K \cdot M$ 种组合，计算并存储各种矩阵行列式和叉积。请注意，方差矩阵只有 K 种不同的。

步骤4，计算参数的剖面似然 $\ell_p(\omega_k, \lambda_k)$，并利用最高剖面似然值获取相关参数和 λ 的单变量剖面似然。

步骤5，通过取最大值，获得 σ 的多个取值对应的 PLL。$\ell_p(\sigma, \omega_k, \lambda_k, \hat{\beta}_{\sigma,\omega_k,\lambda_k}, y)$ 在 k 和 m上。

步骤6，通过在 k 和 m 上最大化 $\ell_p(\beta, \omega_k, \lambda_k, \hat{\sigma}_{\beta,\omega_k,\lambda_k}; y)$，获得 β 的 PLL。第（3.4）节展示了如何获取单个 β_p 或系数对的单变量 PLL。

3.1 步骤1：协方差参数的最大似然估计

步骤1通过对公式 (7) 中的剖面似然 $\ell_p(\omega, \lambda;\ y)$ 进行数值最大化来完成。这需要按顺序计算似然，并使用 geostatsp 软件包 [8] 中实现的 L-BFGS-B 算法在中央处理器上完成。由于似然相对于形状参数 κ 通常非常平坦，因此还会计算两个分别固定 $\kappa = 0.5$ 和 $\kappa = 10$ 的约束最大值。

3.2 步骤2：一组代表性参数

干扰参数的粗略配置会导致剖面曲线 [9], 出现波动；需要大量经过适当且充分采样的参数值，才能使剖面曲线变得平滑。与 ***R-INLA*** 软件包 [10], 类似，我们使用（7）中的剖面似然 $\ell_p(\omega, \lambda;\ y)$ 的二次近似创建一组“配置”点 $\omega_k;\ k = 1 \dots K$。这些点均匀分布在由 $\ell_p(\omega, \lambda; y) = c$指定的曲面周围，其中 c取似然最大值处的不同 χ_2 分位数。如果该近似是精确的，则在 ω_k, λ_k 处评估的似然将等于模型参数似然的 χ^2 分位数。我们寻找一种重新参数化，使精确似然与近似似然尽可能相似。

步骤2包括三个部分：对似然函数进行重新参数化；使用海森矩阵计算二次近似；以及确定代表点。

重新参数化

对模型参数进行重新参数化，是为了使剖面似然尽可能接近二次形式，从而提高近似的准确性。Matérn形状参数 κ 的变换取决于 MLE $\hat{\kappa}$ 是否较大。所使用的重新参数化为

$$\tilde{\kappa} = \begin{cases} \log \kappa, & \hat{\kappa} < 4 \\ 1/\sqrt{\kappa}, & \hat{\kappa} \geq 4. \end{cases}$$

当 $\hat{\kappa}$ 较大时，似然往往较为平坦，而在实际应用中，逆平方根变换比对数变换效果更好。

控制空间相关性衰减速率的三个参数 ϕ_X、ϕ_Y 和 ϕ_A 按照 [11] 提出的方式进行重新参数化，其中

$$\begin{aligned} \gamma_1 &= \log \phi_X + \log \phi_Y, \quad \text{(sumLogRange)} \\ \gamma_2 &= \sqrt{\phi_X/\phi_Y - 1}\cos(2\phi_A), \quad \text{(aniso1)} \\ \gamma_3 &= \sqrt{\phi_X/\phi_Y - 1}\sin(2\phi_A). \quad \text{(aniso2)} \end{aligned}$$

注意，exp($\gamma_1/2$)（组合范围）是 ϕ_X 和 ϕ_Y 的几何平均数。(γ_2, γ_3) 的新参数化对范围之比 ϕ_X/ϕ_Y 和角度 ϕ_A 应用了极坐标变换，[11] 之所以选择这种变换，是因为各向同性模型对应于单个点 $\gamma_2 = \gamma_3 = 0$，从而避免了数据具有各向同性时似然在 ϕ_A 维度上趋于平坦的问题。通过第6节和第7节所介绍的结果，可以更容易地理解采用这种重新参数化的动机。

我们将参数 $\omega =$($\phi_X, \kappa, \nu^2, \phi_Y, \phi_A$) 称为“自然参数”，将重新参数化后的参数 $\omega' =$($\gamma_1, \tilde{\kappa}, \nu, \gamma_2, \gamma_3$) 称为“内部参数”。

海森矩阵

共有六个非线性参数，即 Box-Cox 参数 λ 和 ω' 中的五个协方差参数。为了构造对数似然的二次近似，我们隐含地假设对数^ 似然关于 λ 和 ω' 的二阶导数为零，并重点研究 $\ell_p(\omega', \lambda; y)$ 的 5 维海森矩阵，其中改变 ω'，但将 λ 固定在最大似然估计处（作为 ω' 的函数）。似然关于 λ 的二阶导数用于近似似然关于 Box-Cox 参数的函数。此外，还计算了两个 4 维海森矩阵（分别将 κ 固定为 0.5 和 10）。

所有海森矩阵均使用数值微分进行计算。对于 5 维海森矩阵，我们需要计算 $51 \cdot 3$ ^ 个似然值（^$\omega' \pm \delta$ 和 $\lambda \pm \delta$），而对于 4 维海森矩阵，则需要计算 $33 \cdot 3$ 个似然值。似然值通过图形处理器并行计算。在空间统计学中，似然函数较为平坦的情况相当常见，

[12] 从解析上表明，尺度参数、形状参数和方差参数通常难以分别准确识别。海森矩阵可能不是负定矩阵，或者其特征值非常接近于零。我们对（负）海森矩阵 $-H$（$\hat{\omega}'$）进行特征分解（稍后在获取代表点时需要这些结果）：

$$-H(\hat{\omega}') = EDE^{\mathsf{T}},$$

其中， E 是特征向量矩阵， D 是特征值对角矩阵，并将任何负特征值替换为其绝对值。如果特征值的范围较大，即最大特征值超过 100，而最小特征值接近 0，那么我们将最小特征值替换为 0.1，以避免椭球的某一轴主导点的选择。在多个真实和模拟数据集上的实践表明，这种处理较小和负特征值的方法效果良好。

获取代表点

性质良好的对数似然曲面在 $\hat{\omega}'$ 附近可以用由泰勒展开导出的二次函数很好地近似

$$\ell(\omega') = \ell(\hat{\omega}') + \frac{1}{2}(\omega' - \hat{\omega}')^{\mathsf{T}}(-H(\hat{\omega}'))^{-1}(\omega' - \hat{\omega}'). \tag{10}$$

满足以下条件的实心椭球 ω' 值

$$(\omega' - \hat{\omega}')^{\mathsf{T}}(-H(\hat{\omega}'))^{-1}(\omega' - \hat{\omega}') \le \chi^2_p(\alpha) \tag{11}$$

定义一个概率为 $1-\alpha$ 的置信区域，其中 $\chi^2_p(\alpha)$ 是具有 p 个自由度 [13] 的卡方分布的上侧第 100α 百分位数。为了在超椭球面上获得均匀分布的点，我们首先使用一个随机优化器，寻找以原点为中心、在五维单位超球面上均匀分布的点 $x = (x_1, \ldots x_n)$；该优化器通过最大化最近点对之间的距离 $\max_X \min_{i,j} \|x_i - x_j\|$ 来实现。然后根据公式，将点 x 变换为目标超椭球面上的点 x'

$$x' = E^{-2}D^{1/2}x + \hat{\omega}'.$$

从每个密度等高线层级中选取 726 个代表点 ($n = 726$)。通常将来自 9 到 11 个等高线层级的参数集组合起来，形成最终的相关参数配置矩阵。对于 λ，我们在其于 λ处的一维二次近似的 0.01 和 0.99 分位数之间，设置一个通常包含 30 个等间距值的向量。

由于形状参数在经验上难以估计，当其对数似然明显偏离二次函数时，二次近似可能无法很好地适用于该形状参数。因此，我们额外拟合了若干模型，将 Matérn形状参数固定在 0.5 到 20 之间的若干取值上。这些额外模型的海森矩阵为 4×4，其中每个密度等高线层级都有 $n = 120$ 个代表点。形状参数的似然通常非常平坦，因此这一过程让我们能够探索 κ 取值范围内的似然。如果真实块金效应接近 0，那么根据二次近似抽样时可能会得到块金效应 ν^2 的负值。在这种情况下，我们将其中一半负值设为 0，另一半替换为从区间 (0,2) 中随机抽取的均匀分布数值。真实数据示例中抽样得到的块金效应值及其对应的剖面对数似然见图7c，其中可以看到 0 处聚集了一小堆点。

3.3 步骤3：对数似然计算

计算似然值时，计算开销最大的步骤是对方差矩阵 V 进行 Cholesky 分解，并执行三角回代计算。这些步骤在图形处理器上完成，表 1 中列出的交叉乘积和矩阵行列式则保存后返回中央处理器。随后使用这些似然值组成部分计算不同形式的似然值（即剖面化掉 β、 σ，或同时剖面化掉两者）。

我们通过以下步骤计算参数组 ω'_1 到 ω'_K 对应的表 1 中各项 .

步骤1：使用 *maternBatch()* 函数计算 Matérn 相关矩阵 $V = (V_1, \ldots, V_K)$。

步骤2：使用 *cholBatch()* 对 $V_k = L_k D_k L_k^{\mathsf{T}}$ 进行 Cholesky 分解，其中 L_k 是下三角单位矩阵， D_k 是对角矩阵。随后，该函数取对数 $|V| = 2$、对数 $|L| +$、对数 $|D| =$、对数 $|D|$。*cholBatch()* 返回 L_k、 D_k 以及 log $|V_k| =$ sum(log D_k)。

步骤3：使用 , X 通过 *backsolveBatch()* 求解 L_k^{-1} (y。

组件	R 中的名称	
$\log\lvert V\rvert$	方差行列式	
$\log\lvert X^\intercal V^{-1}X\rvert$	REML 行列式	
$y^\intercal V^{-1}y$	YX 平方和	左上角对角元素
$X^\intercal V^{-1}X$	YX 平方和	右下角对角元素
$X^\intercal V^{-1}y$	YX 平方和	左下块
$(X\hat{\beta})^\intercal V^{-1}(X\ \hat{\beta})$	Betahat 平方和	
$(y-X\hat{\beta})^\intercal V^{-1}(y-X\ \hat{\beta})$	ssqResidual	

表 1：针对每组参数值计算并保留的汇总统计量

步骤4：使用 the 使用上面的内容，通过 *crossprodBatch()* 计算 $(y,X)^\intercal V_k^{-1}(y,X)$，结果保存在 YX 平方和中。

步骤5：使用 *cholBatch()* 对 $X_k^\intercal V_k^{-1}X_k \ = \ Q_kP_kQ_k^\intercal$ 进行 Cholesky 分解，并计算 $\log\lvert X_k^\intercal V_k^{-1}X_k\rvert$。

步骤6：使用 *backsolveBatch()*求解方程 $c=Q_k^{-1}X^\intercal V_k^{-1}y$。

步骤7：通过 *crossprodBatch()*计算 ssqBetahat $=(X\hat{}\beta)^\intercal V^{-1}(X\hat{}\beta)\ =c^\intercal P_k^{-1}c$。

步骤 8：计算 ssqResidual $=(y-X\hat{}\beta)^\intercal V^{-1}(y-X\hat{}\beta)\ =$ ssqYX $-$ ssqBetahat。

GPU 内存限制意味着无法同时处理所有 K 组参数；限制因素是方差矩阵 V_k所占用的内存。此处给出的结果每次处理数百组参数；可处理的最大数量取决于数据集的大小和 GPU 内存容量。

3.4 β_p的剖面似然

此处我们考虑计算第 p个协变量对应的单个回归系数 β_p 的剖面似然。剖面似然为

$$\begin{aligned}\ell_p(\beta_p;y) &= \max_{\beta_{-p},\sigma,\omega,\lambda}\{\ell(\beta_p,\beta_{-p},\sigma,\omega,\lambda;y)\}\\ &\approx \max_{k,m}\{\ell(\beta_p,\hat{\beta}_{-p}(\beta_p,\omega_k,\lambda_\ell),\hat{\sigma}(\beta_p,\omega_k,\lambda_\ell),\omega_k,\lambda_m)\}\end{aligned}\tag{12}$$

$\hat{}$ 其中 β_{-p} 是向量 β 去掉第 p个元素后的结果。函数 $\hat{}\sigma(\cdot)$ 和 $\beta_{-p}(\cdot)$ 表示在给定 β_p、ω 和 λ的值时，β_{-p} 和 σ的最大似然估计；这些估计有闭式解。

当 β_p 固定时，项 $X_p\beta_p$ 成为固定偏移量，模型中涉及协变量的项变为 $\hat{}X_{-p}{}^\intercal\beta_{-p}$，其中 X_{-p}是矩阵 X 去掉第 p列后的结果。计算 $\beta_{-p}(\beta_p,\ \omega,\ \lambda)$ 涉及以下矩阵交叉乘积:

$$\begin{aligned}(y-X_p\beta_p)^\intercal V^{-1}(y-X_p\beta_p) &= y^\intercal V^{-1}y-2\beta_p(X^\intercal V^{-1}y)_{[p,]}+\beta_p^2(X^\intercal V^{-1}X)_{[p,p]},\\ (X_{-p}\hat{\beta_{-p}})^\intercal V^{-1}(X_{-p}\hat{\beta_{-p}}) &= (X^\intercal V^{-1}y)^\intercal_{[-p,]}(X^\intercal V^{-1}X)^{-1}_{[-p,-p]}(X^\intercal V^{-1}y)_{[-p,]}\\ &\quad -2\beta_p(X^\intercal V^{-1}X)^\intercal_{[-p,p]}(X^\intercal V^{-1}X)^{-1}_{[-p,-p]}(X^\intercal V^{-1}y)_{[-p,]}\\ &\quad +\beta_p^2(X^\intercal V^{-1}X)^\intercal_{[-p,p]}(X^\intercal V^{-1}X)^{-1}_{[-p,-p]}(X^\intercal V^{-1}X)_{[-p,p]}.\end{aligned}$$

下标记法 $[-p,-p]$ 表示从矩阵中排除第 p 行和列，而 $[-p,p]$ 表示排除第 p 行后的矩阵第 p 列。计算上述内容只需对表 1 中列出的 ssqYX 矩阵进行行列子集选取，无需额外使用图形处理器即可计算每个 β_p 的剖面似然。

3.5 σ的剖面似然

将 (5) 代入 (4)，即可轻松得到 (σ， ω， λ) 的 PLL：

$$-2\ell_p(\sigma,\omega,\lambda;y) - n\log(2\pi) = (y - X\hat{\beta})^{\mathsf{T}}(\sigma^2 V)^{-1}(y - X\hat{\beta}) + n\log\sigma^2 + \log|V| - 2(\lambda-1)\sum_{i=1}^{n}\log y_i, \tag{13}$$

ˆ其中， β 和 V 都依赖于 ω。(13)中的各项均可根据表 1 中的量计算，唯有最后一项（Box-Cox 雅可比行列式）除外；该项也由图形处理器返回。我们用以下方式近似 σ 的剖面似然：

$$\ell_p(\sigma) \approx \max_{k,m} \ell_p(\sigma,\omega_k,\lambda_m;y),$$

并针对一系列 σ取值计算 $\ell_p(\sigma)$。

3.6 相关参数的剖面似然

我们可以根据所有已计算似然值 $\ell_p(\omega_k, \lambda_m; y)$ 的点云，求得任意相关参数的剖面似然；相关参数既可以采用重新参数化形式 $\omega' = (\gamma_1, \tilde{\kappa}, \nu, \gamma_2, \gamma_3)$，也可以采用原始形式 $\omega = (\phi_X, \kappa, \nu^2, \phi_Y, \phi_A)$，示例如图7所示。以计算形状参数 κ 的剖面似然（PLL）为例，我们根据配置点 ω_k 创建二维散点图，其中 κ_k 位于 x 轴，$\max_m \ell_p(\omega_k, \lambda_m; y)$ 位于 y 轴。通过计算这些点的凸包，得到其中的“最佳”似然值集合，如图7b中的12个黑色圆点所示。图7b中的绿色线条是对这12个点进行线性插值得到的，即估计的 PLL 曲线。Box-Cox 参数 λ 的 PLL 则是对 λ_m 和 $\max_k \ell_p(\omega_k, \lambda_m; y)$ 进行插值得到的。

类似地，对于 ω' 中相关参数 (ω'_i, ω'_j) 任意两个参数的二维剖面似然，我们首先在三维空间中找到能够包围点集 $[\omega'_i, \omega'_j, \max_\lambda \{\ell_p(\omega, \lambda; y)\}]$ 的最小凸包，移除凸包底面上的点，然后对剖面似然曲面进行线性插值。我们在由 (ω'_i, ω'_j) 张成的空间上创建规则网格，并使用插值法预测这些网格上的值。

为了绘制“自然空间”(ω_i, ω_j) 中相关参数的二维剖面似然，我们将“自然空间”中创建的规则网格转换为“内部空间”中的网格，并仍使用在“内部空间”中构建的插值法来预测这些网格点。凸包和插值方法在原始参数空间中的效果较差，因为对数似然在该空间中的钟形特征不如重参数化空间中明显。

4 计算方面的考量

要充分利用图形处理器的并行计算能力，需要对图形处理器内存和工作项（CPU 核心在图形处理器上的对应单位）进行一些手动配置。工作项被划分为工作组，这些工作组共享本地内存（其速度快于全局内存）。我们将工作项视为排列在二维网格上（未使用第三维），并按照表 2 所示，将任务分配给各个工作项。

工作项排列在一个 Nglobal0 × Nglobal1 的网格上，每个工作组包含 Nlocal0 × Nlocal1 个工作项。假设 Nglobal0=6、Nglobal1=4、Nlocal0=3 和 Nlocal1=2，这意味着在二维（维度 0 和维度 1）计算域中（见图 1）有 2×2 个工作组，每个工作组由 3×2 个工作项组成。假设我们共有 500 个参数配置，将其分为 5 组，每组 100 个。在 C++ 函数 *likfitGpu_Backend()* 的每次循环中，图形处理器计算 100 个对数似然，并将相应的 100 个充分统计量保存到图形处理器的全局内存中。

表 2 展示了用于执行矩阵运算的四种主要 GPU 内核。在本例的每个循环中，*maternBatch* 内核的维度 0 上的工作项会遍历方差矩阵中的元素，维度 1 上的工作项则会遍历参数集（每个参数集用于不同的方差矩阵）。在上述情形中，每个工作组会使用 100 个参数集中的一半，并将 50 个参数集复制到本地内存中，因为维度 1 上有两个工作组。每个工作组会将 $\text{NparamPerIter}/2 = 50$ 个参数和 3 个坐标（ Nlocal 的第一个元素）复制到本地内存中。每次迭代时，一个工作组会将三个矩阵元素对应的坐标对复制到本地内存中（因为 Nlocal0=3），并将这些坐标用于两个矩阵（因为 Nlocal1 = 2）。

表 2：主要 GPU 内核列表，展示了它们所执行的任务，以及任务在工作项和工作组之间的划分方式。

内核	公式	group0	组1	局部0	局部1
maternBatch	V	共享坐标 共享参数		矩阵元素	矩阵
cholBatch	$V = LDL^{\intercal}$	矩阵	$unused$	L 的行	L 的列
backsolveBatch	$AC = B$	矩阵	B 和 C 的列	A 和 C 的行	A 的列 , C 的行
crossprodBatch	$C = A^{\intercal}DA$	C 的列	矩阵	D 的元素	C 的行

Cholesky 分解由 *cholBatch* 执行，且只使用维度 1 上的单个工作组（需要 Nglobal1 = Nlocal1）。第一个工作组对方差矩阵 $V_0, V_2, V_4, \ldots$ 执行 Cholesky 分解，第二个工作组则对 $V_1, V_3, V_5, \ldots$ 执行 Cholesky 分解。在每个工作组内，第一行的工作项负责计算 L第一列中的第 $1, 4, 7, \ldots$ 行；同时，第二行的工作项负责计算 L第一列中的第 $2, 5, 8, \ldots$ 行，第三行的工作项则计算 L第一列中的第 $3, 6, 9, \ldots$ 行中的单元格。

在*backsolveBatch* 内核中，同一行的工作组会计算相同的 C_i。在此示例中，上方两个工作组依次计算 C_0, $C_2, C_4, \ldots$，下方两个工作组则依次计算其余的 $C_1, C_3, C_5, \ldots$。每个工作组负责 B 的 ceil(ncol(B)/2) 列，换言之，此处灰色阴影工作组负责每个 C_i 的第 $0, 2, 4, \ldots$ 列，白色工作组则负责每个 C_i 的第 $1, 3, 5, \ldots$ 列。在每个工作组内，第 0 行的工作项会一起计算每个 C_i 的第 $0, 3, 6, \ldots$行；同样地，第 1 行的工作项会同时一起计算每个C_i 的第 $1, 4, 7, \ldots$ 行，第 2 行的工作项则会一起计算每个 C_i 的第 $2, 5, 8, \ldots$ 行。

crossprodBatch 内核计算 $C = A^{\intercal}DA$ 或 $C = A^{\intercal}D^{-1}A$ 或 $C = A^{\intercal}A$。左侧两个带阴影的工作组负责计算 C_O, C_2， C_4，……，而白色工作组负责计算C_1， C_3， C_5，……。维度 0 中的工作组索引决定该工作组中的工作项所计算的 C 的列；在此，上方工作组计算每个 C_i 中的第 0、2、4、……列，下方工作组计算每个 C_i 中的第 1、3、5、……列。在每个工作组内，维度 1 中局部索引相同的工作项计算 C_i 的相同行，因此这里左侧三个工作项计算每个 C_i 的第 0、2、4、……行，右侧三个工作项计算每个 C_i 的第 1、3、5、……行。

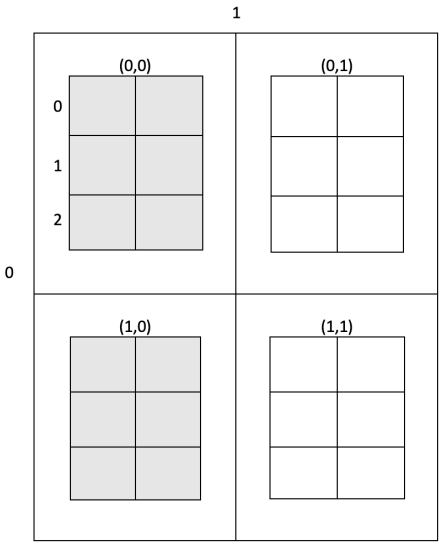


图 1： 用于说明的简单示例中二维计算域的表示。

函数 *likfitLgmGpu()* 中有一个参数 NparamPerIter，用于指定每次调用后端 GPU 函数时，每次迭代同时计算的参数集数量。批大小值（NparamPerIter）受可用 GPU 全局内存的限制。如果待估计的 LGM 有 n 个观测值、 p 个协变量，并且 Box-Cox 参数的向量配置有 m 个值，那么后端 C++ 函数中的每个循环都会刷新 GPU 矩阵 V（存储 Matérn 相关矩阵）；该矩阵有 n*NparamPerIter 行和 n 列；*crossprodBatch* 的结果 ssqYX 是一个具有（$m+p$）*NparamPerIter 行和 $m+p$ 列的 GPU 矩阵。如果要拟合的数据集很大，或者模型中有大量协变量且 λ 的配置包含大量值，则应将 NparamPerIter 设置得相对较小。Nglobal 和 Nlocal 的值最好为 2 的幂，并且不应超过 GPU 上工作项的最大数量；该数量可通过 gpuR::gpuInfo() 查找。

5 两项模拟研究

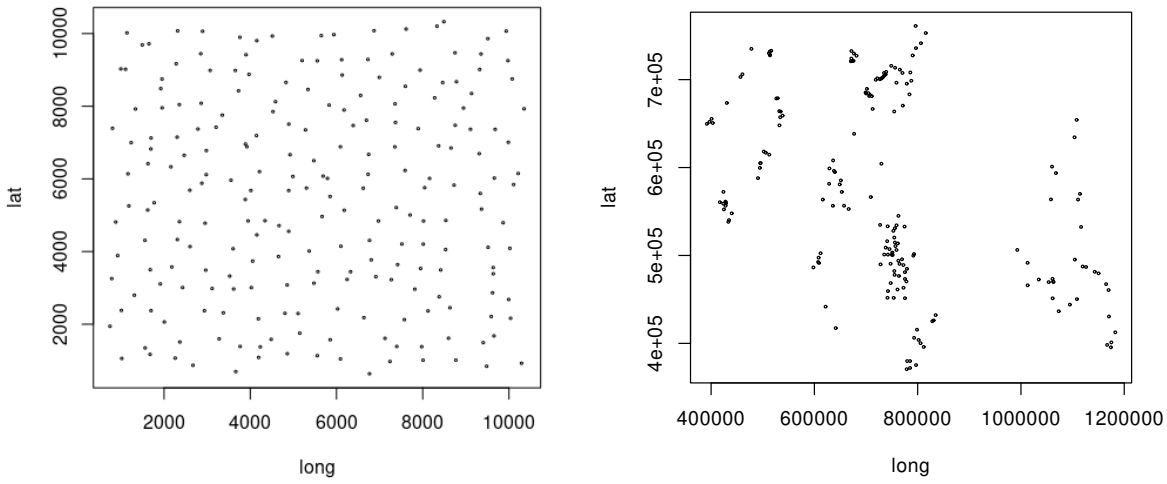
为评估我们的方法在现实情境下产生的置信区间，我们进行了 2 项模拟研究：研究A使用各向异性数据，研究 B 使用各向同性数据。在每项研究中，我们使用模拟坐标进行了 150 次模拟，并使用经典*Loaloa* 数据集中的坐标进行了另外 150 次模拟。

具体而言，在研究A中，我们从模型 $Y_i \sim N(5+1*X_{i1}+1*X_{i2}+U(s_i)\ \tau^2)$（其中 $\tau = 0.8$）生成 150 个实现值 y_1，.....， y_{150} 两次。在两个模拟 A(i) 和 A(ii) 中，具有方差 $\sigma^2 = 1$ 和 Matérn 相关函数（参数为 $\kappa = 2$、 $\phi_R = 2$、 $\phi_A = 0.2$ ）的平稳高斯随机场 $U(s)$ 在不同坐标上生成，两个解释变量也相应地以不同方式定义。未对 Y 进行 Box-Cox 变换，因此参数 $\lambda = 1$。

对于研究 A(i)， $U(x)$是在 $9km \times 9km$ 正方形上的 224 个点位置（Ngrid = 15）生成的（如图 2a 所示）。图 2a 中的点位置是随机生成的，我们移除了一个与另一个点非常接近的点。对于这些坐标，范围参数 ϕ_X 设为 1,000。协变量 X_{i1} 和 X_{i2} 根据坐标值定义为： $X_{i1} = coordsX_i/10000, X_{i2} = (coordsX_i/10000)^2,$ ，其中 $coordsX_i$ 表示位置 s_i的 x 坐标。

对于研究 A(ii)， $U(s)$是在*Loaloa* 坐标上生成的，如图 2b 所示。*Loaloa* 是一个真实数据集（可在***geostatsp*** 软件包中获取），包含 190 个位置（来自 190 次村庄调查），其中许多位置聚集在一起。对于 A(ii)， ϕ_X 设为 50,000。相应地，两个解释变量定义为： $X_{i1} = \log(elevationLoa_i); X_{i2} = eviLoa_i * 10^{-7}$，其中 $elevationLoa_i$ 表示位置 s_i处海拔的对数。

类似地，对于研究 B，我们从模型 $Y_i \sim N(2 + 1 * X_{i1} + 1 * X_{i2} + U(s_i)\ \tau^2)$ 中生成总计 300 个实现值 $y_1, \ldots, y_{300}$，其中 $\tau = 0.8$ 和 $U(s)$ 定义在与研究 A 中相同的两种类型的坐标上。 $U(s)$ 的参数设置与模拟 A 中相同，但 $U(s)$ 为各向同性。同样，研究 B(i) 的实现值是在模拟坐标上生成的，而研究 B(ii) 则使用*Loaloa* 坐标。



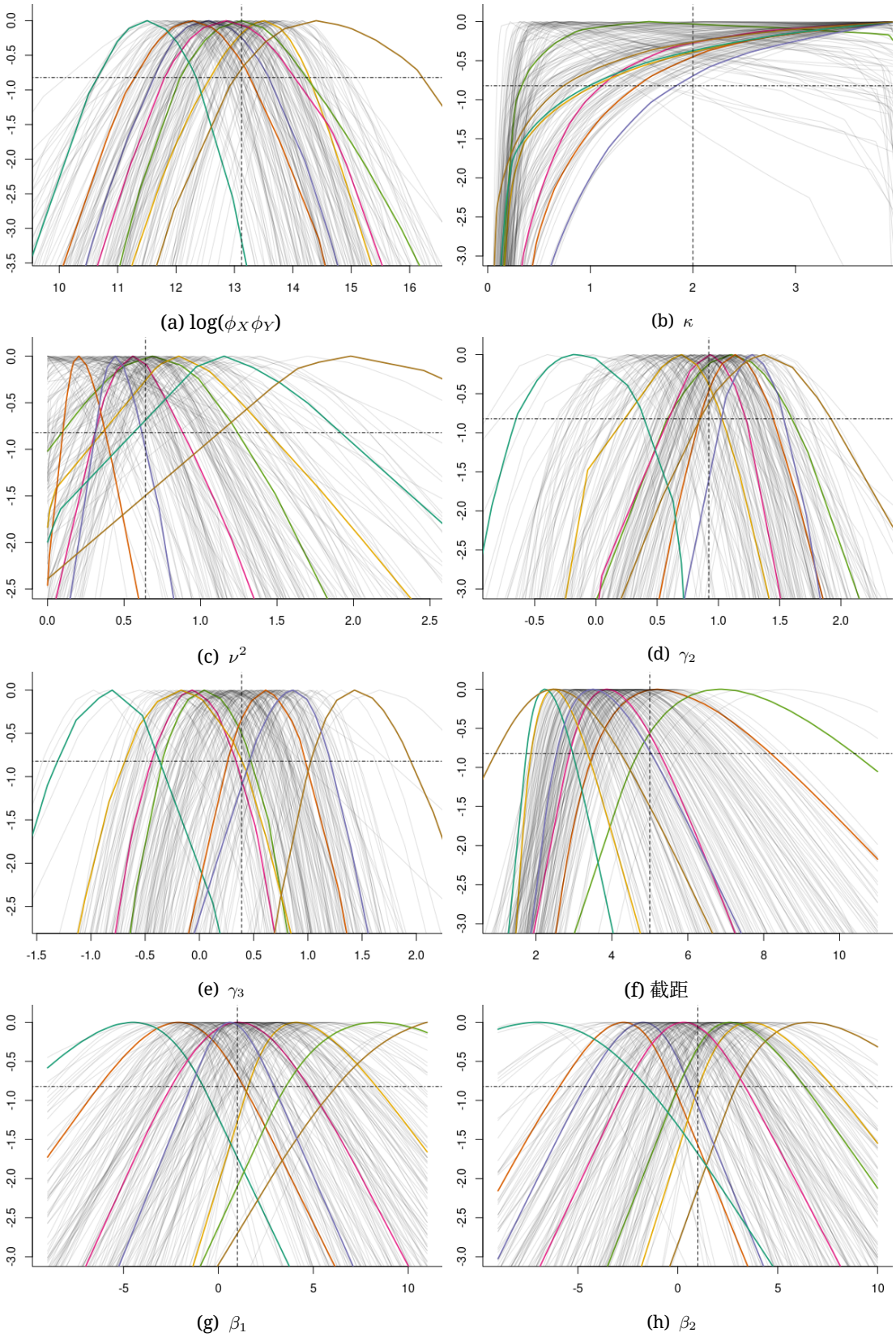
(a) 模拟数据 (b) Loaloa

图 2： $U(s)$ 的模拟坐标。

我们计算了每个参数的 80% 置信区间覆盖率，并将这些参数的置信区间覆盖率与使用 ***geostatsp*** 计算的结果进行了比较，见表3。研究 A(i) 和 B(i) 中，使用模拟坐标计算的 8 个参数各自对应的 150 条剖面似然曲线分别如图3和图4所示。各向异性参数的 80% 置信区域二维图如图5所示。

结果表明，***gpuLik*** （或基于剖面似然的置信区间）的覆盖率更高，也更接近期望值，优于***geostatsp***（或基于信息量的置信区间）。对于回归参数 β's 以及基于真实坐标*Loaloa*进行的模拟，这种优势更加明显。

对于使用各向同性数据的模拟，我们还统计了二维 80% 基于剖面似然的置信区域 (γ_2， γ_3) 的覆盖率，结果见图5。在 150 次模拟中，有 103 个置信区域覆盖了参数 (γ_2， γ_3) 的真实值（比率为 $= 0.687$）。



(a) $\log(\phi_X \phi_Y)$ (b) κ

(c) ν^2 (d) γ_2

(e) γ_3 (f) 截距

(g) β_1 (h) β_2

图3：模拟研究A（i）中模型参数的剖面对数似然图（各向异性数据和模拟坐标）。

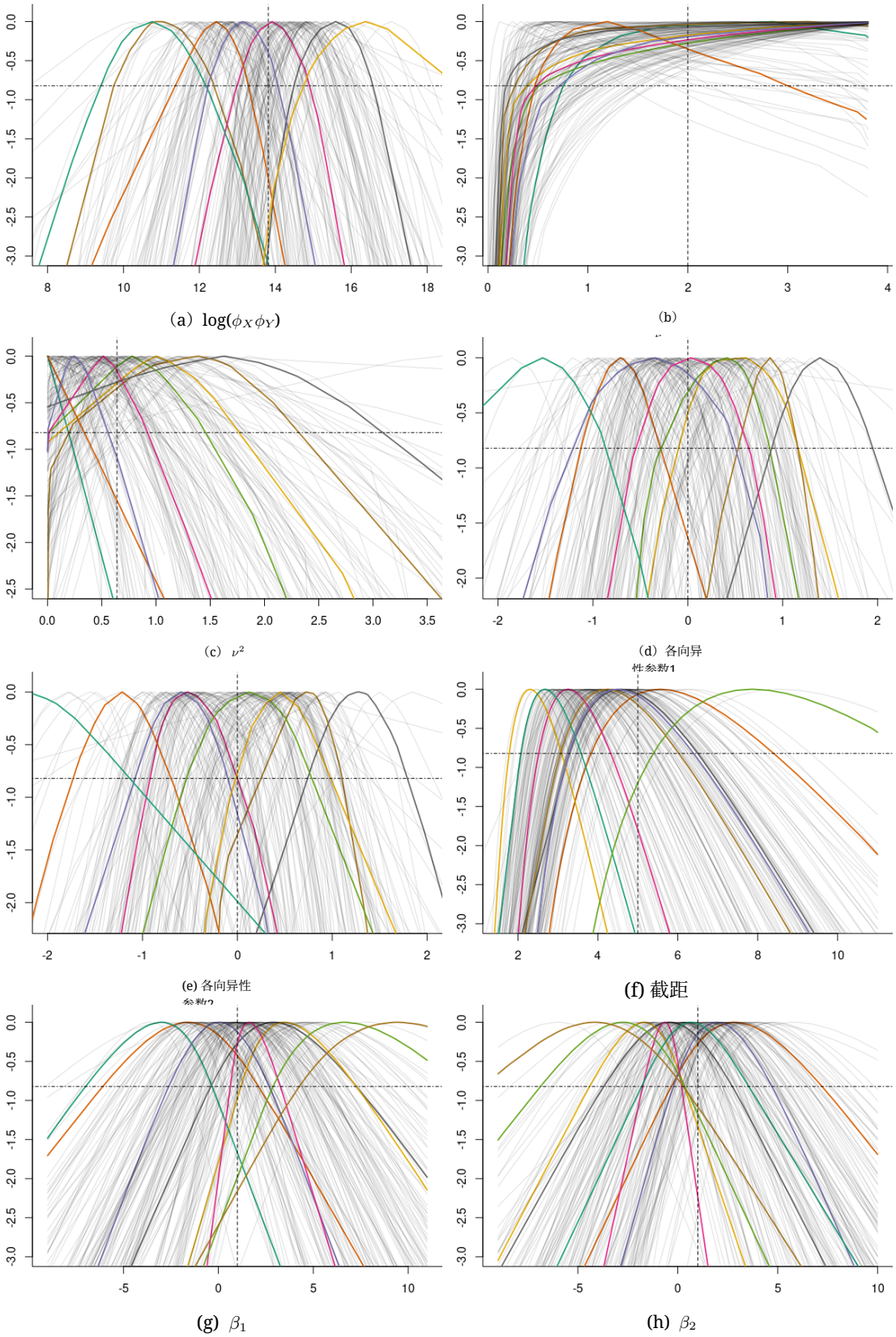
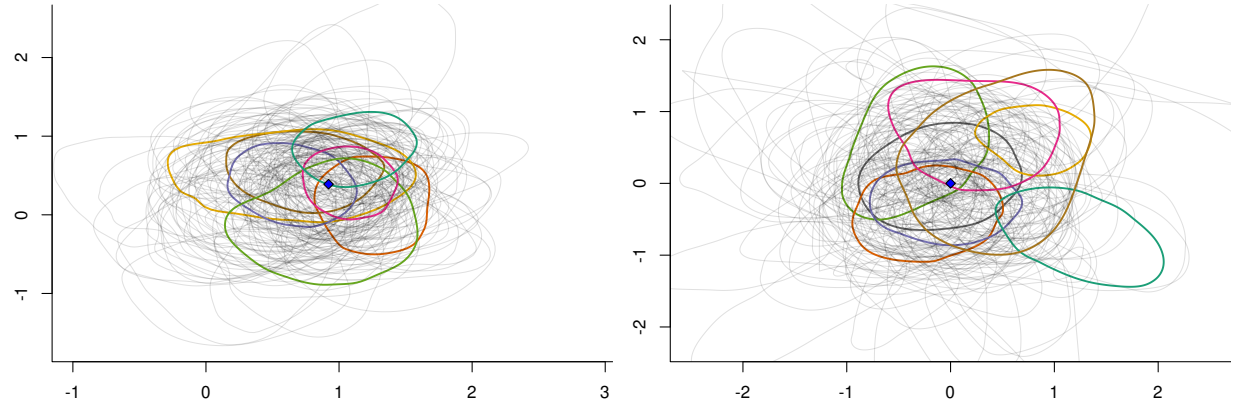


图 4：模拟研究 B (i) 中模型参数的剖面对数似然（各向同性数据和模拟坐标）。



(a) γ_2 与 γ_3的关系 (b) γ_2 与 γ_3的关系

图5：分别展示研究A中150次模拟得到的 γ_2 （横轴）和 γ_3（纵轴）的二维80%剖面对数似然置信域： (i)（左图）使用各向异性数据和规则坐标，B (i)（右图）使用各向同性数据和规则坐标。

表3：比较模拟研究A和B中模型参数的150个80%置信区间的覆盖率，分别采用***gpuLik*** 和 ***geostatsp*** 提供的方法。

	各向异性				各向同性			
	模拟坐标		Loaloa坐标		模拟坐标		Loaloa坐标	
参数	基于似然的	Wald	基于似然的	Wald	基于似然的	Wald	基于似然的	Wald
截距	0.600	0.340	0.740	0.513	0.666	0.340	0.686	0.533
β_1	0.800	0.786	0.813	0.173	0.800	0.786	0.813	0.180
β_2	0.800	0.766	0.820	0.180	0.793	0.746	0.813	0.266
ϕ_X 或 $\sqrt{\phi_X\phi_Y}$	0.760	0.460	0.760	0.706	0.420	0.193	0.513	0.400
κ	0.967	0.800	0.926	0.766	0.980	0.393	0.933	0.506
ν或 τ	0.860	0.780	0.826	0.713	0.753	0.320	0.826	0.466
γ_2 或 ϕ_R	0.780	0.713	0.660	0.580	0.653	0.186	0.660	0.320
γ_3 或 ϕ_A	0.713	0.680	0.686	0.573	0.680	0.213	0.733	0.353
λ	0.806	0.720	0.793	0.686	0.813	0.326	0.820	0.480

6 示例：瑞士降雨数据

现在，我们使用 ***geostatsp*** 软件包中的瑞士降雨数据来实施我们的方法（有关该数据和 SIC-97 项目的详细说明，请参见 [8, 14] ）。图 6 展示了这些数据，其中包含 1986年5月在瑞士采集的 100 个每日降雨测量值。我们使用线性地统计模型对这些数据进行拟合

$$Y_i' \sim \mathrm{N}[X(s_i)^\top \beta + U(s_i), \tau^2]$$
$$\mathrm{COV}[U(s+h), U(s)] = \sigma^2 \rho(h; \phi_X, \phi_Y, \phi_A, \kappa)$$

其中， Y_i'是第 i 个位置的（经 Box-Cox 变换的）降雨测量值，协变量 X（s）由截距和位置 s 的海拔组成，而 U（s）是零均值高斯过程。相关函数 ρ是具有范围参数 ϕ_X 和 ϕ_Y、旋转参数 ϕ_A以及形状参数 κ 的各向异性 Matérn。下面使用 ***geostatsp*** 中的 swissRain 和 swissAltitude 对象拟合 LGM，并通过 *geostatsp::lgm()* 函数计算模型参数的最大似然估计。

```
R> 数据("swissRain", 软件包 = "geostatsp") R> swissRain$海拔 =
raster::extract(swissAltitude, swissRain) R> mle1 =geostatsp::lgm(公式 = rain ~ 海
拔, 数据 = swissRain,
```

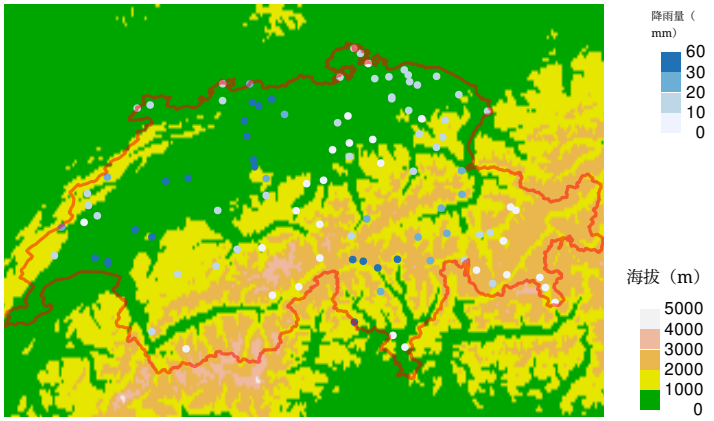


图 6：瑞士降雨数据（蓝色点及顶部图例）与海拔（背景颜色及底部图例）。

+ 协变量 = swissAltitude, fixBoxcox = FALSE, fixShape = FALSE, +
fixNugget = FALSE, aniso = TRUE, reml =FALSE, grid = 20)

如第 3 节算法的步骤1所述，我们通过将 Matérn 形状参数固定为 κ0.5、0.9、10、20 和 100（使用参数 fixNuggeet = FALSE 和 shape = 0.5），拟合另外五个模型。将包含这六个 *lgm()* 输出的列表传递给*gpuLik::configParams()* 函数，以创建代表点。

R> 重新参数化 = gpuLik::configParams(list(mle1, mle2, mle3, mle4, mle5,+
mle6), alpha = c(0.00001, 0.01, 0.1, 0.2, 0.25, 0.3, 0.5, 0.8,+ 0.9, 0.95,
0.99,0.999)) R> 使用的参数 <- 重新参数化$代表性参数 R> b <- 重新参数化$
Box-Cox变换

我们为 α 选择了 12 个显著性水平，其范围介于 0。00001 和 0。999 之间，用于近似似然面的等高线，从而得到 15318 行代表点（包括来自 6 个拟合线性地统计模型的 6 个最大似然估计）。

这些代表性参数的似然值通过下面的 *gpuLik::likfitLgmGpu()* 函数进行评估。boxcox 的取值为 ˆ Wald 置信区间的第 0.01 分位数和第 0.99 分位数之间等间距分布的 33 个值，以及最大似然估计 λ。NparamPerIter 参数指定同时评估 400 个似然值，这意味着需要一个 40,000 × 100 的矩阵来存储这 400 个协方差矩阵。由 *crossprodBatch()* 返回的矩阵 ssqYX $=(y,X)^{\intercal}V^{-1}(y,X)$包含 $36\cdot 400$ 行和 36 列，其中 34 列对应 34 个 Box-Cox 参数，其余两列分别对应两个协变量。

R> 结果 <-gpuLik::likfitLgmGpu(模型 = 最大似然估计1, 参数 = paramsUse,+ 要估计的参数 =
+ c('尺度参数', '组合范围', '形状', '块金效应', '各向异性参数1', '各向异性参数2', 'Box-Cox变换'),+
Box-Cox变换 = seq(b[1], b[9], len = 33), 置信区间水平 = 0.9,+ 每次迭代的参数数目 = 400, 全局 = c(128, 128), 局部 = c(16, 16), 局部缓存 = 2800)

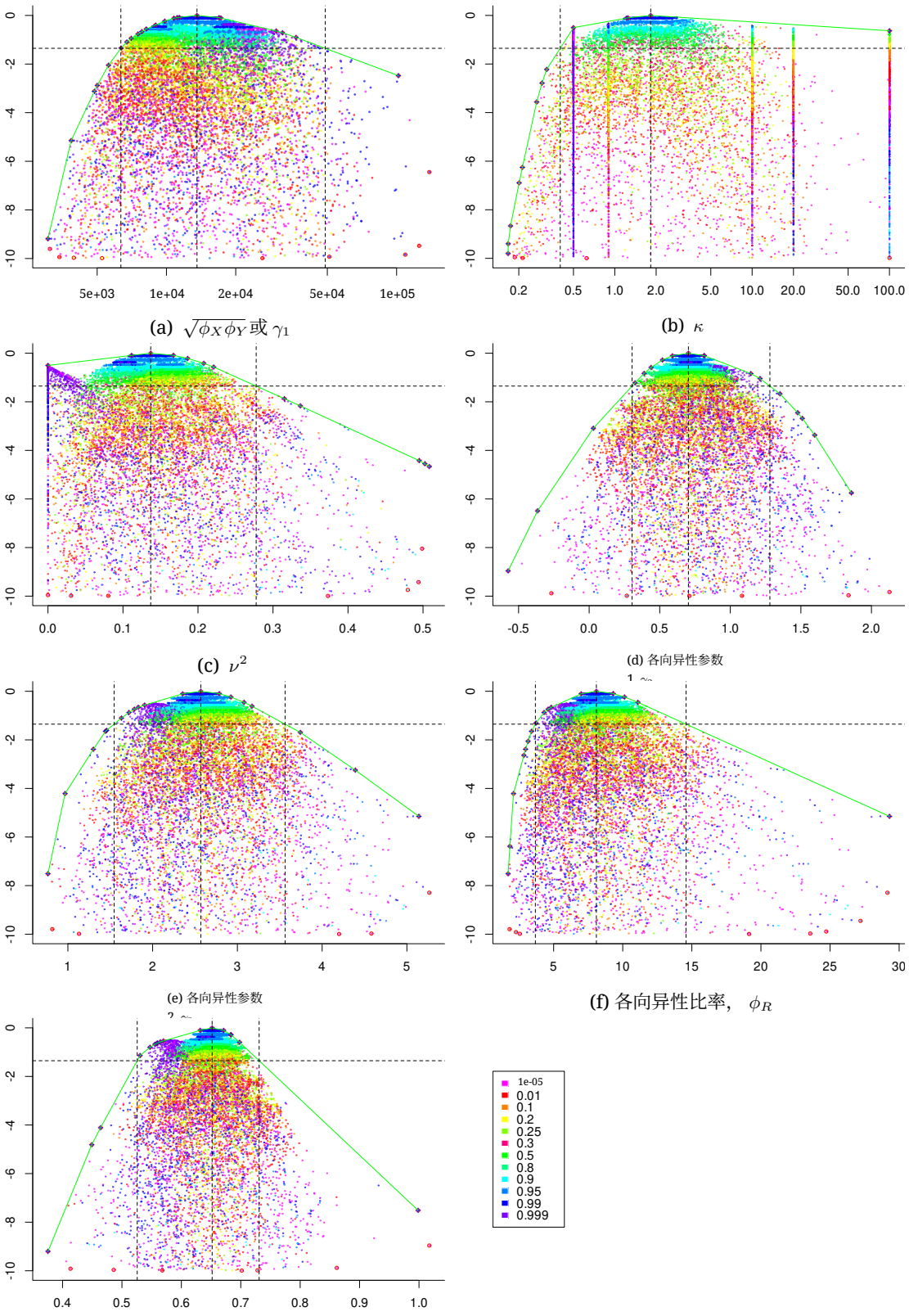
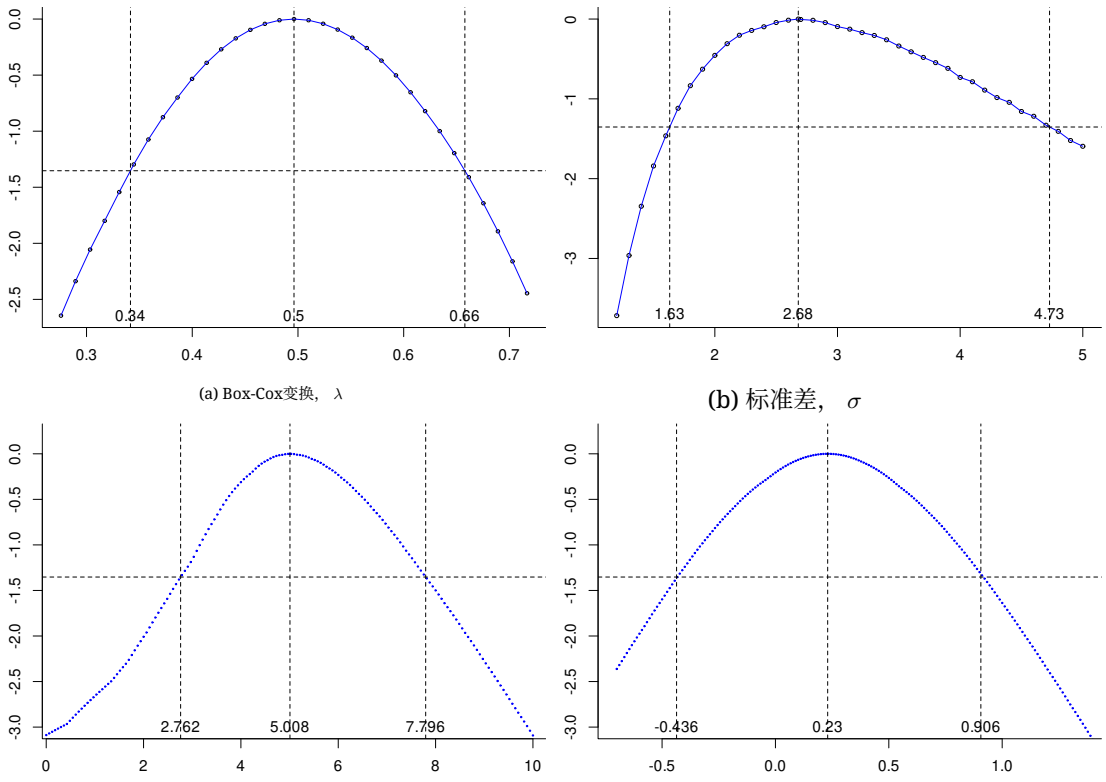


图7：根据瑞士降雨数据估计的相关参数的剖面对数似然。纵轴为对数似然减去最大似然估计处的似然值。虚线显示最大似然估计和90%置信区间。点云为代表点及其似然值，颜色显示每个点所绘制等高线的 α 分位数。

NlocalCache 参数与图形处理器上可用的本地内存容量有关，任一时刻将有 2,800 个实数存储在本地内存中。每个工作组由一个 16×16 工作项的网格组成，共包含一个 8×4 的工作组网格（总计 $16 \cdot 8 \times 16 \cdot 4$ 个工作项）。

图7展示了协方差参数的剖面似然。点云表示代表性参数及其对数似然（减去最大似然估计处的对数似然），颜色对应于绘制这些点所依据的等高线显著性水平α。图7b中显示的 κ 的剖面似然曲线在其最大值右侧相当平坦，似然的二次近似效果较差。将 κ 固定为 0.5、0.9、10、20 和 100 时的配置点清晰可见，呈现为垂直线。对于 ν^2 的剖面似然，可以看到大量点聚集在 0 处：我们将样本中一半为负的 ν^2 设为 0，另一半则从区间 (0，2) 中随机抽取。重新参数化后的各向异性参数 γ_2 和 γ_3 的对数似然相当接近二次函数，这支持采用极坐标变换。数据明显具有各向异性，尽管比率参数存在相当大的不确定性，其值可能小至 4，也可能高达 15。



(a) Box-Cox变换，λ　　(b) 标准差，σ

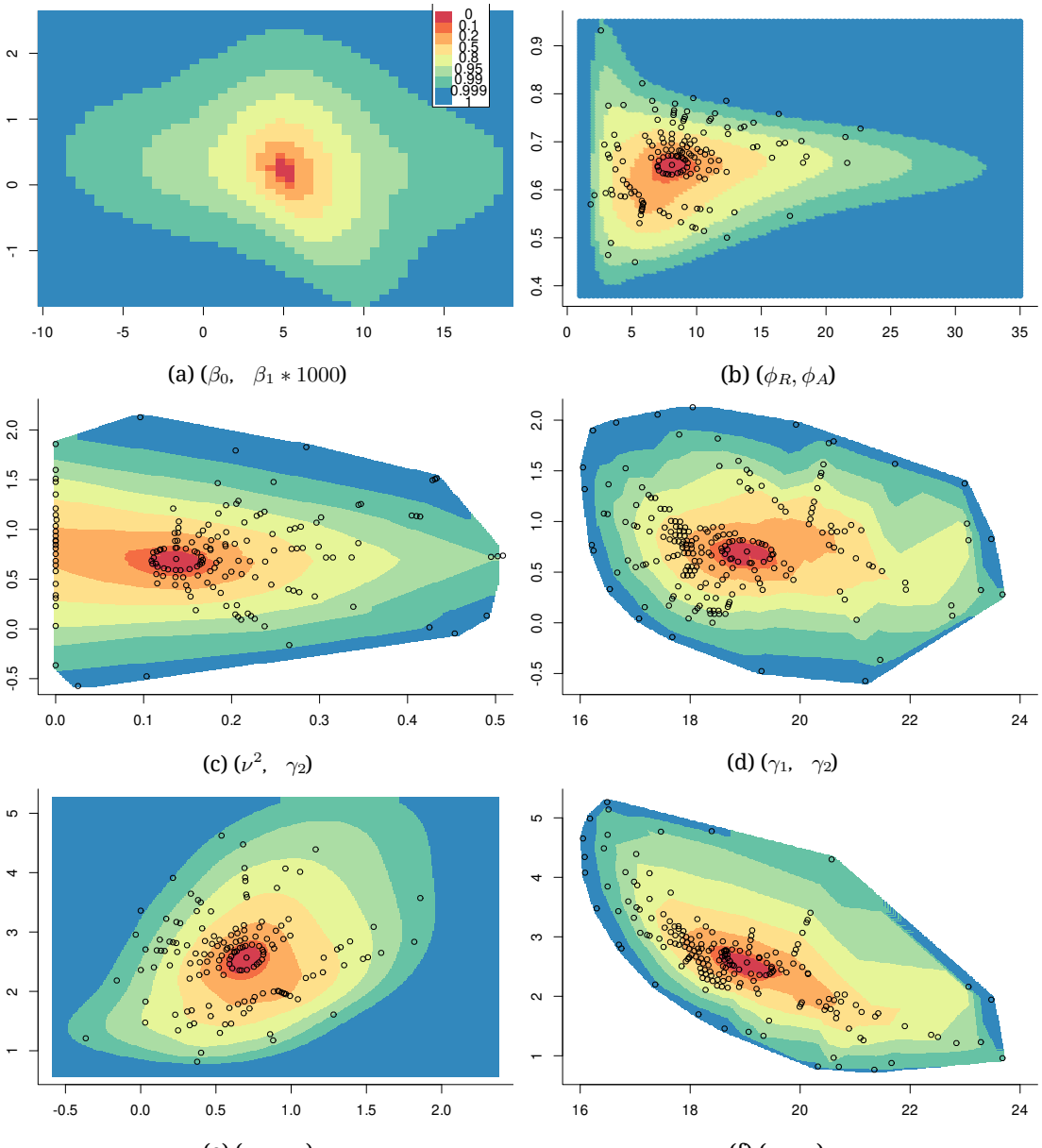
(c) 截距，β_0　　(d) 海拔，$\beta_1 \cdot 1000$

图8：根据瑞士降雨数据计算得到的 Box-Cox 变换参数、标准差和回归参数的剖面对数似然。每幅图中的水平虚线定义了近似的 90%置信区间。

图 9 中.....的二维 PLL 等高线图显示，数据具有各向异性。它们展示了重新参数化的优势：与自然参数的 PLL 等高线相比，内部参数的等高线形状更接近椭圆。随着各向异性比率的值趋近于 1（各向同性），各向异性弧度的置信区间估计值会变宽，最终覆盖整个区间（$-\pi/2, \pi/2$）。

表4显示，***gpuLik*** 得到的基于剖面的置信区间比***geostatsp***得到的基于Wald的置信区间更宽。两种方法得到的 λ置信区间非常接近。

表4将模型参数的基于似然的置信区间与***geostatsp***的基于 Wald 的置信区间进行了比较。图7至图9展示了模型参数的 PLL图以及一些二维 PLL图。



(a) (β_0, $\beta_1 * 1000$)

(b) (ϕ_R, ϕ_A)

(c) (ν^2, γ_2)

(d) (γ_1, γ_2)

(e) (γ_2, γ_3)

(f) (γ_1, γ_3)

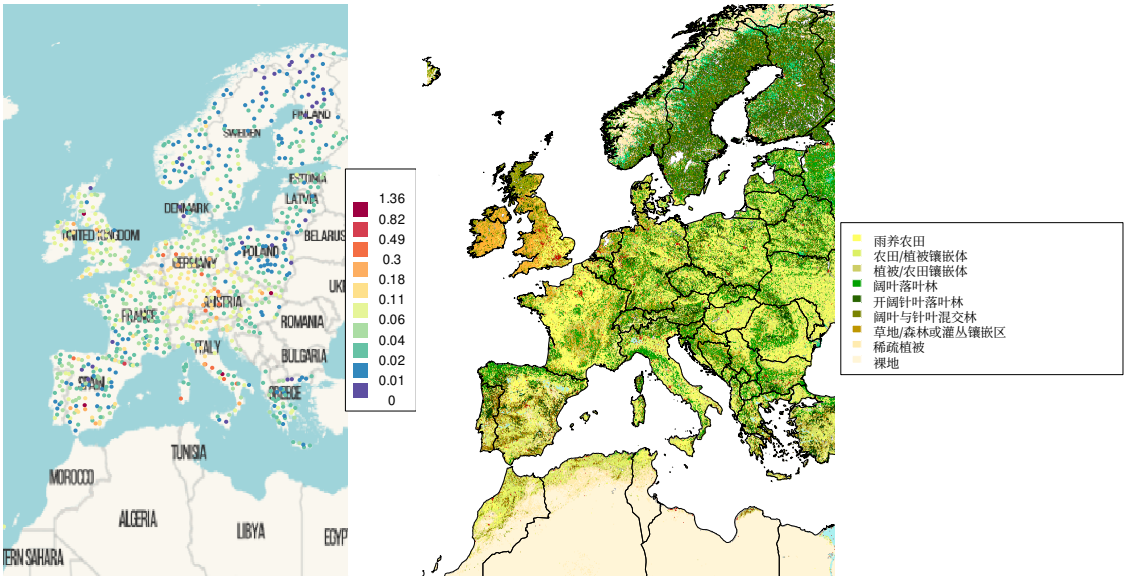
图 9：根据瑞士降雨数据计算的模型参数对的二维剖面对数似然等高线图。圆形绘图符号显示凸包上代表点的位置。

表4：瑞士降雨数据。Wald置信区间与基于剖面似然的置信区间的比较。

			基于似然的		Wald	
	符号表示法	估计值	ci0.05	ci0.95	ci0.05	ci0.95
(Intercept)	β_1	5.01	2.76	7.8	3.32	6.69
海拔*1000	$\beta_2 * 1000$	0.23	-0.44	0.91	-0.36	0.82
空间标准差	σ	2.68	1.63	4.74	1.57	4.58
尺度参数/1000	$\phi_X/1000$	38.62	21.79	118.63	18.54	80.42
组合范围/1000	$\sqrt{\phi_X\phi_Y}/1000$	13.58	6.34	48.96	NA	NA
各向异性比率	ϕ_R	8.09	3.68	14.58	4.73	13.86
形状	κ	1.83	0.40	> 100	-1.25	4.91
块金效应	ν^2	0.14	0.00	0.28	NA	NA
块金效应标准差	τ	0.99	0.00	1.41	0.62	1.60
各向异性角度（弧度）	ϕ_A	0.65	0.53	0.73	0.58	0.73
各向异性参数1	γ_2	0.70	0.30	1.28	NA	NA
各向异性参数2	γ_3	2.57	1.55	3.56	NA	NA
Box-Cox 变换	λ	0.50	0.34	0.66	0.34	0.65

7 示例：土壤汞数据

为通过更大的数据集说明我们的方法，我们将该方法应用于规模大得多的土壤汞数据集，该数据集包含829个观测值和29个预测变量。土壤汞数据取自“EuroGeoSurveys - FOREGS地球化学基线数据库”[15]，，可从 http://weppi.gtk.fi/publ/ foregsatlas/ForegsData.php 下载。图10a绘出了表层土壤中的汞浓度测量值。协变量 X 包含：海拔、土地类别（见图10b和10c）、夜间灯光、植被指数。



（a）土壤中的汞 （b）土地类型 (c) 土地类型

图 10：欧洲各位置的土壤汞测量值。

此流程与瑞士降雨数据示例中使用的流程相同。我们使用*geostatsp::lgm()*拟合 4 个模型以获得最大似然估计：第一个模型使用 Matérn 形状参数的最大似然估计，最后三个模型则将该参数固定为 0.5、0.8 和 1。*gpuLik::configParams()* 根据每个模型的最大似然估计配置候选点。所选等高线水平与瑞士降雨数据示例相同，参数配置矩阵有 12316 行（包括 4 个拟合线性地统计模型的最大似然估计、基于下方 *hgRes1* 对象的 $12 \cdot 726$ 个配置，

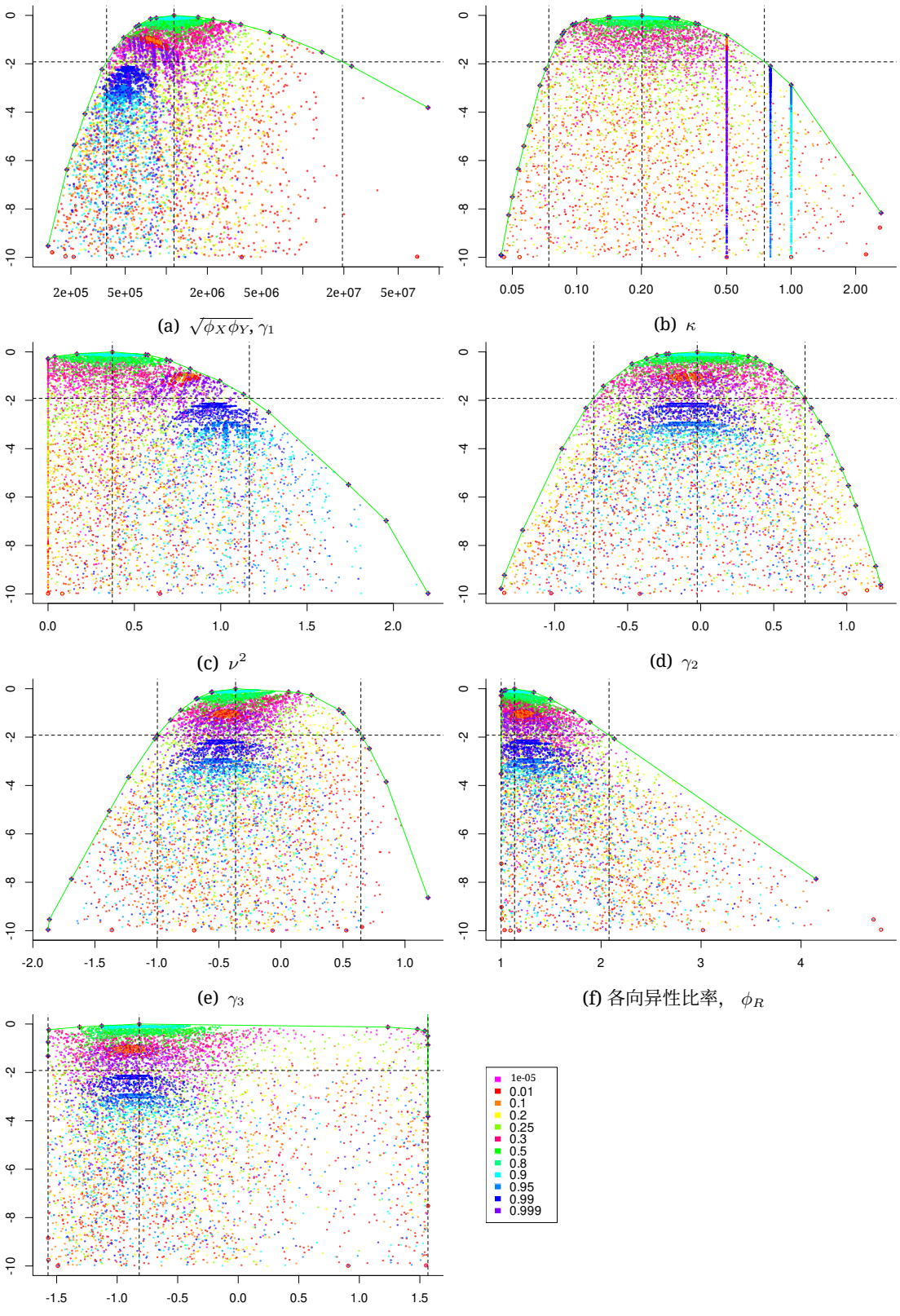
以及基于其他 3 个模型的 $3 \cdot 10 \cdot 120$ 个配置）和 5 列。通过在近似高斯分布的第 0.01（b[1]）和第 0.99（b[2]）分位数之间选择 31 个等间距值，配置 Box-Cox 变换的取值。我们使用下面的代码，通过*gpuLik::likfitLgmGpu()* 获得对数似然值。

```
R> hgRes1 =geostatsp::lgm(公式 = HG ~ 海拔 + 土地 + 夜间 + evi,+ 数据 = hgm,
grid = 20, 协变量 = covList, fixBoxcox = FALSE,+ fixShape = FALSE, fixNugget =
FALSE, reml = FALSE, aniso = TRUE)

R> reParamaters <- gpuLik::configParams(list(hgRes1, hgRes2, hgRes3, hgRes4),+
alpha = c(0.00001, 0.01, 0.1, 0.2, 0.25, 0.3, 0.5, 0.8, 0.9, 0.95,+ 0.99,0.999))
R> paramsUse <- reParamaters$representativeParamaters[, 1:5]R> b <-
reParamaters$Box-Cox变换

R> 结果 <- gpuLik::likfitLgmGpu(模型 = hgRes1, 参数 = paramsUse,+ Box-Cox变
换 = seq(b[1],b[9], len = 31), 置信区间水平 = 0.95, 类型 ="double",+ 用于
beta 的凸包 =FALSE, 每次迭代的参数数量 = 400, 全局数量 = c(256,+ 256),
局部数量 = c(16, 16), 局部缓存 = 2800, 详细程度 = c(1,+ 0))
```

土壤汞数据近似各向同性，且各向异性角度参数的置信区间涵盖整个定义域（$-\pi/2, \pi/2$）。图12中的二维剖面对数似然等高线图 (a)、(b) 使用内部参数，确实能显著改善似然值的二次近似。表5比较了参数估计值和置信区间与 Wald置信区间。对于大多数参数，剖面似然给出的置信区间比 Wald 方法更宽；对于各向异性弧度、各向异性比率和形状等参数，***gpuLik*** 给出的置信区间更好，因为它考虑了参数边界。同样，两种方法得到的变换参数置信区间非常相似。



(a) $\sqrt{\phi_X \phi_Y}, \gamma_1$

(b) κ

(c) ν^2

(d) γ_2

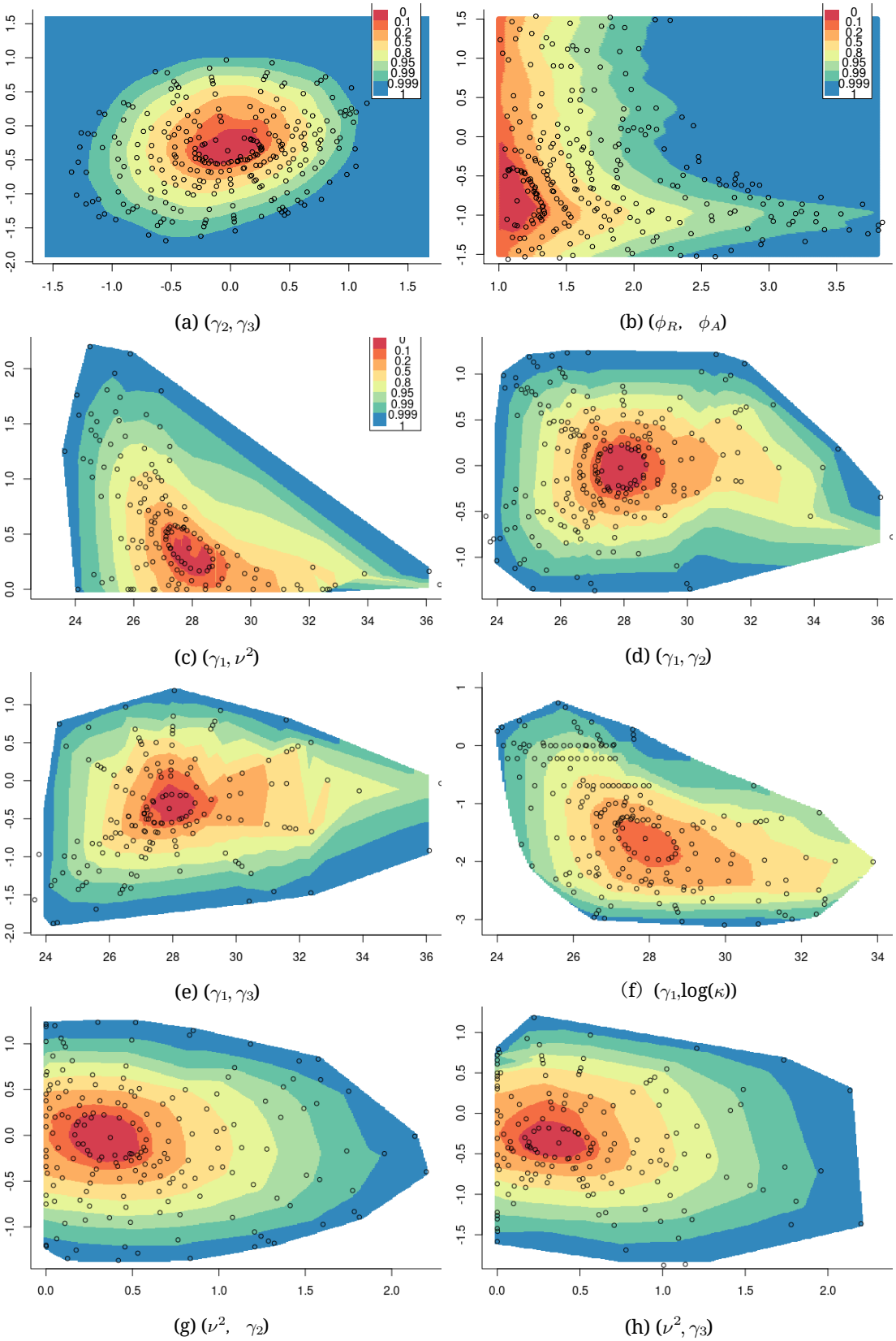
(e) γ_3

(f) 各向异性比率， ϕ_R

(g) 各向异性弧度， ϕ_A

(h) 图例

图 11：根据土壤汞数据估计的相关参数的剖面对数似然。纵轴表示对数似然减去最大似然估计处的似然值。虚线表示最大似然估计和 95%置信区间。点云表示代表点及其似然值，颜色表示每个点所取自的等高线的 α 分位数。



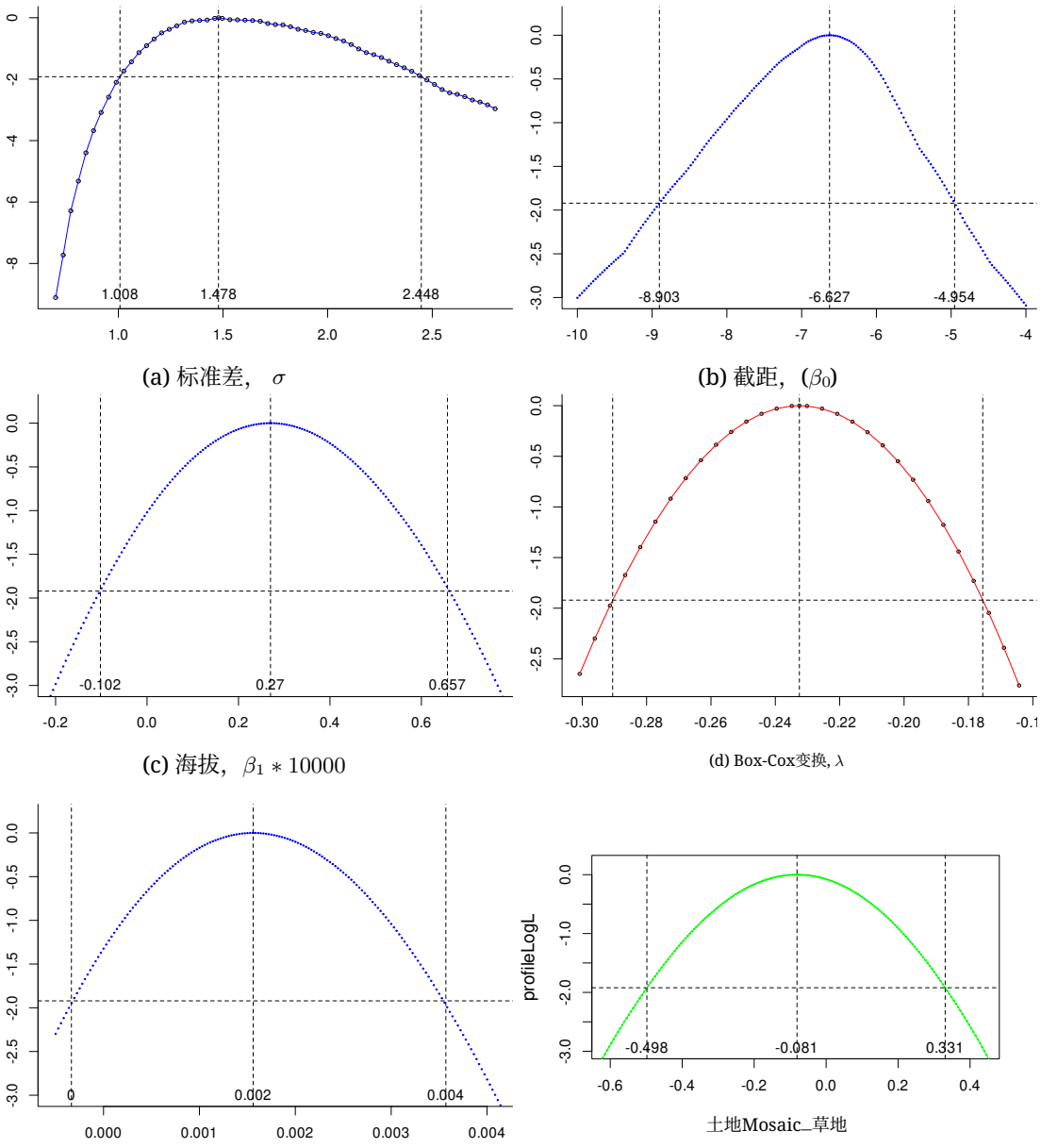
(a) (γ_2, γ_3) (b) $(\phi_R, \quad \phi_A)$

(c) (γ_1, ν^2) (d) (γ_1, γ_2)

(e) (γ_1, γ_3) (f) $(\gamma_1, \log(\kappa))$

(g) $(\nu^2, \quad \gamma_2)$ (h) (ν^2, γ_3)

图12：根据土壤汞数据计算得到的模型参数对的二维剖面对数似然等高线图。圆形绘图符号显示凸包上代表点的位置。



(a) 标准差， σ

(b) 截距， (β_0)

(c) 海拔， $\beta_1 * 10000$

(d) Box-Cox变换, λ

(e) 夜间, β_{16}

(f) 土地拼图_草地,β_5

图13：根据土壤汞数据计算的标准差、部分回归参数以及 Box-Cox 变换参数的剖面对数似然。每幅图中的水平虚线表示近似的 95%置信区间和最大似然估计。

表5：土壤汞数据的 LGM 拟合中模型参数的 95% 置信区间。展示的是由***gpuLik*** 通过图形处理器计算的剖面似然置信区间，以及由***geostatsp***计算的传统 Wald 置信区间。

			基于似然的		Wald	
	符号表示法	估计值	置信区间0.025	置信区间0.975	置信区间0.025	置信区间0.975
截距	β_1	-6.63	-8.90	-4.98	-7.80	-5.45
海拔*10000	$\beta_2 * 10000$	2.72	-1.02	6.57	-1.00	6.40
阔叶落叶林	β_3	-0.28	-0.65	0.08	-0.63	0.08
农田/植被镶嵌体/植被	β_4	-0.08	-0.43	0.25	-0.42	0.25
开阔针叶落叶林	β_5	-0.63	-1.21	-0.10	-1.16	-0.10
草地镶嵌体	β_6	-0.08	-0.50	0.33	-0.49	0.33
镶嵌植被 / 耕地	β_7	-0.37	-0.80	0.03	-0.78	0.03
混合阔叶林	β_8	-0.51	-1.01	-0.04	-0.97	-0.05
稀疏植被	β_9	-1.02	-1.60	-0.50	-1.52	-0.52
草本植被	β_{10}	0.00	-0.58	0.59	-0.57	0.58
森林镶嵌体或灌木地	β_{11}	-0.28	-0.91	0.32	-0.88	0.31
针叶树叶状的常绿林	β_{12}	-0.01	-0.60	0.57	-0.58	0.56
灌木地	β_{13}	0.20	-0.49	0.90	-0.48	0.88
草地或木本植被	β_{14}	-0.69	-1.64	0.20	-1.58	0.20
人工地表	β_{15}	-0.44	-2.38	1.47	-2.32	1.45
水体	β_{16}	-1.17	-3.78	1.35	-3.67	1.34
夜间*1000	$\beta_{17} * 1000$	1.57	-0.34	3.57	-0.32	3.44
evi	β_{18}	2.07	1.00	3.28	1.03	3.11
空间标准差	σ	1.48	1.01	2.45	0.91	2.40
尺度参数/1000	$\phi_X/1000$	1212.73	450.28	22100.81	220.56	6667.96
组合范围/1000	$\sqrt{\phi_X\phi_Y}/1000$	1138.60	365.27	19530.13	NA	NA
各向异性比率	ϕ_R	1.13	1.00	2.08	0.64	2.00
形状	κ	0.20	0.07	0.75	-0.14	0.54
块金效应	ν^2	0.37	0.00	1.17	NA	NA
块金效应标准差	τ	0.90	0.00	1.60	0.38	2.13
各向异性角度（弧度）	ϕ_A	-0.82	-1.57	1.57	-2.55	0.92
各向异性参数1	γ_2	-0.02	-0.73	0.71	NA	NA
各向异性参数2	γ_3	-0.37	-1.00	0.64	NA	NA
Box-Cox 变换	λ	-0.23	-0.29	-0.18	-0.29	-0.17

8 讨论

由于涉及大量非线性参数（两个范围参数、各向异性参数、块金效应），各向异性地统计模型的推断颇具挑战；若还将 Matérn 形状参数和 Box-Cox 参数视为未知，难度则更大。许多地统计分析软件无法计算模型参数的剖面似然，而是采用插件式基于信息的（Wald）置信区间来处理回归参数（将相关参数固定在其最大似然估计值上）。这些置信区间没有考虑其他参数的不确定性对 β 的不确定性所产生的影响。此外，基于信息量的置信区间并非总是可行，例如当参数估计值接近边界时，信息矩阵可能不是正定的。

计算并展示剖面似然有助于评估空间模型参数，而基于似然的置信区间相比基于信息量的置信区间，覆盖范围更广、覆盖率也更高。要获得任一参数的剖面似然，需要反复进行大量似然值评估；这项计算密集型工作

借助图形处理器进行并行计算会更容易完成。图形处理器旨在同时处理数千个并行线程，擅长重复进行似然评估，并能显著缩短运行时间。此处介绍的基于图形处理器的技术也可应用于其他含有多个方差参数的变换高斯模型，例如包含多个随机效应的层次模型、具有潜在随机游走的平滑模型，或包含额外随机效应的更复杂空间模型。

我们的工作着重于使用不进行近似的稠密矩阵空间算法，并借助图形处理器，使标准的基于Cholesky分解的似然评估能够应用于包含 800 个空间位置、规模相当大的土壤汞数据集。扩展到更大数据集时，主要的计算限制因素是图形处理器内存。以双精度存储 10,000 个空间位置对应的 400 个方差矩阵，需要 300GB 内存；在撰写本文时，这种容量只有专用且昂贵的图形处理器硬件才具备。第二个限制是最大化似然函数所需的时间，因为这一过程需要在中央处理器上按顺序多次评估似然。第三，对于包含大量彼此接近的点（相对于范围参数而言）的较大数据集，舍入误差可能导致方差矩阵不再正定。不过，只要使用足够强大的图形处理器，任何能够通过现有稠密矩阵Cholesky方法获得最大似然估计的数据集，都可以采用这种基于图形处理器的剖面似然方法进行处理。

当数据集规模更大（或时间限制更严格），超出稠密 Cholesky 算法的处理能力时，就需要采用 SPDE方法 [16], 谱近似 [17], 低秩近似 [18, 19], 或协方差截尾 [20, 21] 等近似方法。不过，并行创建代表点并计算似然值的统计方法论与这类近似方法完全兼容；但要实现这些方法，就需要开发新的 GPU 代码，以高效执行并行计算。SPDE方法已成为贝叶斯推断中占主导地位的近似方法，而在当前的***gpuLik***软件包中加入与批量矩阵函数对应的稀疏矩阵实现，将能基于 SPDE 模型为更大的数据集计算 GPU 剖面似然。

将方法论扩展到具有非高斯响应分布的潜在高斯模型，是一项很有价值但复杂得多的工作。似然没有闭式表达式，采用 [22] 和 [10] 所用的拉普拉斯近似将是最简单的方法。要在 GPU 上实现这一点，就需要高效地批量优化联合概率 $\pi(Y, U)$中的“内层”问题，这将是一项艰巨的工作。

总之，我们的方法利用图形处理器上的并行似然评估，是一种可行且便捷的方法，可用于理解各向异性空间模型中协方差参数的不确定性，并将这种不确定性传递到对 β 回归参数的推断中。这包括经常被忽视的Matérn形状参数；我们发现，尽管 κ 的剖面似然通常较为平坦，但通常存在可识别的最大值，因此可以将其剖面化消去。此外，我们证明，对于中等规模的空间数据集，使用精确形式的似然进行频率学派推断（不近似方差矩阵）是可行的。

9 附录：受限最大剖面似然

最大似然估计的一个问题是方差会被低估。考虑一个简单情形： Y的协方差矩阵为 $\sigma^2 I$，那么

$$\hat{\sigma}^2 = \frac{1}{n}(y - X\hat{\beta})^\top (y - X\hat{\beta}),$$
$$\mathbb{E}[\hat{\sigma}^2] = \frac{n-p}{n}\sigma^2 < \sigma^2,$$

其中， $p = \text{rank}(X)$是 β中的元素个数。估计偏差源于没有考虑用于估计 β的自由度。由 [23] 和 [24] 提出的限制最大似然（REML）是一种通常能够得到无偏估计量或比 MLE 有偏程度更小的估计量的方法。REML 消除了 X 对 $\hat{}\sigma^2$ 和 $\hat{}\omega$的影响。在该方法中，我们找到一个满秩矩阵 A，使得 $AX = 0$。 $n \times n$ 投影矩阵 $S = I - X(X^\top X)^{-1}X^\top$满足 $SX = 0$，然而， S是退化的，因为 $\text{rank}(S) = n - p$，因此我们仅从 S 中取出 $n-p$个线性无关的行，构成矩阵 $A_{(n-p)\times n}$。将观测值 Y线性变换为 $Y^* = AY$，则 $\mathbb{E}(Y^*) = AX\beta = 0$。回顾式（1），我们得到 $Y \sim N_n(X\beta,\ \sigma^2 V)$，因此

$$Y^* \sim N_n(0, \sigma^2 AVA^\top).$$

REML 方法的原理是通过最大化下方给出的限制对数似然函数来估计方差参数 σ^2，其中 y^* 表示随机变量 Y^*的一次实现，

$$-2\ell^*(\sigma^2,\omega;y^*)=y^{*\intercal}(\sigma^2AVA^\intercal)^{-1}y^*+\log|\sigma^2AVA^\intercal|+(n-p)\log(2\pi), \tag{14}$$

对 σ^2 求导并令其等于零，可得 σ^2 的无偏 REML 估计值

$$\hat{\sigma}^2_{reml}(\omega)=\frac{y^{*\intercal}(AVA^\intercal)^{-1}y^*}{n-p}.$$

REML 准则基于 y^* 的似然函数 (14)，其中不包含 X。REML 避免估计 β，因此能够为 σ^2 产生无偏估计值。

利用 [25] 在第 5.1 节中推导出的两个恒等式，我们可以用观测数据 y 表示 (14)：

$$|A(\sigma^2V)A^\intercal|=|\sigma^2V||X^\intercal(\sigma^2V)^{-1}X|$$

and

$$y^*(\sigma^2AVA^\intercal)^{-1}y^*=(y-X\hat{\beta})^\intercal(\sigma^2V)^{-1}(y-X\hat{\beta}),$$

其中 $\hat{}\,\beta=(X^\intercal V^{-1}X)^{-1}X^\intercal V^{-1}y$。于是，$y$ 的限制性对数似然为

$$\begin{aligned}-2\ell^*(\sigma^2,\omega;y)=&(y-X\hat{\beta})^\intercal(\sigma^2V)^{-1}(y-X\hat{\beta})+(n-p)\log\sigma^2+\log|V|\\&+\log|X^\intercal V^{-1}X|+n\log(2\pi).\end{aligned}$$

如果对 y 应用 Box-Cox 变换，则纳入 Box-Cox 变换参数后的 y 的一般限制对数似然为

$$\begin{aligned}-2\ell^*(\sigma^2,\omega,\lambda;y)=&(y'-X\hat{\beta})^\intercal(\sigma^2V)^{-1}(y'-X\hat{\beta})+(n-p)\log\sigma^2+\log|V|\\&+\log|X^\intercal V^{-1}X|+n\log(2\pi)-2(\lambda-1)\sum_{i=1}^{n}\log y_i,\end{aligned} \tag{15}$$

该式不依赖于矩阵 A 的选择。对（15）关于 σ^2 求导并令其等于零，可得到以给定参数 ω 和 λ 以及变换后的数据 y' 表示的最大似然估计 $\hat{}\,\sigma^2_{reml}$，

$$\hat{\sigma}^2_{reml}(\omega,\lambda)=\frac{(y'-X\hat{\beta})^\intercal V^{-1}(y'-X\hat{\beta})}{n-p}. \tag{16}$$

将（16）代入（15），得到（ω, λ）的剖面限制对数似然

$$\begin{aligned}-2\ell^*_p(\omega,\lambda;y)=&(n-p)\log\frac{(y'-X\hat{\beta})^\intercal V^{-1}(y'-X\hat{\beta})}{n-p}+\log|V|+\log|X^\intercal V^{-1}X|\\&-2(\lambda-1)\sum_{i=1}^{n}\log y_i+n\log(2\pi)+n-p.\end{aligned} \tag{17}$$

与（7）中给出的“ml”剖面似然相比，有效维数从 $\ell(\cdot)$ 的 n 降低为 $\ell^*_p(\cdot)$ 的（$n-p$），如果 p 较大，这一区别十分重要。

9.1 使用限制性最大似然求 β_p 的剖面似然

~如果采用 REML 方法，那么 $\ell_p(\beta_i,\beta_{\beta_i},\hat{}\,\sigma_{\beta_i},\omega,\lambda;\,y)$ 将等价于

$$\begin{aligned}-2\ell^*_p(\omega,\lambda;y)=&(n-p)\log\frac{(\tilde{y}-\tilde{X}\hat{\tilde{\beta}}_{\omega,\lambda})^\intercal V^{-1}(\tilde{y}-\tilde{X}\hat{\tilde{\beta}}_{\omega,\lambda})}{n-p}+\log|V|+\log|X^\intercal V^{-1}X|\\&-2(\lambda-1)\sum_{i=1}^{n}\log y_i+n\log(2\pi)+n-p.\end{aligned} \tag{18}$$

参考文献

[1] George EP Box 和 David R Cox. 变换分析. 英国皇家统计学会会刊：B辑（方法论）, 26(2):211–243, 1964.[2] Dionissios T Hristopulos. 空间数据建模随机场：科学家和工程师入门. 施普林格·自然, 2020.[3] Michael L Stein. 空间数据插值：克里金法的一些理论. 施普林格科学与商业媒体, 1999.[4] Peter J Diggle 和 Emanuele Giorgi. 全球公共卫生的基于模型的地统计学：方法与应用. Chapman and Hall/CRC, 2019.[5] Kathryn A Haskard, Brian R Cullis, 和 Ar¯unas P Verbyla. 各向异性 Matérn 相关与使用限制性最大似然进行空间预测. 农业、生物与环境统计学杂志, 12(2):147–160, 2007.[6] Charlotte Baey, Paul-Henry Cournède, 和 Estelle Kuhn. 非线性混合效应模型中方差分量似然比检验统计量的渐近分布. 计算统计学与数据分析, 135:107–122, 2019.[7] Peter M Visscher. 关于用于检验方差分量的似然比检验渐近分布的注记. 双生子研究与人类遗传学, 9(4):490–495, 2006.[8] Patrick E Brown. 基于模型的地统计学：简易方法. 统计软件杂志, 63:1–24, 2015.[9] PD Dauncey, M Kenzie, N Wardle, 和 GJ Davies. 处理背景形状的不确定性：离散剖面分析法. 仪器学杂志, 10(04):P04015, 2015.[10] Håvard Rue, Sara Martino, 和 Nicolas Chopin. 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. 英国皇家统计学会会刊：B辑（统计方法论）, 71(2):319–392, 2009.[11] Patrick E Brown Kamal Rai. 使用分块复合似然对空间模型进行估计和预测. 2022. 手稿准备中.[12] Hao Zhang. 基于模型的地统计学中的不一致估计与渐近等价插值. 美国统计学会会刊, 99(465):250–261, 2004.[13] Richard Arnold Johnson, Dean W Wichern, 等. 应用多元统计分析，第6卷. 培生，英国伦敦：, 2014.[14] Patrick E Brown. R 软件包‘geostatsp’. 2021.[15] R Salminen, MJ Batista, M Bidovec, A Demetriades, B De Vivo, W De Vos, M Duris, A Gilucis, V Gre- gorauskiene, J Halamic, 等. 欧洲地球化学图集。第1部分——背景信息、方法论和地图. 芬兰地质调查局. 网址：http://www. gtk. fi/publ/foregsatlas/index. php *[30*/03/*2010]*,2005年.[16] Finn Lindgren, Håvard Rue, 和 Johan Lindström. 高斯场与高斯马尔可夫随机场之间的显式联系：随机偏微分方程方法. 英国皇家统计学会会刊：B辑（统计方法论）, 73(4):423–498, 2011.[17] Benjamin M Taylor, Tilman M Davies, Barry S Rowlingson, 和 Peter J Diggle. lgcp：用于时空对数高斯考克斯过程推断的 R 软件包. arXiv 预印本 arXiv:1110.6054, 2011.[18] Sudipto Banerjee, Alan E Gelfand, Andrew O Finley, 和 Huiyan Sang. 大型空间数据集的高斯预测过程模型. 英国皇家统计学会会刊：B辑（统计方法论）, 70(4):825–848, 2008.[19] Noel Cressie 和 Gardar Johannesson. 超大型空间数据集的固定秩克里金法. 英国皇家统计学会会刊：B辑（统计方法论）, 70(1):209–226, 2008.[20] Reinhard Furrer, Marc G Genton, 和 Douglas Nychka. 用于大型空间数据集插值的协方差截尾. 计算与图形统计学杂志, 15(3):502–523, 2006.

[21] Cari G Kaufman、Mark J Schervish 和 Douglas W Nychka。用于大型空间数据集基于似然的估计的协方差截尾。美国统计学会会刊，103(484):1545–1555，2008年。[22] Alex Stringer、Patrick E Brown 和 Jamie Stafford。扩展潜在高斯模型中后验分布的快速、可扩展近似。*arXiv* 预印本 *arXiv:2103.07425*，2021年。[23] H Desmond Patterson 和 Robin Thompson。当区组大小不等时恢复区组间信息。*Biometrika*，58(3):545–554，1971年。[24] David Roxbee Cox 和 Nancy Reid。参数正交性与近似条件推断。英国皇家统计学会会刊：*B*辑（方法论），49(1):1–18，1987年。[25] SR Searle、RL Quaas 等。方差分量笔记：近期方差分量估计方法的详细描述及其在动物育种中的应用。1978年。