arXiv:2602.03483v1 [stat.ME] 3 Feb 2026

通过惩罚近邻选择实现大规模数据集的克里金

Francisco Cuevas-Pacheco[1*]和 Jonathan Acosta[2†]

[1*]费德里科·圣玛丽亚技术大学数学系，Avenida España 1680，智利瓦尔帕莱索，邮编2340000。[2]智利天主教大学统计系，Avenida Vicuña Mackenna 4860，马库尔，邮编7820436，圣地亚哥，智利。

*通讯作者。电子邮件：francisco.cuevas@usm.cl；贡献作者：jonathan.acosta@uc.cl；

†这些作者对本研究的贡献相同。

摘要

克里金是空间预测的基本工具，但其计算复杂度为 $O(N^3)$ ，对于大规模数据集而言变得难以承受。虽然使用 K最近邻的局部克里金可以解决这一问题，但 K 的选择通常依赖于特设准则，未能考虑空间相关结构。我们提出一种惩罚克里金框架，将LASSO型惩罚直接纳入克里金方程，以实现自动的、数据驱动的邻居选择。我们进一步将其扩展到自适应LASSO，使用考虑空间相关结构的数据驱动惩罚权重。我们的方法通过 ℓ_1 正则化确定哪些观测值贡献非零权重，并借助基于有效样本量的新准则选择惩罚参数，从而在预测精度与信息冗余之间取得平衡。数值实验表明，惩罚克里金能够根据潜在的空间相关性自动调整邻域结构，对更平滑过程选择更少的邻居，对高度可变的场选择更多邻居，同时以大幅降低的计算成本保持与全局克里金相当的预测精度。

关键词： 空间插值、局部克里金、LASSO 惩罚项、有效样本量、邻居选择、稀疏克里金

1 引言

通过克里金进行空间预测已成为地统计学中的基本工具，用于在不同科学领域对空间现象进行插值与预测（Cressie1993b）。然而，标准克里金方程需要对一个 $N \times N$ 协方差矩阵求逆，从而导致 $O(N^3)$ 的计算复杂度（Cressie 和 Johannesson 2008），随着数据集规模增长到数千乃至数百万个观测值，这一复杂度变得难以承受。事实上，现代应用经常产生海量空间数据集：MODIS 和 Landsat 等卫星遥感平台在数百万个位置生成多光谱影像，ERA5 等气候再分析产品在数十万个空间点上提供格点场，激光雷达测量则生成包含数百万个高程测量值的点云。海洋观测系统体现了这一规模：仅 COBE 海表温度数据集就包含超过 43,000 个空间位置，以 1° × 1°分辨率覆盖全球（Ishii等 2005），而现代卫星高度计测量和 Argo 浮标网络产生的数据集甚至更大。

为应对这一计算挑战，地统计学界已提出多种策略。最广泛采用的方法之一是使用局部邻域，仅利用数据集的子集来计算克里金预测 K，即使用最近邻居，而非整个数据集（Gramacy 和 Apley，2015）。其他策略包括通过协方差渐消使协方差矩阵呈现稀疏性（Furrer 等，2006）和低秩近似（Cressie 和 Johannesson，2008）。Heaton 等人（2019）对这些方法进行了全面比较，结果表明，尽管基于邻域的方法计算效率较高，但其性能在很大程度上取决于邻域大小和配置的选择。

尽管 K-最近邻克里金被广泛使用，但邻域大小的选择 K 仍然存在问题。选择通常依赖用户指定的标准或交叉验证程序，而对于大规模数据集而言，这两种方法的计算成本都很高。Emery （2009）指出，固定的邻域选择策略未必能在具有相同观测值数量的所有可能选择中实现克里金方差最小化。 K最近邻方法未能考虑空间相关结构和信息冗余：在空间相关性强的区域，较小的邻域可能就足够；而在空间变异性高的区域，为了达到相当的预测精度，可能需要更多邻居。

惩罚回归方法，尤其是最小绝对收缩和选择算子（LASSO），由Tibshirani （1996）提出，通过 ℓ_1 正则化为自动变量选择提供了一个原则性框架。

LASSO 惩罚项会使回归系数变得稀疏，从而在估计模型参数的同时有效地执行变量选择。已有多位作者将这一方法拓展到克里金领域：Hung(2011) 提出了惩罚盲克里金，用于在泛克里金中自动选择均值函数，并具有预言机性质；而 Park (2021) 将 LASSO 正则化应用于克里金，以进行全局趋势模型选择，并证明该方法提高了预测准确度和计算效率。然而，这些方法主要关注均值函数设定，而非旨在提升计算效率的邻居选择。

在这些研究成果的基础上，我们提出一种惩罚克里金框架，将 LASSO型惩罚直接纳入克里金方程，以实现稀疏邻域选择。与传统的 K最近邻方法不同，后者预先固定邻居的数量和身份；我们的方法会自动确定在每个预测位置上，哪些观测值会对克里金预测器产生非零权重。稀疏度由调优参数 η 控制，我们通过一种基于有效样本量概念的准则来选择该参数，该准则会明确考虑空间相关结构。此准则在预测精度与空间数据中的信息冗余之间取得平衡，为计算开销较大的交叉验证提供了一种替代方案。

我们的主要贡献有三点。首先，我们重新表述了克里金问题，引入 ℓ_1 惩罚项，使克里金权重具有稀疏性，同时保持克里金固有的无偏性约束。其次，我们提出一种基于有效样本量的准则，用于自动选择惩罚参数，并明确考虑空间相关性和信息冗余，从而无需计算成本高昂的交叉验证。第三，通过对模拟数据和真实数据进行数值实验，我们证明，所提出的惩罚克里金方法能够根据底层空间相关性自动调整邻域结构：对于更平滑的过程选择较少的邻居，而在空间变异性较高的区域选择较多的邻居，同时以大幅降低的计算成本，达到与全局克里金相当的预测精度。

本文其余部分安排如下。第 2 节回顾克里金方程，并阐明块金效应与岭惩罚之间的联系，从而说明采用惩罚方法的动机。第3 节介绍我们的 LASSO惩罚克里金形式、提出自适应LASSO扩展，并分析解路径随惩罚参数的变化。第 4节提出用于选择调优参数的有效样本量准则。第 5 节通过数值实验考察邻居选择如何适应不同的协方差结构和空间配置。第 6 节将该方法应用于两个空间尺度不同的真实数据集：汝拉重金属数据集和大尺度 COBE 海表温度数据集。最后，第7 节讨论研究结果的意义及未来研究方向。

2 克里金方程

我们首先定义符号并回顾克里金框架，它构成了我们惩罚方法的基础。本节还将说明惩罚如何通过块金效应自然地出现在克里金中，从而为我们对 LASSO 的扩展提供理论依据。

令 $\{\boldsymbol{s}_i\}_{i=1}^N \subset \mathbb{R}^d$ 为一组 N 空间位置，令 $\boldsymbol{Z} = (Z(\boldsymbol{s}_1) \ldots, Z(\boldsymbol{s}_N))^\top$ 为高斯随机场 Z 在相应位置上的观测值向量，其均值函数为 $\mu(\boldsymbol{s}) = \mathbb{E}[Z(\boldsymbol{s})]$ ，协方差函数为 $C(\boldsymbol{s}, \boldsymbol{s}') = \text{cov}(Z(\boldsymbol{s}) Z(\boldsymbol{s}'))$。我们希望利用观测值向量 $\boldsymbol{Z}$ 中的信息，预测未观测位置 $\boldsymbol{s}_0$ 处的 Z 。我们引入以下记号：$\Sigma = \text{Var}[\boldsymbol{Z}]$, $\boldsymbol{c}_0 = \text{cov}(Z(\boldsymbol{s}_0) \boldsymbol{Z}) = (C(\boldsymbol{s}_0, \boldsymbol{s}_1) \ldots, C(\boldsymbol{s}_0, \boldsymbol{s}_N))^\top$，以及 $\sigma_0^2 = \text{Var}[Z(\boldsymbol{s}_0)]$。

空间预测通过提出不同空间位置处观测变量的线性组合来实现。这一过程称为克里金。形式上，我们考虑以下预测器

$$P(\boldsymbol{Z}, \boldsymbol{s}_0) = \lambda_0 + \sum_{i=1}^{N} \lambda_i Z(\boldsymbol{s}_i) = \lambda_0 + \boldsymbol{\lambda}^\top \boldsymbol{Z},$$

其中， λ_0 是用作截距的实数， $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)^\top$ 是用作权重的实数向量。预测器 $P(\boldsymbol{Z}, \boldsymbol{s}_0)$ 旨在通过最优组合观测值来预测位置 $\boldsymbol{s}_0$ 处的未观测值，其中权重 $\boldsymbol{\lambda}$ 的选择以最小化预测误差并确保无偏性为目标。

通常，通过最小化均方误差（MSE）函数来获得系数向量 $\boldsymbol{\lambda}$：

$$\mathrm{MSE}(P(\boldsymbol{Z}, \boldsymbol{s}_0)) := \mathbb{E}\left[\left(Z(\boldsymbol{s}_0) - P(\boldsymbol{Z}, \boldsymbol{s}_0) \right)^2 \right] = \boldsymbol{\lambda}^\top \Sigma \boldsymbol{\lambda} - 2\boldsymbol{\lambda}^\top \boldsymbol{c}_0 + \sigma_0^2, \tag{1}$$

然而，$\mathrm{MSE}(P(\boldsymbol{Z}, \boldsymbol{s}_0))$ 不依赖于 λ_0。因此，有必要施加所谓的无偏性条件，即 $\mathbb{E}[P(\boldsymbol{Z}, \boldsymbol{s}_0)] = \mathbb{E}[Z(\boldsymbol{s}_0)]$。因此，克里金预测器的权重通过求解约束最小化问题来获得

$$\begin{aligned} \min_{\boldsymbol{\lambda} \in \mathbb{R}^N} \quad & \boldsymbol{\lambda}^\top \Sigma \boldsymbol{\lambda} - 2\boldsymbol{\lambda}^\top \boldsymbol{c}_0 + \sigma_0^2 \\ \text{s.t.} \quad & \mathbb{E}[P(\boldsymbol{Z}, \boldsymbol{s}_0)] = \mathbb{E}[Z(\boldsymbol{s}_0)] \end{aligned} \tag{2}$$

根据对均值函数所作的假设，可得到不同类型的克里金：简单克里金、普通克里金和泛克里金（Schabenberger和Gotway 2005）。

对于均值函数已知这一假设，容易看出（2）的解为 $\widehat{\boldsymbol{\lambda}} = \Sigma^{-1}\boldsymbol{c}_0$，且 $\widehat{\lambda}_0 = \mu(\boldsymbol{s}_0) - \widehat{\boldsymbol{\lambda}}^\top \boldsymbol{\mu}$，其中 $\boldsymbol{\mu} = (\mu(\boldsymbol{s}_1) \ldots, \mu(\boldsymbol{s}_N))^\top$。在该假设下，我们得到简单克里金估计量，

$$\widehat{P}_{\mathrm{sk}}(\boldsymbol{Z}, \boldsymbol{s}_0) = \mu(\boldsymbol{s}_0) + \boldsymbol{c}_0^\top \Sigma^{-1}(\boldsymbol{Z} - \boldsymbol{\mu}).$$

该预测器对已知均值 $\mu(\boldsymbol{s}_0)$ 进行校正，校正量为观测位置处相对均值偏差的加权组合，权重由空间协方差结构确定。当然，已知均值这一假设在实践中并不现实。处理该假设的经典方法是通过统计建模并对均值进行估计。因此，所提出的均值结构会带来一组新的约束，这些约束必须纳入最小化问题（Gaetan等人2010）。例如，当我们假设均值为常数且未知时，该预测器称为普通克里金；而泛克里金则假设均值是已知协变量与未知系数的线性组合。

假设均值函数遵循线性结构， $\mu(\boldsymbol{s}) = \boldsymbol{x}^\top(\boldsymbol{s})\boldsymbol{\beta}$，其中 $\boldsymbol{x}(\boldsymbol{s}) = (x_1(\boldsymbol{s}) \ldots, x_p(\boldsymbol{s}))^\top \in \mathbb{R}^p$ 包含协变量函数，且

$\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top \in \mathbb{R}^p$ 包含相应的系数。无偏性约束变为 $\mathbb{E}[Z(\boldsymbol{s}_0)] = \boldsymbol{x}_0^\top \boldsymbol{\beta}$ 和 $\mathbb{E}[P(\boldsymbol{Z}, \boldsymbol{s}_0)] = \boldsymbol{\lambda}^\top \boldsymbol{X}\boldsymbol{\beta}$，其中 $\boldsymbol{x}_0 = \boldsymbol{x}(\boldsymbol{s}_0)$，而 $\boldsymbol{X}$ 是设计矩阵，其元素为 $X_{ij} = x_j(\boldsymbol{s}_i)$，适用于 $i = 1, \ldots, N$ 以及 $j = 1, \ldots, p$。等价地，$\boldsymbol{X}$ 的第 i 行为 $\boldsymbol{x}^\top(\boldsymbol{s}_i)$。通常，对于所有 $\boldsymbol{s}$，$x_1(\boldsymbol{s}) = 1$，因此无偏性条件变为 $\widehat{\lambda}_0 = 0$ 和 $\boldsymbol{X}^\top \boldsymbol{\lambda} = \boldsymbol{x}_0$，也就是说，该条件等价于对参数 $\sum_{i=1}^N \lambda_i x_j(\boldsymbol{s}_i) = x_j(\boldsymbol{s}_0)$ 施加 p 个线性约束，适用于 $j = 1, \ldots, p$。在这种情况下，式（2）的解为

$$\widehat{\boldsymbol{\lambda}} = \Sigma^{-1}\left[\boldsymbol{c}_0 + \boldsymbol{X}(\boldsymbol{X}^\top \Sigma^{-1} \boldsymbol{X})^{-1}(\boldsymbol{x}_0 - \boldsymbol{X}^\top \Sigma^{-1} \boldsymbol{c}_0)\right],$$

从而得到泛克里金估计量，

$$\widehat{P}_{\mathrm{uk}}(\boldsymbol{Z}, \boldsymbol{s}_0) = \boldsymbol{x}_0^\top \widehat{\boldsymbol{\beta}} + \boldsymbol{c}_0^\top \Sigma^{-1}(\boldsymbol{Z} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}), \quad \text{where} \quad \widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \Sigma^{-1} \boldsymbol{X})^{-1} \boldsymbol{X}^\top \Sigma^{-1} \boldsymbol{Z}.$$

注意，若 $x_1(\boldsymbol{s}) = 1$，则第一个约束为 $\sum_{i=1}^N \lambda_i = 1$。此外，若这是唯一的协变量，即 $p = 1$ 且 $x_1(\boldsymbol{s}) = 1$，则该预测器称为普通克里金，最优估计量的解为

$$\widehat{P}_{\mathrm{ok}}(\boldsymbol{Z}, \boldsymbol{s}_0) = \left[\boldsymbol{c}_0^\top + \frac{1 - \boldsymbol{c}_0^\top \Sigma^{-1} \mathbf{1}_N}{\mathbf{1}_N^\top \Sigma^{-1} \mathbf{1}_N} \mathbf{1}_N^\top\right] \Sigma^{-1} \boldsymbol{Z}.$$

所得预测器还可以计算预测方差，该方差取决于所施加的约束类型。简单克里金方差为

$$\begin{aligned} \sigma^2_{\mathrm{sk}}(\boldsymbol{s}_0) &= \mathrm{MSE}(\widehat{P}_{\mathrm{sk}}(\boldsymbol{Z}, \boldsymbol{s}_0)) \\ &= \sigma_0^2 - \boldsymbol{c}_0^\top \Sigma^{-1} \boldsymbol{c}_0, \end{aligned}$$

泛克里金方差为

$$\begin{aligned} \sigma^2_{\mathrm{uk}}(\boldsymbol{s}_0) &= \mathrm{MSE}(\widehat{P}_{\mathrm{uk}}(\boldsymbol{Z}, \boldsymbol{s}_0)) \\ &= \sigma_0^2 - \boldsymbol{c}_0^\top \Sigma^{-1} \boldsymbol{c}_0 + (\boldsymbol{x}_0 - \boldsymbol{X}^\top \Sigma^{-1} \boldsymbol{c}_0)^\top (\boldsymbol{X}^\top \Sigma^{-1} \boldsymbol{X})^{-1} (\boldsymbol{x}_0 - \boldsymbol{X}^\top \Sigma^{-1} \boldsymbol{c}_0). \end{aligned}$$

2.1 与惩罚回归的联系

促使我们采用惩罚克里金方法的一个关键见解是：当协方差呈现块金效应时，惩罚实际上已经隐含在标准克里金中。这种联系在地统计学文献中尚未得到广泛认可，但为我们将 LASSO 扩展到此处提供了有力的理论依据。

考虑带块金效应的克里金，其中 $\mathrm{Var}[\boldsymbol{Z}] = \Sigma_\tau$ 且 $\Sigma_\tau = \Sigma + \tau^2 \boldsymbol{I}$，以及 $\mathrm{Var}[Z(\boldsymbol{s}_0)] = \sigma_0^2 + \tau^2$。值得注意的是，与无块金情形相比，克里金方差保持不变，但问题（2）的目标函数可写为

$$\begin{aligned} \mathrm{MSE}(P(\boldsymbol{Z}, \boldsymbol{s}_0)) &= \boldsymbol{\lambda}^\top \Sigma_\tau \boldsymbol{\lambda} - 2\boldsymbol{\lambda}^\top \boldsymbol{c}_0 + \sigma_0^2 + \tau^2 \\ &= \boldsymbol{\lambda}^\top (\Sigma + \tau^2 \boldsymbol{I}) \boldsymbol{\lambda} - 2\boldsymbol{\lambda}^\top \boldsymbol{c}_0 + \sigma_0^2 + \tau^2 \end{aligned}$$

$$= (\boldsymbol{\lambda}^\top \Sigma \boldsymbol{\lambda} - 2\boldsymbol{\lambda}^\top \boldsymbol{c}_0 + \sigma_0^2) + \tau^2(1 + \|\boldsymbol{\lambda}\|_2^2),$$

其中 $\|\cdot\|_2$ 为欧几里得范数。因此，带块金效应的克里金问题等价于求解如下问题

$$\begin{aligned} \min_{\boldsymbol{\lambda} \in \mathbb{R}^N} \quad & \boldsymbol{\lambda}^\top \Sigma \boldsymbol{\lambda} - 2\boldsymbol{\lambda}^\top \boldsymbol{c}_0 + \sigma_0^2 + \tau^2(1 + \|\boldsymbol{\lambda}\|_2^2) \\ \text{s.t.} \quad & \mathbb{E}[Z(\boldsymbol{s}_0)] = \mathbb{E}[P(\boldsymbol{Z}, \boldsymbol{s}_0)] \end{aligned} \tag{3}$$

问题（3）是问题（2）的一个惩罚版本，带有岭（ℓ_2）惩罚，其中惩罚参数 τ^2 通过最大似然或加权最小二乘从数据中估计得出。这表明，只要存在块金效应，岭惩罚就已隐含在标准地统计学实践中。

2.2 局部克里金

计算 $\widehat{\boldsymbol{\lambda}}$（2）需要 $O(N^3)$ 次运算来求逆 $N \times N$ 协方差矩阵 Σ，对于高 N 而言，这在计算上难以承受，并可能导致数值不稳定性。局部克里金通过仅使用可用数据中的一部分 K观测值，并对该子集应用相同的优化问题（2）来解决这一问题 $\boldsymbol{Z}_K$。虽然选择这些 K 观测值有多种选择准则，但最常见的方法是选取预测位置的 K 最近邻 $\boldsymbol{s}_0$。不失一般性，我们按观测位置到预测位置的欧几里得距离递增排序 $\boldsymbol{s}_0$，因此 K 最近邻对应于排序后位置中的前 K 个位置。由此得到线性预测量

$$P(\boldsymbol{Z}, \boldsymbol{s}_0, K) = \lambda_0 + \sum_{i=1}^{K} \lambda_i Z(\boldsymbol{s}_{(i)}) = \lambda_0 + \boldsymbol{\lambda}^\top \boldsymbol{Z}_K,$$

其中，$\boldsymbol{s}_{(i)}$ 表示所选子集中的第 i 个位置。通过定义 $\boldsymbol{Z}_K = (Z(\boldsymbol{s}_{(1)}), \ldots, Z(\boldsymbol{s}_{(K)}))^\top$、$\mathrm{Var}[\boldsymbol{Z}_K] = \Sigma_K$ 和 $\boldsymbol{c}_{0K} = \mathrm{Cov}(\boldsymbol{Z}_K, Z(\boldsymbol{s}_0))$，局部克里金权重可通过求解以下问题获得

$$\begin{aligned} \min_{\boldsymbol{\lambda} \in \mathbb{R}^K} \quad & \boldsymbol{\lambda}^\top \Sigma_K \boldsymbol{\lambda} - 2\boldsymbol{\lambda}^\top \boldsymbol{c}_{0K} + \sigma_0^2 \\ \text{s.t.} \quad & \mathbb{E}[Z(\boldsymbol{s}_0)] = \mathbb{E}[P(\boldsymbol{Z}, \boldsymbol{s}_0, K)] \end{aligned} \tag{4}$$

局部克里金仅使用 K 个 N 可用观测值来构建稀疏预测量。其主要局限在于，必须事先选择 K，而不考虑数据中的空间相关结构或信息冗余。这促使我们引入 LASSO 型惩罚，以获得数据驱动的稀疏解，使稀疏度由惩罚参数自动确定，而不是由用户固定。

3 惩罚克里金方程

现在，我们介绍主要的方法学贡献：一种通过稀疏正则化自动选择邻居的 LASSO惩罚克里金框架。回顾第 2 节，我们按照观测位置到预测位置的欧几里得距离递增顺序对观测位置进行了排序 $\boldsymbol{s}_0$。这种排序确保前 p 个位置是最近邻，并且由于我们将在下文引入的惩罚结构，这些位置将在解中得以保留。

当我们认识到，对于给定数量的邻居 K，求解局部克里金问题（4）等价于

$$\begin{aligned} \min_{\boldsymbol{\lambda}\in\mathbb{R}^N} \quad & \boldsymbol{\lambda}^\top \Sigma \boldsymbol{\lambda} - 2\boldsymbol{\lambda}^\top \boldsymbol{c}_0 + \sigma_0^2 \\ \text{s.t.} \quad & \mathbb{E}[Z(\boldsymbol{s}_0)] = \mathbb{E}[P(\boldsymbol{Z}, \boldsymbol{s}_0)] \\ & \boldsymbol{\lambda}_{-K} = \mathbf{0}_{N-K} \end{aligned} \tag{5}$$

其中 $\mathbf{0}_m$ 是 m维零向量， $\boldsymbol{\lambda}_{-K} = [\lambda_{K+1}, \ldots, \lambda_N]^\top$ 是最后（$N-K$）个系数构成的向量。限制条件 $\boldsymbol{\lambda}_{-K} = \mathbf{0}_{N-K}$ 意味着，方程$\widehat{\quad}$ $\widehat{\quad}$ $\widehat{\quad}$（5）的解具有 $\boldsymbol{\lambda}_N = (\boldsymbol{\lambda}_K,$ $\mathbf{0}_{N-K}$）的形式，其中 $\boldsymbol{\lambda}_K$ 是方程（4）的解。

根据 Emery (2009)， K 最近邻并不一定是在所有具有相同观测值数量的可能选择中，使克里金方差最小的最优选择。为了实现稀疏性，同时保留克里金预测器的统计性质，我们在克里金方程中加入 LASSO 惩罚项。一个自然的初步方法是对所有权重添加 ℓ_1 惩罚项：

$$\begin{aligned} \min_{\boldsymbol{\lambda}\in\mathbb{R}^N} & \boldsymbol{\lambda}^\top \Sigma \boldsymbol{\lambda} - 2\boldsymbol{\lambda}^\top \boldsymbol{c}_0 + \sigma_0^2 + \eta\|\boldsymbol{\lambda}\|_1 \\ \text{s.t.} & \mathbb{E}[Z(\boldsymbol{s}_0)] = \mathbb{E}[P(\boldsymbol{Z}, \boldsymbol{s}_0)] \end{aligned} \tag{6}$$

其中 η 为非负常数，且 $\|\cdot\|_1$ 为 ℓ_1范数。问题（6）等价于约束LASSO（Gaines等（2018））。然而，该表述存在一个关键局限。解的存在性与唯一性强烈依赖于惩罚参数 η 与由无偏性约束所定义的可行域之间的相互作用。对于 η的某些取值，该约束优化问题可能无解或有多个解，尤其是当约束集与 ℓ_1 球的边界相交于不可微点时。这促使我们对问题（6）加以修正。

为确保对所有 $\eta \geq 0$都存在唯一解，我们修改惩罚结构。关键在于，无偏性约束使我们能够将克里金权重中的 p表示为其余权重的函数。回顾第2节，在 $\mu(\boldsymbol{s}) = \boldsymbol{x}^\top(\boldsymbol{s})$ $\boldsymbol{\beta}$ 具有 p协变量的线性均值结构下，无偏性约束变为 $\boldsymbol{X}^\top\boldsymbol{\lambda} = \boldsymbol{x}_0$。我们将设计矩阵分块为 $\boldsymbol{X} = [\boldsymbol{X}_p^\top, \boldsymbol{X}_{-p}^\top]^\top$，其中 $\boldsymbol{X}_p \in \mathbb{R}^{p\times p}$ 包含前 p 行（对应于 p 个最近邻），并假设其可逆；而 $\boldsymbol{X}_{-p} \in \mathbb{R}^{(N-p)\times p}$ 包含其余行。我们可以通过以下关系分离出 p中的 $\boldsymbol{\lambda}$ 个分量：

$$\boldsymbol{\lambda}_p = \boldsymbol{X}_p^{-1}(\boldsymbol{x}_0 - \boldsymbol{X}_{-p}\boldsymbol{\lambda}_{-p}),$$

其中 $\boldsymbol{\lambda}_p = [\lambda_1, \ldots, \lambda_p]^\top$ 和 $\boldsymbol{\lambda}_{-p} = [\lambda_{p+1}, \ldots, \lambda_N]^\top$ 构成 $\boldsymbol{\lambda}$的一个划分。仅对 $\boldsymbol{\lambda}_{-p}$施加惩罚，可确保前 p 个邻居（根据第2节中的排序假设，它们是最近邻）始终获得非零权重，而距离较远的邻居则可能被收缩至零。这便得到我们提出的问题：

$$\begin{aligned} \min_{\boldsymbol{\lambda} \in \mathbb{R}^N} \; & \boldsymbol{\lambda}^\top \Sigma \boldsymbol{\lambda} - 2\boldsymbol{\lambda}^\top \boldsymbol{c}_0 + \sigma_0^2 + \eta \|\boldsymbol{\lambda}_{-p}\|_1 \\ \text{s.t.} \quad & \boldsymbol{X}\boldsymbol{\lambda} = \boldsymbol{x}_0 \end{aligned} \tag{7}$$

对于固定的正 η。注意，我们只需求解 $\boldsymbol{\lambda}_{-p}$，从而降低问题的维数，并确保对所有 $\eta \geq 0$都存在解。^令 $\boldsymbol{\lambda}(\eta)$ 为固定7的解，则稀疏克里金预测器为 η

$$\widehat{P}_\eta(\boldsymbol{Z}, \boldsymbol{s}_0) = \widehat{\boldsymbol{\lambda}}(\eta)^\top \boldsymbol{Z}.$$

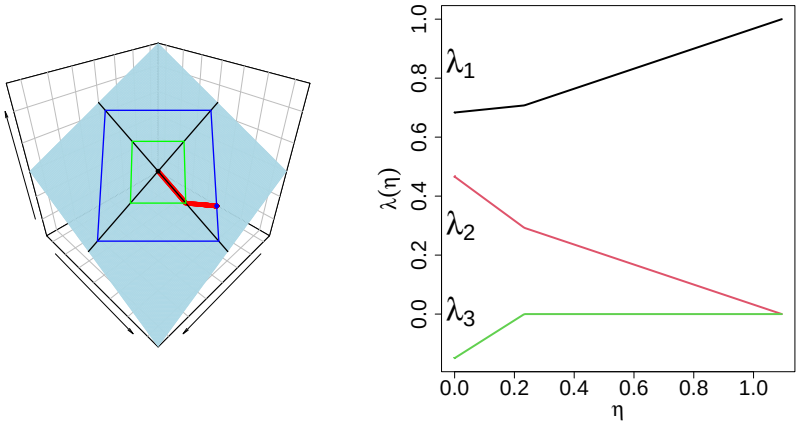
^为了解解相对于 η的变化规律，注意到 $\boldsymbol{\lambda}(0)$与式（2）中的标准克里金预测器一致；而当 η^足够大时，存在某个值 M ，使解向量变得完全稀疏：对所有 $\eta > M$，均有 $\boldsymbol{\lambda}(\eta) = [\boldsymbol{X}_p^{-1}\boldsymbol{x}_0, \boldsymbol{0}_{N-p}]^\top$。这意味着 $\widehat{\boldsymbol{\lambda}}(\eta)$在克里金预测器与仅使用 p个最近邻居的预测器之间建立了连续的联系。参数 η 控制稀疏性与精度之间的权衡：较小的 η 会产生接近标准克里金的解（使用较多邻居），而较大的 η 则会产生高度稀疏的^解（使用较少邻居）。解路径 $\boldsymbol{\lambda}(\eta)$关于 η呈分段线性，其折点出现在邻居进入或离开活动集时。

对于普通克里金情形，我们有 $\boldsymbol{X}_p = 1$ 和 $\boldsymbol{x}_0 = 1$，因此 $\boldsymbol{X}_p^{-1}\boldsymbol{x}_0 = 1$。通过将第一个权重单独移项，可以将 (7) 中的约束写为 $\lambda_1 = 1 - \sum_{i=2}^N \lambda_i$。因此，$\widehat{\boldsymbol{\lambda}}(M) = [1, \boldsymbol{0}_{N-1}]^\top$，这意味着 $\widehat{P}_M(\boldsymbol{Z}, \boldsymbol{s}_0)$ 就是最近邻预测器。

我们通过一个简单示例说明这一解路径行为。假设我们希望在未观测位置 $\boldsymbol{s}_0 = (0.4, 0)$ 处进行预测，已知观测值位于 $\boldsymbol{s}_1 = (0, 0)$、$\boldsymbol{s}_2 = (1, 0)$ 和 $\boldsymbol{s}_3 = (2, 0)$。假设未知常数均值（普通克里金）和指数协方差 $C(\|\boldsymbol{s}_i - \boldsymbol{s}_j\|) = \exp(-\|\boldsymbol{s}_i - \boldsymbol{s}_j\|/0.75)$，则该解必须满足 $\lambda_1 + \lambda_2 + \lambda_3 = 1$，等价于 $\lambda_1 = 1 - \lambda_2 - \lambda_3$。图 1(a) 给出了该问题的几何解释，其中可以看到约束集和惩罚项。图 1(b) 描绘了由 $\boldsymbol{\lambda}(\eta)$ 生成的解路径。

我们框架的一个重要特征是，它能够自然地处理协方差结构中的块金效应。事实上，利用第2.1节中所采用的相同论证，带有块金效应的 τ^2 问题（

$$\begin{aligned} \min_{\boldsymbol{\lambda} \in \mathbb{R}^N} \; & \boldsymbol{\lambda}^\top \Sigma \boldsymbol{\lambda} - 2\boldsymbol{\lambda}^\top \boldsymbol{c}_0 + \sigma_0^2 + \eta \|\boldsymbol{\lambda}_{-p}\|_1 + \tau^2(1 + \|\boldsymbol{\lambda}_p\|_2^2 + \|\boldsymbol{\lambda}_{-p}\|_2^2) \\ \text{s.t.} \quad & \boldsymbol{X}\boldsymbol{\lambda} = \boldsymbol{x}_0. \end{aligned} \tag{8}$$



(a) LASSO曲面问题　　(b) 轨迹图

图 1 左：曲面 $\lambda_1 = 1 - \lambda_2 - \lambda_3$ 由问题 7 在均值未知的假设下（普通克里金设置）得到。绿色和蓝色曲线分别是 LASSO 约束 $|\boldsymbol{\lambda}_{-1}| = 0.291$ （两个位置的普通克里金）和 $|\boldsymbol{\lambda}_{-1}| = 0.615$ （精确普通克里金解）与约束平面的交线。红色曲线是解路径 $\boldsymbol{\lambda}(\eta)$，连接解 $\boldsymbol{\lambda}(M) = [1, 0, 0]^\top$ 与普通克里金解 $\widehat{\boldsymbol{\lambda}}(0) = [0.683, 0.466, -0.149]^\top$。右：$\widehat{}$根据解路径 $\boldsymbol{\lambda}(\eta$绘制的轨迹图，展示该方法如何将最远位置的权重设为零。

由于 $\boldsymbol{\lambda}_p = \boldsymbol{X}_p^{-1}(\boldsymbol{x}_0 - \boldsymbol{X}_{-p}\boldsymbol{\lambda}_{-p})$，问题（8）对应于加权弹性网络问题（Zou 和 Hastie 2005）。对于问题（7）和（8），选择 η 对于平衡稀疏性与统计效率至关重要。第4节将对此进行讨论。

3.1 自适应 LASSO 表述

虽然式（7）中的标准 LASSO 惩罚项能够自动进行邻居选择，但它对所有远距离邻居一视同仁，不考虑它们与预测位置之间的空间关系。实际上，由于空间相关性模式不同，一些远距离邻居可能比其他邻居更有信息量。例如，在空间相关性较强的方向上，距离适中的邻居可能比在相关性较弱方向上距离更近的邻居更有价值。这促使我们采用一种能够考虑空间协方差结构的自适应加权方案。

问题（7）对所有权重 $\eta|\lambda_i|$ λ_i $i > p$采用统一的惩罚，因此对所有远距离邻居一视同仁。这忽略了它们对于在 $\boldsymbol{s}_0$ 处进行预测的相对重要性，因为对预测贡献较大的邻居应比贡献较小的邻居受到更轻的惩罚。遵循 Zou (2006）的做法，我们提出一种两阶段程序，第二阶段使用基于初始估计的数据驱动权重。

自适应LASSO采用两阶段程序。第一阶段，我们使用所有观测值，通过标准克里金获得˜初始估计 $\boldsymbol{\lambda}$ 。第二˜阶段，我们根据 λ_i的大小定义自适应权重：

$$w_i = \frac{1}{|\tilde{\lambda}_i|}, \quad i > p. \tag{9}$$

˜较大 $|\lambda_i|$ 的邻居（表明其对预测的贡献较大）会获得较小的˜权重 w_i （惩罚较轻），而较小 $|\lambda_i|$ 的邻居则会获得较大的权重（收缩力度更大）。自适应LASSO克里金权重通过求解以下问题获得：

$$\begin{aligned} \min_{\boldsymbol{\lambda}\in\mathbb{R}^N} \quad & \boldsymbol{\lambda}^\top \Sigma \boldsymbol{\lambda} - 2\boldsymbol{\lambda}^\top \boldsymbol{c}_0 + \sigma_0^2 + \eta \sum_{i>p} w_i |\lambda_i| \\ \text{s.t.} \quad & \boldsymbol{X}\boldsymbol{\lambda} = \boldsymbol{x}_0. \end{aligned} \tag{10}$$

将参数设为 $w_i = 1$ （对所有 i ）即可恢复标准最小绝对收缩和选择算子问题（7）。

问题（10）采用带加权˜软阈值处理的改进梯度下降方法求解，详见附录C （算法2）。初始估计 $\boldsymbol{\lambda}$ 也为梯度下降优化提供了极佳起点，与冷启动初始化相比，梯度下降迭代次数减少了40%–50%。当初始估计使用全部 N个观测值时，总复杂度为 O（N^3+TN），其中 T 为梯度下降迭代次数（通常为 $T < 100$）。当阶段 1 采用具有 K个邻居的局部克里金时，复杂度降至 O（$K^3 + TN$）。对于满足 $K \ll N$的大规模数据集，与 O（N^3）所对应的标准克里金相比，可显著节省计算量。在克里金背景下，自适应LASSO能够识别实现准确预测所需的最小邻域，同时使预测方差保持接近标准克里金，如第6节所示。

4 调优参数的选择

我们现在着手解决选择惩罚参数 η这一关键问题。该参数控制着预测精度与邻域稀疏性之间的权衡：较小的值会产生类似于标准克里金的稠密解，而较大的值则会产生高度稀疏的解，从而可能牺牲预测质量。

对于标准回归设定中的独立数据，调优参数的选择通常依赖于交叉验证或信息准则，例如赤池信息准则或贝叶斯信息准则（Tibshirani 1996; Hastie等人，2009; Friedman等人，2010）。然而，这些标准方法在克里金背景下面临两个根本性挑战。首先，它们计算代价高昂。与为整个数据集选择单一惩罚参数的标准回归不同，我们的惩罚克里金框架要求选择惩罚参数 η ，以便在每个预测位置适应局部空间相关结构。对于拥有 M 个预测位置的数据集，若要评估 G个候选η 值（每个位置），则需要求解 $O(MG)$个惩罚克里金问题，每个问题涉及代价为 $O(N^3)$（标准克里金）或 O (K^3)（当使用K 个邻居时）的矩阵运算，从而使交叉验证代价高得难以承受。

其次，更根本的是，克里金方差不取决于观测值，而取决于观测值的空间配置和协方差结构（Heuvelink 等人，2002）。这与标准回归中的预测误差不同。因此，数据拆分方法对于了解预测

不确定性的信息较少，因为留出位置的克里金方差仅取决于该位置与训练位置之间的空间关系，而非实际观测值。因此，交叉验证可能会选择能使经验预测误差最小化的 η，却未能恰当考虑空间相关结构和信息冗余。这些局限表明，我们需要一种能够明确考虑空间相关性且计算高效的不同方法。

4.1 有效样本量与信息冗余

我们这一方法的核心洞见在于，呈正空间相关的观测值携带冗余信息。在完全相关的极端情形下，由 N 个观测值构成的样本所提供的信息并不多于单个观测值。相反，在独立条件下，全部 N 个观测值均贡献独特信息。对于中等程度的空间相关性，有效信息量介于 1 与 N 之间。

这一概念通过有效样本量（ESS）加以形式化，它量化了与给定空间相关样本等价的独立观测数量（Cressie 1993a，第 15 页）。在克里金背景下，ESS 为邻居选择提供了自然准则，因为我们寻求的稀疏度（由 η 控制）应去除冗余邻居，同时保留准确预测所需的有效信息量。

Cressie （1993a）通过比较样本平均 Z 在不相关观测值与空间相依观测值下的方差，推导出有效样本量的表达式，得到：

$$n^{\mathrm{ESS}} = \frac{N^2}{\mathbf{1}_N^\top \boldsymbol{R} \mathbf{1}_N} = \frac{N^2}{\sum_{i=1}^{N}\sum_{j=1}^{N} \rho(\boldsymbol{h}_{ij})}, \quad \boldsymbol{h}_{ij} = \boldsymbol{s}_i - \boldsymbol{s}_j, \tag{11}$$

其中 $\rho(\cdot)$ 是有效的相关函数，$\boldsymbol{R}$ 是元素为 $\boldsymbol{R}_{ij} := \rho(\boldsymbol{h}_{ij})$ 的相关矩阵，$\mathbf{1}_N$ 是一个大小为 N 的全 1 列向量。

例如，考虑等相关结构，其中 $\mathrm{cov}(Z_i, Z_j) = \sigma^2\rho$ 对于 $i \neq j$，且 $\sigma^2 = \mathrm{var}(Z_i)$；并且 $\rho > 0$。那么 $n^{\mathrm{ESS}} = N/(1 + (N-1)\rho)$，因此，如果 $\rho = 0$，那么 $n^{\mathrm{ESS}} = N$（独立性保留完整样本量）；如果 $\rho = 1$，那么 $n^{\mathrm{ESS}} = 1$（完全相关不会带来额外信息）。这种极端情况凸显了空间相关性对有效信息的影响：强正相关会大幅降低有效样本量，而独立性则能保留完整样本量。对于现实中的空间过程，有效样本量通常介于这两个极端之间，具体取决于空间相关性的变程和强度。这些性质不仅适用于等相关结构，也适用于任何有效的相关结构（更多有效样本量性质，参见 Acosta 等人 2018）。

对于邻居选择，我们关注互补量 $r^{\mathrm{ESS}} = N - n^{\mathrm{ESS}}$，它衡量大小为 N 的样本中的冗余信息。这种冗余量化了在不显著降低有效信息量的情况下可以移除多少个观测值。请注意，当观测值相互独立时，$r^{\mathrm{ESS}} = 0$

（没有冗余）；而当观测值完全相关时，$r^{\mathrm{ESS}} = N - 1$（冗余最大）。一般而言，$0 \le r^{\mathrm{ESS}} \le N - 1$。

屏蔽效应（Stein2002）为克里金中冗余为何重要提供了直观解释：当观测值高度相关时，邻近的邻居会提供重叠的信息，导致克里金权重不稳定地围绕零振荡，尽管满足无偏约束。LASSO 惩罚项通过明确移除冗余邻居来解决这一问题。

4.2 局部克里金与惩罚克里金的有效样本量

回顾一下，全局克里金使用全部 N 个空间位置，而局部克里金使用 $K \le N$个空间位置。设 $\boldsymbol{e}_{N,\,K}$ 为一个 N维二元向量，其中第 i个元素在位置 $\boldsymbol{s}_i$ 用于局部克里金时取 1，否则取 0，则$\|\boldsymbol{e}_{N,\,K}\|_0 = \sum_{i=1}^{N} \mathbb{I}(e_{N,\,K,\,i} \neq 0) = K$，其中 $\mathbb{I}(\cdot)$ 是示性函数，$\|\cdot\|_0$ 是 ℓ_0范数。约束$\|\boldsymbol{e}_{N,\,K}\|_0 = K$ 确保恰好选取 K 个邻居。该子集的有效样本量为

$$n_{\mathrm{loc}}^{\mathrm{ESS}} = \frac{K^2}{\boldsymbol{e}_{N,K}^{\top} \boldsymbol{R} \boldsymbol{e}_{N,K}},$$

其中考虑了选定邻域内的相关结构，而冗余为 $r_{\mathrm{loc}}^{\mathrm{ESS}} = K - n_{\mathrm{loc}}^{\mathrm{ESS}}$。

现在，回顾一下，$\widehat{P}_\eta(\boldsymbol{Z}, \boldsymbol{s}_0) = \widehat{\boldsymbol{\lambda}}(\eta)^{\top} \boldsymbol{Z}$ 是通过求解（7）得到的预测器。$\widehat{\boldsymbol{\lambda}}(\eta)$ 的稀疏性取决于 η，因此选定位置的数量及其有效样本量也取决于此。我们定义二元指示向量 $\boldsymbol{e}_{N,\widehat{\eta}}$，其中 $e_{N,\eta,i} = \mathbb{I}(\lambda_i(\eta) \neq 0)$，用于标识惩罚水平$\eta$下的活动集邻居。作为 η 的函数，有效样本量随后为

$$n_{\eta}^{\mathrm{ESS}} = \frac{\|\widehat{\boldsymbol{\lambda}}(\eta)\|_0^2}{\boldsymbol{e}_{N,\eta}^{\top} \boldsymbol{R} \boldsymbol{e}_{N,\eta}},$$

并且冗余为 $r_{\eta}^{\mathrm{ESS}} = \|\widehat{\boldsymbol{\lambda}}(\eta)\|_0 - n_{\eta}^{\mathrm{ESS}}$。注意，随着 η 增大，$\|\widehat{\boldsymbol{\lambda}}(\eta)\|_0$ 减小（邻居更少），但 η 与 n_{η}^{ESS} 之间的关系更加复杂：移除高度相关的邻居可能会增加每个观测值的有效样本量，而移除具有信息量的邻居则会降低总体信息量。

由于信息和冗余都取决于相关结构，因此比较不同 η 值需要进行归一化。我们提出了归一化冗余度量

$$s_\eta = \frac{r_{\eta}^{\mathrm{ESS}}}{r_{0}^{\mathrm{ESS}}} = \frac{\|\widehat{\boldsymbol{\lambda}}(\eta)\|_0 - n_{\eta}^{\mathrm{ESS}}}{N - n_{0}^{\mathrm{ESS}}},$$

它表示当前冗余占全局克里金解中总冗余的比例（$\eta = 0$）。这种归一化确保 $0 \le s_\eta \le 1$，其中 $s_\eta =$

1 对应全局克里金（包含全部冗余），而 $s_\eta \to 0$ 则表示 $\eta \to \infty$（通过稀疏化移除冗余）。

重要的是，s_η 不一定会随 η单调递减。尽管邻居数量 $\|\widehat{\boldsymbol{\lambda}}(\eta)\|_0$ 会单调减少，但移除高度冗余的邻居后，有效样本量 n_η^{ESS}可能增加，导致s_η 呈现非单调变化，并在邻居进入或离开活动集时对应的 η 取值处出现不连续跳变。

4.3 η的最优选择

单凭s_η 无法确定 η 的最优值，除非固定某个特定量——例如 $s_\eta = 0.5$，它表示 50% 的相对冗余度。要确定η，我们需要一种能够量化拟合质量的度量。正如第2节所讨论的，预测方差 $\sigma_\eta^2 := \mathrm{MSE}(\widehat{P}_\eta(\boldsymbol{Z},\ \boldsymbol{s}_0))$由克里金方程得出。该方差用作衡量统计性能的指标，并与全局克里金方差 $\sigma_{\mathrm{gk}}^2 := \sigma_{\mathrm{gk}}^2(\boldsymbol{s}_0)$进行比较。随着 η 增大，参与预测的邻居减少，预测方差 σ_η^2 单调增大。当 $\eta \to 0$ 时，我们有 $\sigma_\eta^2 \to \sigma_{\mathrm{gk}}^2$（全局克里金达到最小方差）；而当 $\eta \to \infty$时，我们有 $\sigma_\eta^2 \to \sigma_\infty^2$，其中 σ_∞^2 对应于仅使用 p 个未经惩罚的最近邻时的方差。为了使方差项与稀疏度度量具有可比性，我们将其归一化为

$$v_\eta = \frac{\sigma_\eta^2 - \sigma_{\mathrm{gk}}^2}{\sigma_\infty^2 - \sigma_{\mathrm{gk}}^2},$$

其范围从 $v_\eta = 0$ （全局克里金，最小方差）到 $v_\eta = 1$ （最大稀疏性，最大方差）。

我们的目标是选择 η，使两个相互竞争的目标达到最佳平衡：

- **最大化稀疏性** （较大的 s_η）：移除冗余邻居，以降低计算成本和信息冗余
- **尽量降低方差增幅** （较小的 v_η）：使预测精度接近全局克里金

对s_η 和v_η 取简单平均并不恰当，因为这会将两个目标视为同等重要，尤其是在边界情况下：稀疏性高但精度差的解（例如 $s_\eta = 0.9$、$v_\eta = 0.1$）与稀疏性低但精度极佳的解（例如 $s_\eta = 0.1$、$v_\eta = 0.9$）得分会相同。相反，我们采用调和平均数，它会对不平衡情况施以较大惩罚。只有当两项指标都相对较大且较为均衡时，调和平均数才会达到最大值。

为说明这一点，请注意，对于固定的和 $s_\eta + v_\eta = c$，当 $s_\eta = v_\eta = c/2$ 时，调和平均数 $\frac{2 s_\eta v_\eta}{s_\eta + v_\eta} = \frac{2 s_\eta v_\eta}{c}$ 达到最大值。当其中一项远小于另一项时（例如 $s_\eta \approx 0$ 或 $v_\eta \approx 0$），无论另一项的值是多少，调和平均数都接近于零。因此，调和平均数尤其适用于我们的多目标优化问题。因此，我们提出通过最大化调和平均数来选择 η：

$$\eta^* = \mathrm{argmax}_{\eta \geq 0} \left\{ \frac{2 s_\eta v_\eta}{s_\eta + v_\eta} \right\}. \tag{12}$$

目标函数 $\frac{2 s_\eta v_\eta}{s_\eta + v_\eta}$ 的取值范围为 0 到 1；当稀疏性和方差损失达到平衡时，目标函数取得最大值。该准则倾向于选择能够消除大量冗余（s_η 较大），同时不会施加过度方差惩罚（v_η 不太大）的方案。

准则（12）计算效率高，只需：
1. 在候选 η 值处计算惩罚克里金目标函数 2. 计算 $n_\eta^{\mathrm{ESS}} = \|\widehat{\boldsymbol{\lambda}}(\eta)\|_0^2 / (\boldsymbol{e}_{N,\,\eta}^\top \boldsymbol{R} \boldsymbol{e}_{N,\,\eta})$，这需要对活动集的相关矩阵 $\boldsymbol{R}$ 中的元素求和 3. 计算克里金方差 σ_η^2，该值由克里金方程得出

所有所需量都是求解惩罚克里金问题的副产品，因此无需额外进行矩阵求逆或模型拟合。这与K折交叉验证形成鲜明对比；后者针对 G 个候选 η 值和 K 折，需要重新拟合 $O(GK)$个模型。

在实践中，我们通过将搜索限制在有界区间 $[\eta_{\min}, \eta_{\max}]$ 内来避免计算 σ_{gk}^2（这需要对完整的 $N \times N$ 矩阵求逆），其中 $\eta_{\min}$ 足够小，使得 $\sigma_{\eta_{\min}}^2 \approx \sigma_{\mathrm{gk}}^2$。通常，我们将 $\eta_{\max}$ 设为普通克里金权重的最大绝对值，以确保至少保留 p 个邻居，并且 $\eta_{\min} = 10^{-9} \times \eta_{\max}$。

最优值 η^* 通过自适应网格搜索算法确定，该算法结合了粗略的初步探索与迭代细化。算法见算法1。该过程通常需要对准则进行 200–300 次评估，准则

算法1 用于选择 η 的两阶段搜索过程

1:在跨越 $[\eta_{\min}, \eta_{\max}]$、包含 κ_0 个点的粗略对数网格上评估准则。2: 找出准则值最高的 5 个候选值。3: **对于**每个候选值 η_k **执行**4: 在区间 $[\eta_k/2, 2\eta_k]$ 内构建一个包含 κ_1 个点的局部网格。5: 在该局部网格上评估准则。6: **结束循环**7: **返回**使准则达到最大值的 η^*。

这在计算上可忽略不计，与交叉验证所需的矩阵求逆相比尤其如此。步骤 1 中的对数间距确保了对多个数量级的良好覆盖，而步骤 3–4 中的局部细化则能在最优值附近实现精确搜索。

例如，图2 展示了指数型协方差模型在三个实际变程（PR）下的准则表现；实际变程定义为相关性衰减至 0.05 时的距离。可以观察到几个规律：(1) 随着 η 增大，方差比 v_η（红色）单调上升，这符合预期，因为保留的邻居更少；而稀疏度度量 s_η 则呈现相反趋势。(2) 调和

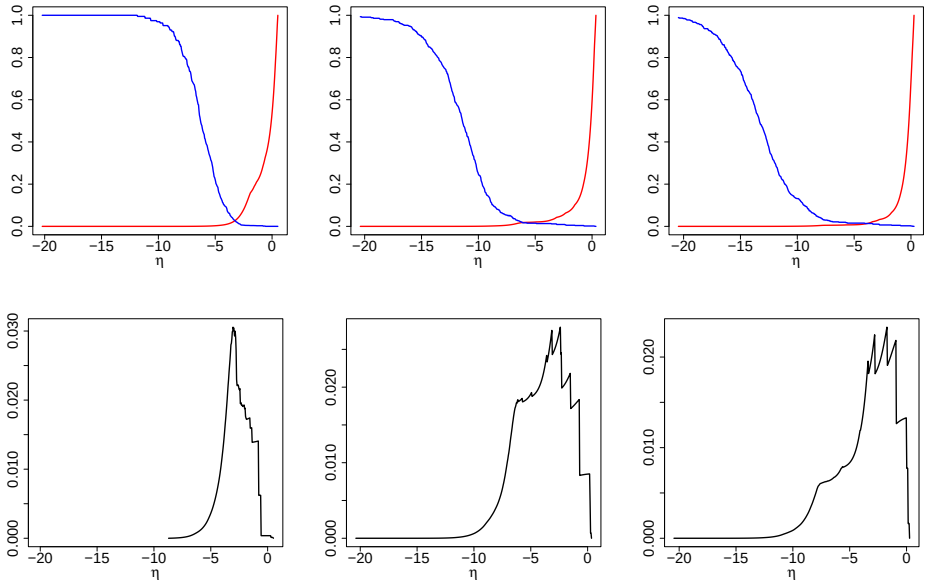


图 2 实际变程 (PR) 为 0.05（左）、0.15（中）和 0.2（右）的指数协方差模型的调优参数选择准则的变化。方差比 v_η（红线）随 η 单调递增，而稀疏度度量 s_η（黑线）呈现非单调变化。调和平均准则（蓝线）在中间值处达到最大值；对于更平滑的过程（更大的 PR），最优 $\eta^\star$ 会增大。

平均数（黑色）呈现非单调变化，每当有邻居从活动集中移除时，都会出现不连续的跃变。(3) 调和平均数在中间的 η 取值处达到最大，避开了两个极端：方差缩减可忽略不计的稠密解，以及效率损失显著的过度稀疏解。(4) 对于 PR 较大的过程（图 2右图），最优 $\eta^\star$ 较高；由于空间相关性增强，表示因而更加稀疏。(5) 对于更粗糙的过程（PR 较小，图 2左图），最优 $\eta^\star$ 较小，因此会保留更多邻居，以充分捕捉局部变异性。总的来说，基于 ESS 的准则能够根据底层相关结构自适应地调整邻域大小，无需用户调参或进行交叉验证。

5 数值实验

在所有数值实验中，我们考虑均匀分布在 $[0,1]\times[0,1]$上的 500 个固定位置。我们在五个具有不同邻域配置的代表性位置进行预测：（FN）最远邻居，定义为所有点中最近邻距离最大的那个位置；（AN）平均邻居，其最近邻距离最接近全局平均值；（DN）最密邻居，位于点密度最高的区域；（SN）位于 (0，0.5) 的侧边邻居；以及（CN）位于 (0，0) 的角点邻居。这些配置如图 3 所示。

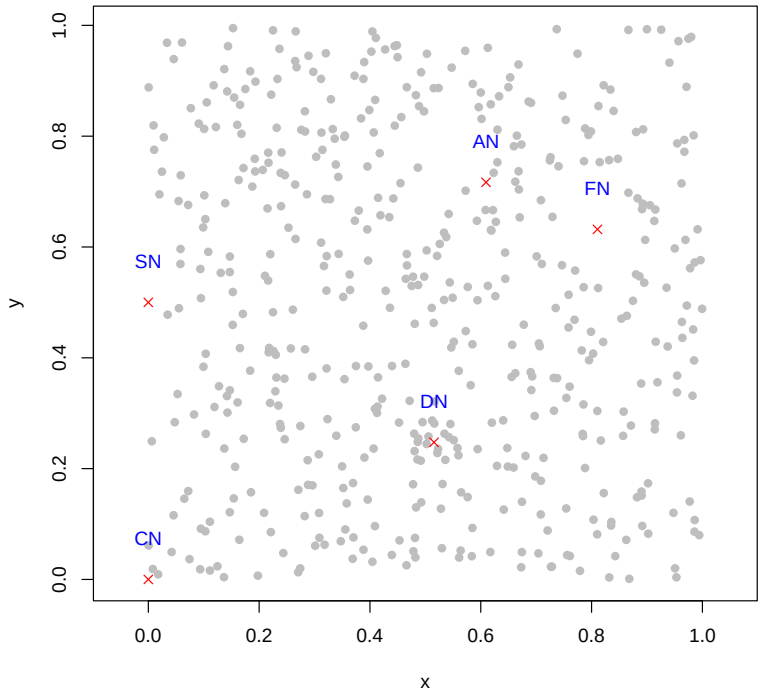
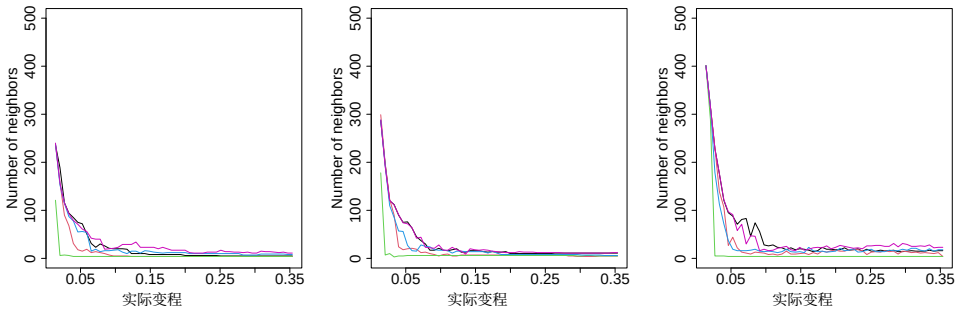


图 3 模拟实验中使用的位置。灰色点作为训练集，红色点则是用于探索惩罚克里金所选邻域结构的预测位置。标签分别表示最远邻居（FN）、平均邻居（AN）、最密邻居（DN）、侧边邻居（SN）和角点邻居（CN）。

底层高斯随机场采用三种协方差模型生成：指数型、球状和 Mat´ern，光滑参数为 $\nu = 1.5$。为确保不同模型之间具有可比性，我们对所有协方差函数进行参数化，使其具有相同的实际变程（PR）；实际变程定义为相关性降至 0.05 时对应的距离。针对每个预测点，我们通过分析若干关键特征，评估第 3 节介绍的惩罚克里金方法，以及第 4 节所述的调节参数选择策略。

图 4 总结了在不同协方差结构和预测位置下，选定的空间位置数量如何随实际变程变化。我们观察到，随着空间相关性增强，算法会系统性地减少邻居数量。例如，在指数模型下，平均邻居数（AN）从 $PR = 0.014$时的 237 个降至 $PR = 0.053$时的 15 个，并在 $PR \geq 0.117$时稳定为 4 个邻居。这意味着邻居数量从所有可用观测值的 47% 降至不足 1%，同时保持了预测精度。这种减少的幅度很大程度上取决于预测点周围的局部构型。对于

例如，由于附近的观测值已能提供充分信息，额外的邻居变得冗余，因此 DN 在较小的 PR 值下便达到最小邻域大小。相比之下，在相近的 PR 水平下，SN 和 CN 需要更大的邻域。总体而言，不同协方差模型之间没有显著差异，这表明邻域大小的减少主要由空间配置和相关性强度决定，而非特定的协方差族。



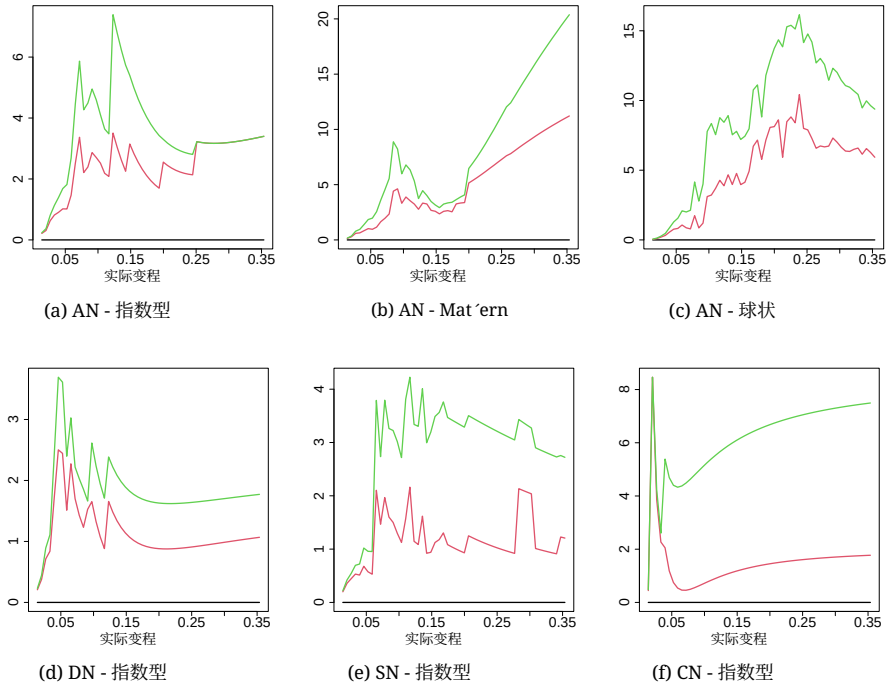
(a) 指数型　(b) Mat´ern ($\nu = 1.5$)　(c) 球状

图4 在三种协方差模型下，惩罚克里金选定的邻居数量随实际变程的变化。每条曲线代表五个预测位置之一（FN：黑线；AN：红线；DN：绿线；CN：蓝线；SN：洋红线）。对于所有模型，随着实际变程增大，选定邻居的数量都会减少，且减少速度取决于过程平滑度和预测位置的几何形状。

图5 比较了惩罚克里金（PK）和局部克里金（LK）在不同实际变程下的相对方差增幅。为确保公平比较，局部克里金采用与惩罚克里金相同的邻域大小，并选取相应的 K 最近邻。对于所有考察的设置，惩罚克里金的表现均不逊于局部克里金。

顶行展示了在预测位置固定为 AN 时，不同协方差模型对方差比较的影响。底行展示了在协方差模型固定为指数型时，预测位置对比较结果的影响。附录A中提供了协方差模型与预测位置全部 15 种组合的完整结果。

图5 中的结果表明，尽管邻域大小大幅缩减，惩罚克里金的方差增幅仍显著较低，并且接近使用所有可用观测值的全局克里金（GK）基准。对于 $PR = 0.053$ 处的指数协方差模型，平均邻居（AN）仅需 15 个邻居（约占数据的 3%），相较于 GK，方差增幅仅为 2.44%。在 $PR = 0.354$ 处，AN 仅依赖 4 个邻居，方差增幅为 1.07 %。对于平滑参数为 $\nu = 1.5$ 的 Mat´ern 模型，也观察到了类似的表现：在 $PR = 0.117$ 处，AN 使用 5 个邻居，方差增幅仅为 0.88%。这些结果表明，所提出的方法在大幅减少预测所涉及观测值数量的同时，保留了全局克里金所利用的大部分有效信息。



(a) AN - 指数型　(b) AN - Mat´ern　(c) AN - 球状

(d) DN - 指数型　(e) SN - 指数型　(f) CN - 指数型

图 5 使用全部 500 个观测值时，相较于全局克里金（GK）的相对方差增幅（%）。红线表示惩罚克里金（PK），绿线表示使用与 PK 相同数量的 K-最近邻的局部克里金（LK），黑色水平线表示全局克里金基线（0%）。上排显示预测位置固定为 AN 时协方差模型的影响；下排显示指数协方差下预测位置的影响。

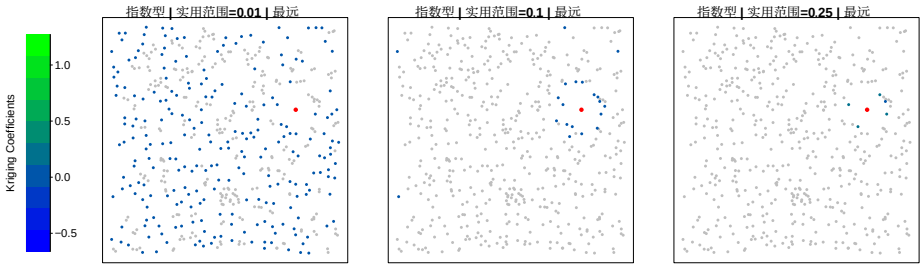
与使用相同数量 K 个最近邻的局部克里金（LK）相比，惩罚克里金在所有配置下都能持续实现更高的预测效率。图 5清楚地展示了这一优势：PK 曲线始终系统性地位于 LK 曲线下方。例如，在 $PR = 0.053$ 处的指数协方差模型中，使用 15 个邻居的惩罚克里金方差增幅为 2.44%，而使用相同邻域大小的 LK 增幅为 3.61%，相对改善幅度为 32%。最密邻居（DN）的差异更加显著：惩罚克里金仅用 4 个邻居，方差增幅便达到 0.75%；而使用相同数量邻居的 LK，增幅为 4.40%。这表明，仅依据距离选择邻居会造成大量冗余，而基于ESS的准则能够识别出信息量最高的观测值。

在所有协方差模型、实际变程和预测位置下，惩罚克里金始终优于局部克里金。虽然多个面板以指数模型作为代表性参照，但 Mat´ern（$\nu = 1.5$）和球状协方差模型也呈现出类似的模式。特别是对于 Mat´ern 模型，在中等和高实际变程下，惩罚克里金通常能取得相当或更大的

相对改进。球状协方差凸显了紧支集的影响：尽管两种方法在较小实际变程下表现良好，但随着变程增大、邻居纳入变得更加离散，惩罚克里金仍保持优势。在不同预测位置中，最密邻居（DN）从所提出的方法中获益最大，因为基于ESS的选择能有效过滤高度聚集区域中的冗余信息；而边界位置（SN 和 CN）则呈现出较小但稳定的提升。

总体而言，这些结果证实，惩罚克里金在稀疏性与预测精度之间实现了有效的折中。该方法利用空间依赖结构，而非仅依赖邻近程度，因此显著优于局部克里金；同时，它仅需一小部分计算成本，便能达到与全局克里金相近的预测性能。通过自动适应相关结构、过程平滑度和局部几何形状，所提出的方法为解决空间预测中的邻居选择问题提供了一种原理可靠且计算高效的方案。

为展示该方法如何进行空间自适应，图6 展示了在三种实际变程水平下，指数协方差条件中最远邻居（FN）位置的选定邻居及其系数大小。空间模式揭示了自适应选择机制的几个关键特征。首先，选定邻居



(a)　(b) 低实用范围（0.014）　(c) 中实际变程（0.104）　(d) 大实用范围（0.251）

图 6 三个实用变程水平下，指数型协方差中最远邻居（FN）位置的空间邻居选择。灰色点表示克里金系数为零的观测值（未选定），彩色点表示选定邻居，颜色深浅表示系数的大小（见图例，蓝色 = 小，绿色 = 大），红点表示预测位置。随着实际变程从左向右增大，选定邻域逐渐变得更加稀疏。

这些观测值构成了一个围绕预测位置、具有空间相干性的集合，而不是随机散布在整个区域中。其次，系数大小（以颜色深浅表示）会随着与预测位置的距离增加而平滑衰减：较近的观测值获得较大的权重（绿色），较远的观测值获得较小的权重（蓝色），直至被完全剔除。这种空间相干性表明，LASSO 惩罚项与基于ESS的准则相结合，能够在遵循潜在空间相关结构的同时，有效识别信息量最大的邻居子集。

然而，这项分析也凸显了所提出的选择方法与最近邻方法的不同之处。尽管该方法明显倾向于选择预测位置周围的观测值，但当更远处的点能够提供非冗余信息时，该方法也可能保留这些信息丰富的点。在所有协方差模型、实际变程和预测位置下，均可观察到这一现象；完整的空间系数可视化见附录 B。

6 数据分析

我们通过两个涵盖不同空间尺度的真实数据应用，展示惩罚克里金方法的实用价值。汝拉数据集（359 个位置）代表一个中等规模问题，在该问题中，全局克里金仍具备计算可行性，因此我们能够将方法的准确性与理论最优结果进行比较。COBE 海表温度数据集（43,799 个位置）代表一个大规模问题，在该问题中，全局克里金的计算成本高到难以承受，从而展示了该方法的可扩展性和实际必要性。总体而言，这些应用说明，惩罚克里金能够自动适应不同的空间尺度、相关结构和预测情景。

6.1 汝拉数据集

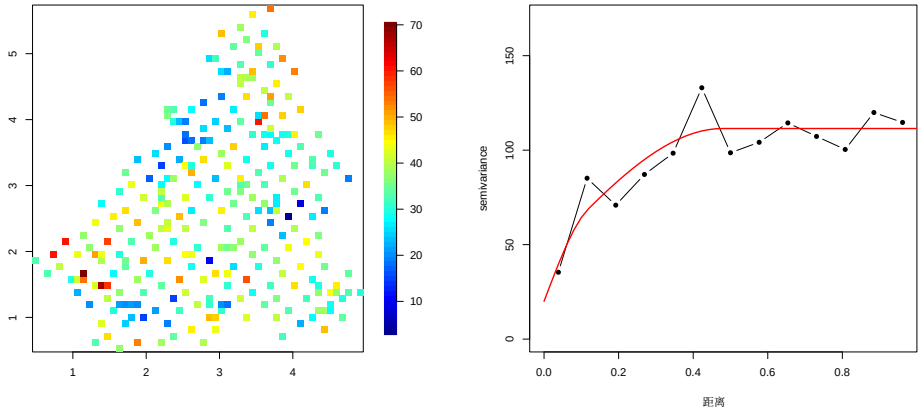
我们使用汝拉数据集（Atteia 等，1994；Webster 等，1994；Goovaerts，1997）评估惩罚克里金方法的性能。该数据集是 R 语言 `compositions` 软件包中的一个经过深入研究的基准数据集（van den Boogaart 等，2025），包含瑞士汝拉地区 359 个空间位置的重金属浓度测量值。我们重点研究铬（Cr）浓度。该变量表现出显著的空间变异，因此需要谨慎进行邻居选择，以实现准确预测；Emery （2009）在最优邻域选择的背景下讨论了这一问题。该数据集规模较小，全局克里金在计算上是可行的，因此可以将我们方法的准确性和效率直接与理论最优结果进行比较。数据如图 7(a)所示。

经验变异函数呈现出两个不同尺度上清晰的空间结构，因此采用嵌套球状协方差模型。

$$C(h) = C_1(h;\theta_1) + C_2(h;\theta_2) + \tau^2 \mathbb{I}_{h=0}, \tag{13}$$

其中，$C_i(h;\ \theta_i)$ 表示部分基台为 σ_i^2、变程为ϕ_i的球状协方差函数，而 τ^2 表示测量误差（块金效应）。第一项捕捉短程变化（局部地质特征），第二项则捕捉区域趋势。模型通过加权最小二乘法拟合（Cressie 1993b），使

$$\sum_{j=1}^{n_b} w_j \left(\hat{\gamma}(h_j) - \gamma(h_j;\boldsymbol{\theta})\right)^2,$$



(a) 汝拉地区的铬观测值　　(b) 经验变异函数和拟合变异函数

图 7 左：汝拉地区的铬观测值；右：估计结果（黑点）和拟合变异函数（红线）

其中，权重 $w_j = n_j$ 根据每个距离分箱中的数据对数量确定。估计参数为 $\hat{}$ $\sigma_1^2 = 50.0$、$\hat{}\phi_1 = 0.15$、$\hat{}\sigma_2^2 = 50.0$、$\hat{}\phi_2 = 0.65$和 $\hat{}\tau^2 = 20.0$，总基台值约为 120 个单位。估计的变异函数如图 7(b)所示。

为了评估空间预测性能，我们在观测位置边界框上构建了一个规则的 50×50 网格。移除落在观测值凸包之外的 1,023 个网格点（以避免过度外推）后，我们得到 1,477 个预测位置，并在这些位置比较三种克里金方法。首先，普通克里金（ GK ）使用全部 359 个观测值，代表理论上的最优方法。其次，惩罚克里金（PK）采用稀疏克里金，并通过自适应LASSO惩罚自动选择邻居。第三，局部克里金（LK）使用与 PK 选出的数量相同的最近邻。

$\hat{}$ 对于 PK，自适应 LASSO 方法采用的惩罚权重为 $w_j = 1/|\lambda_j|\hat{}$ 对于 $j = 2, \ldots, n$，其中 λ_j 为普通克里金权重。最优惩罚参数 η 通过第 4 节所述的自适应网格搜索算法，在范围 $[\eta_{\min}, \eta_{\max}]$内选取，其中 $\eta_{\max} =$ 最大值$_j\, |\hat{}\lambda_j|$ 且 $\eta_{\min} = \eta_{\max} \times 10^{-9}$。网格搜索先采用包含 150 个点的初始粗网格，再在有希望的区域通过迭代细化增加 50 个点。

图 8展示了三种方法的预测图（上排）和克里金方差图（下排）。目视检查表明，预测铬浓度的空间特征高度一致，三种方法均成功捕捉了主要特征：研究区域东北部浓度较高，西南部数值较低；无观测点处方差高，靠近观测位置处方差低。

惩罚克里金所选邻居的数量在空间上有所变化（图 9），在连通良好的内部区域约为 3–5 个邻居，在边界附近以及局部变化复杂的区域则为 10–25 个邻居。这种空间自适应体现了其优势，即能够根据局部几何和相关结构自动调整邻域大小，这与最近邻方法不同。

从方差图中可以看到一些细微差异。首先，普通克里金（图8，左下）在所有空间位置都实现了最低的预测方差（见图9）。其次，惩罚克里金（图8，下方中间）以较少的空间位置再现了普通克里金的方差变化，其方差也小于局部克里金（图8，右下）。当克里金试图在观测空间位置内部进行外推时，差异最大。正如模拟研究所示，惩罚克里金会通过选取更多空间位置来补偿这一效应（见图9）。此外，三种方法的结果总体一致。

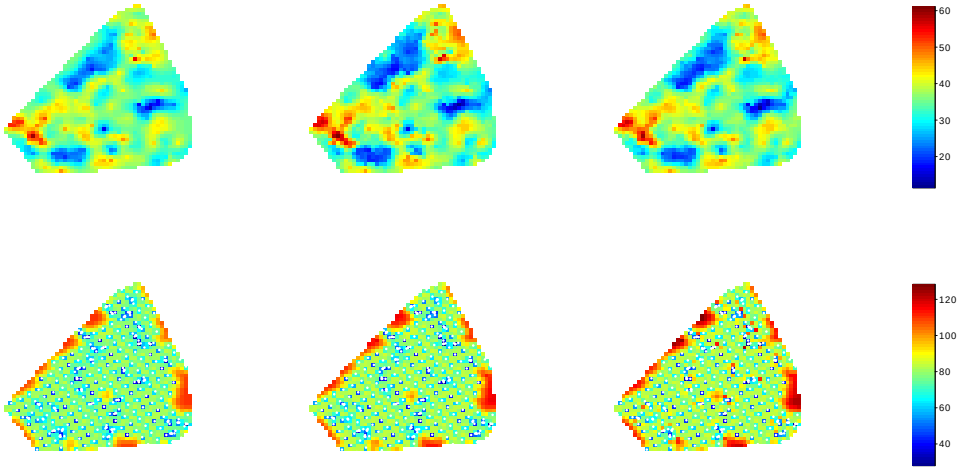


图8 汝拉数据集的预测图（上排）和克里金方差图（下排），分别使用普通克里金（左列）、惩罚克里金（中列）和局部克里金（右列）得到。颜色表示相对大小（蓝色 =低，红色 =高）。预测图和方差图使用各自独立的连续色标，不同排之间的色标不可比较。

图 10 通过相对方差增幅的箱线图，对方差进行比较，其计算公式为（$\sigma^2_{\text{method}} - \sigma^2_{\text{GK}}$）$/\sigma^2_{\text{GK}}$。自适应 LASSO 方法在所有位置上的表现都远优于局部克里金。事实上，在 98.5% 的位置上，自适应 LASSO 的方差小于或等于局部克里金方差。

请注意，惩罚克里金通过平均使用6个邻居实现了这一性能。在局部相关性较强的某些内部位置，惩罚克里金最少仅使用3个邻居，同时将方差保持在普通克里金的5%以内，这表明我们可以通过智能的邻居选择获得精确的稀疏解。

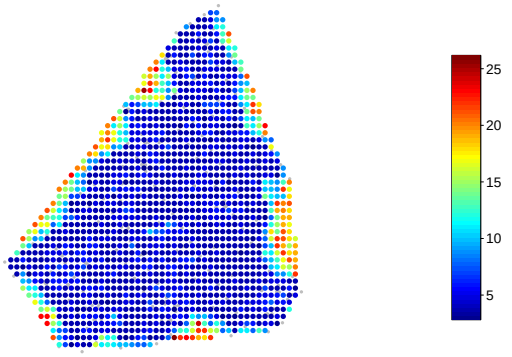


图 9 汝拉数据集采用惩罚克里金选定邻居数量的空间分布。选定的邻域大小从最少 3 个邻居到 25 个邻居不等。在内部区域，该方法通常选定约 5 个邻居；而在靠近区域边界处，邻居数量会增加至 10 至 25 个，体现了该方法能够自动适应局部空间配置。

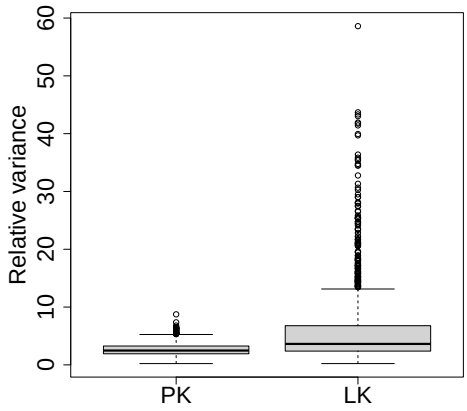


图 10 汝拉数据集中惩罚克里金（PK）和局部克里金（LK）相对于普通克里金（GK）的相对克里金方差增幅（%）箱线图。

6.2 COBE 海表温度数据集：大规模应用

在一个全局克里金可行的中等规模数据集上证明了该方法的性能之后，我们现在考察一个大规模应用，其中全局克里金在计算上变得不可行。我们分析了来自科罗拉多州博尔德NOAA物理科学实验室的COBE海表温度数据集（Ishii等2005），其提供

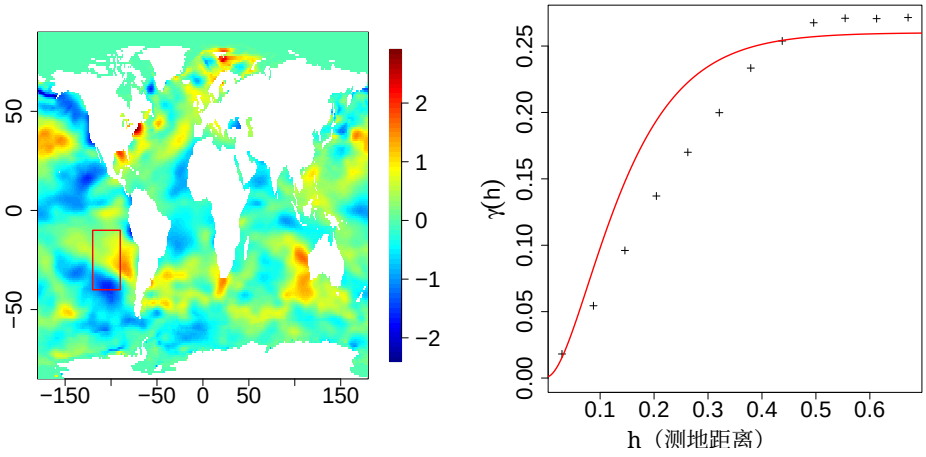
全球 $1^\circ \times 1^\circ$ 分辨率的月度网格化海表温度分析数据。该数据集包含覆盖全球海洋的43,799个空间位置，因此，在标准计算资源上进行全局克里金的矩阵求逆（$O(N^3) \approx 10^{14}$次运算）是不可行的。该数据集为测试我们方法的可扩展性以及处理大规模空间预测问题的能力提供了现实场景。

我们关注2012年3月的平均温度异常（图11a），其呈现出复杂的空间模式，包括热带太平洋拉尼˜娜事件的残余影响。我们在东南太平洋的一个矩形区域内进行降尺度分析（图11a，红框），经度范围从 -120° 至 -90°，纬度范围从 -40° 至 -10°。选择该区域有几个原因：(1) 其温度异常表现出显著的空间变化；(2) 同时包含近海和远洋区域，具有不同的相关结构；(3) 其面积足够大（$30^\circ \times 30^\circ$），既能展示可扩展性，又便于进行详细可视化。在该区域内构建了分辨率为 $0.25^\circ \times 0.25^\circ$ 的预测网格，共得到14,650个预测位置——与原始 $1^\circ \times 1^\circ$ 数据相比，空间分辨率提高了四倍。这代表了一种现实的降尺度场景，即对粗分辨率的全球模型输出进行细化，以服务于区域应用。

我们遵循Alegr´ıa （2024）的方法，拟合一个均值为常数、采用 Mat´ern 协方差函数且光滑参数 $\nu = 1.5$ 固定并包含块金效应的高斯随机场，并使用弦距离（Guinness and Fuentes 2016）来考虑地球球面几何。协方差模型为 $C(h) = \sigma^2 \mathcal{M}(h; \phi, 1。5) + \tau^2 \mathbb{I}(h = 0)$，其中$\mathcal{M}$ 表示 Mat´ern 相关函数，h表示位置之间的弦距离。

由于该数据集规模极其庞大，我们无法使用最大似然来估计协方差参数。因此，我们采用复合似然估计（Bevilacqua et al.2012；Caama˜no-Carrillo et al. 2024）。具体而言，我们对所有可用数据采用包含三个最近邻的成对似然方法（在Bevilacqua （2025）中实现）。估计得到的参数为 $\hat{}\mu_0 = 0.002$、$\hat{}\sigma^2 = 0.2592$、$\hat{}\phi = 0.0767$、$\hat{}\tau^2 = 0.0008$，实际变程约为 0。363 弧度（地球表面约 2,312 千米），块金基台比为 $\hat{}\tau^2/(\hat{}\sigma^2 + \hat{}\tau^2) \approx 0.003$，这表明测量误差约占总变异的 0。3%。图 11b 显示了经验变异函数及其拟合模型，两者吻合较好。整个 COBE 数据集的估计有效样本量为$n^{\text{ESS}} = 67.15$，尽管共有 $N = 43,799$ 个位置，仍表明存在高度的信息冗余（99。85%）。这一结果凸显了大型空间数据集中严重的信息冗余，并强调了自适应邻域选择策略对于实现可扩展克里金的实际意义。

在非常大的数据集上实施惩罚克里金的步骤如下。对于每个预测位置，$_0$, 我们首先根据球面几何形状，从包含43,799个位置的完整数据集中确定 $K = 300$ 个最近邻。选择$K = 300$ 这一数值是基于估计的有效样本量，作为候选邻域大小的保守且实用的上限，以确保候选集包含所有信息性观测，而非将其作为严格的选择规则。



(a) 海表温度平均异常值 2012年3月

(b) 经验变异函数和拟合变异函数

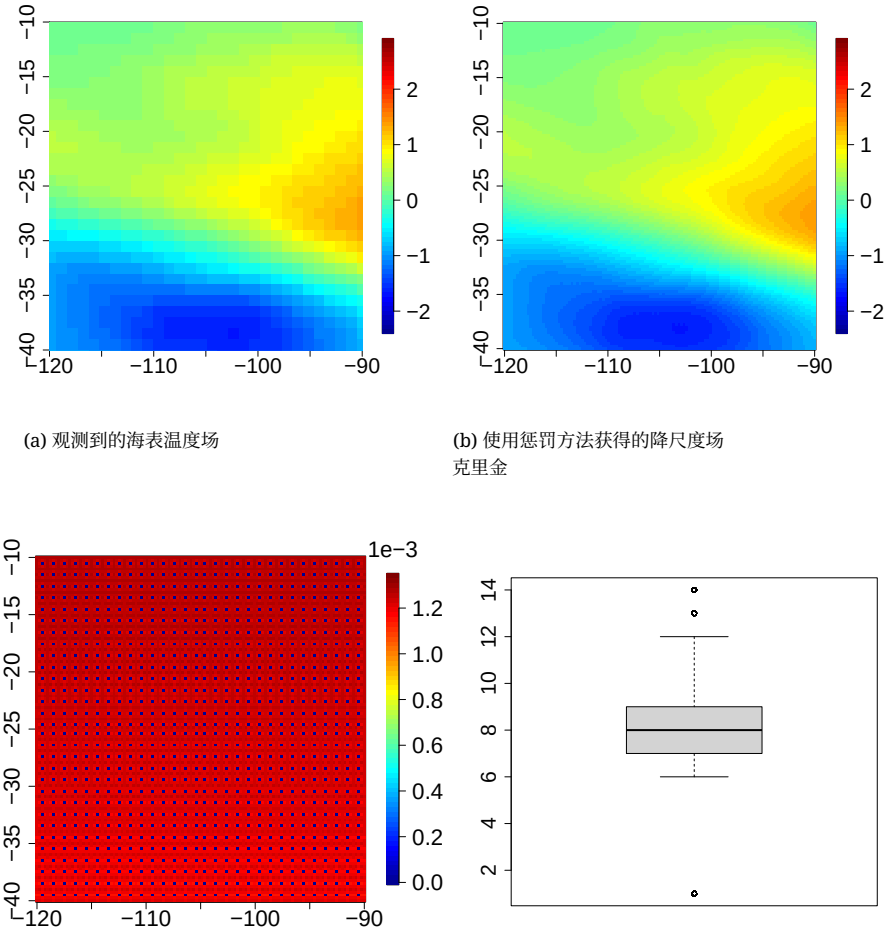
图11 左：COBE数据集中的2012年3月海表温度平均异常值（$N = 43,799$个位置）。红框标示东南太平洋中经度从-120 到 -90、纬度从 -40 到 -10 的降尺度预测区域。右：经验变异函数（叉号）以及通过复合似然估计的拟合 Mat´ern 模型（ $\nu = 1.5$ ，实线）。

惩罚克里金（PK）中使用的自适应权重计算如下： $w_i = 1/|\lambda_i|$，其中 λ_i 表示使用300个最近邻通过普通克里金（GK-300）获得的克里金权重。惩罚参数 η 根据（12）中定义的准则选取。

图12展示了12(a)观测海表温度数据的空间分布，12(b) 使用惩罚克里金得到的降尺度空间场，12(c)相应的克里金方差，以及12(d) 所选邻居数量的箱线图。

降尺度后的场呈现出平滑且符合物理规律的空间模式，较暖与较冷区域之间过渡渐进，与大尺度海洋环流相一致。尽管依赖的局部邻域相对较小，所得预测仍与原始温度场非常相似。具体而言，惩罚克里金在每个预测位置平均选择约 7 个邻居，大多数取值集中在 7 到 9 之间，如图 12(d)所示。这凸显了该方法在大幅减少所用观测值数量的同时实现有效空间细化的能力。相应的克里金方差图表明，整个区域的预测不确定性较低且空间连贯；尽管邻域大小大幅缩减，其水平仍与普通克里金相近。

这两个真实数据应用共同展示了惩罚克里金在不同空间尺度下的实用价值。对于中等规模的数据集（汝拉，359 个位置），该方法在保持精度与全局克里金相差不超过 4% 的同时，实现了较高的计算效率。对于大规模数据集（COBE，43,799 个位置），全局克里金



(a) 观测到的海表温度场

(b) 使用惩罚方法获得的降尺度场克里金

(c) 克里金方差 $\times 10^{-3}$

(d) 选定邻域大小的分布

图 12 COBE 海表温度数据集的结果。各面板分别显示12(a) 观测到的海表温度场，12(b) 使用惩罚克里金获得的降尺度场，12(c) 相应的克里金方差（$\times 10^{-3}$），以及12(d) 选定邻域大小的箱线图。观测数据和预测使用相同的色标。

在无法实施的情况下，该方法能够智能地进行邻居选择，大幅减少准确预测所需的邻居数量。

7 结论

我们开发了一种惩罚克里金框架，将 ℓ_1 正则化直接融入克里金方程，从而无需

用户指定邻域大小或交叉验证程序即可自动进行邻居选择。所提出的方法解决了大型空间数据集克里金计算方面的挑战，同时保持预测准确度，为传统邻域选择策略提供了一种有原则且可扩展的替代方案。

本研究有三项主要贡献。首先，我们使用 ℓ_1 惩罚重新构建克里金问题，使其符合克里金的无偏性约束，并将这一形式扩展为带有数据驱动惩罚权重的自适应LASSO版本。其次，我们提出一种基于有效样本量的准则，用于自动选择惩罚参数；该准则通过组合归一化信息度量的调和形式，在稀疏性与预测方差之间取得平衡，同时避免交叉验证带来的计算负担。第三，通过模拟和真实数据应用，我们证明该方法能够根据空间相关范围、过程平滑度以及预测位置的几何形状自动调整邻域结构。

模拟实验展示了该方法在各种情形下的自适应表现。随着空间相关性增强，所选邻域逐渐变得稀疏，这反映了相邻观测值之间冗余的增加。在相关范围相近时，与粗糙过程相比，更平滑的过程所需的邻居更少；而靠近区域边界的预测位置所需的邻域则比内部位置更大。这些规律自然地源自惩罚形式，无需临时调参或人工干预。

这些发现与以下研究的结论密切相关：Emery （2009），该研究通过大量组合模拟表明，选择最近邻并不一定能使克里金方差最小。在这项研究中，通过使用克里金更新方程迭代评估多种数据配置，确定了最优邻域。相比之下，本文提出的方法通过一种截然不同的机制得出了相同的实质性结论——仅凭距离远近不足以实现邻居选择的最优性：该方法将单步惩罚式形式与有效样本量准则相结合。这样既避免了穷举组合搜索，又能在计算上扩展至超大规模数据集。

所提出的框架存在若干局限，也为未来研究指明了方向。该框架依赖于根据数据估计的有效协方差模型，而模型设定错误可能会影响邻居选择。当前方法假设高斯过程和平稳性；将其扩展到非高斯情形（如转高斯或指示克里金），以及非平稳协方差结构，仍需进一步研究。后续工作可探究空间自适应LASSO的理论性质，将ESS准则与正式的模型选择框架联系起来，并把该方法扩展到多变量及时空过程。总体而言，基于ESS的惩罚克里金框架在稀疏性与预测精度之间实现了有效折衷，为空间预测中长期存在的邻居选择问题提供了一种有理论依据且计算高效的解决方案。此外，还应研究LASSO克里金下条件模拟的行为，而不是简单克里金。

致谢

本研究由智利国家研究与发展署资助，资助项目为 ANID FONDECYT 初创研究基金 11230502（J. Acosta）和 ANID FONDECYT 初创研究基金 11240330（F. Cuevas-Pacheco）。

参考文献

Atteia, O.,Dubois, J.-P., Webster, R.：瑞士汝拉地区土壤污染的地质统计分析。《环境污染》**86**(3), 315–327 (1994) https://doi.org/10.1016/0269-7491(94)90168-6

Alegr´ıa, A.：评估无矩阵分块似然估计在空间模型中的竞争力。arXiv 预印本 arXiv:2401.11265 (2024)

Acosta, J.,Vallejos, R.,and, D.G.：关于有效地理样本量。《统计计算与模拟杂志》**88**(10), 1958–1975 (2018) https://doi.org/10.1080/00949655.2018.1428977 https://doi.org/10.1080/00949655.2018.1428977

Bevilacqua, M.：GeoModels：高斯与非高斯地质统计数据分析程序。 (2025). https://doi.org/10.32614/CRAN.package. GeoModels。R 软件包版本 2.2.0。 https://CRAN.R-project.org/package=GeoModels

Bevilacqua, M., Gaetan, C., Mateu, J., Porcu, E.：针对大型数据集估计空间与时空协方差函数：一种加权复合似然方法。《美国统计学会杂志》**107**(497), 268–280 (2012)

Caama˜no-Carrillo, C., Bevilacqua, M., L´opez, C., Morales-O˜nate, V.：基于数据对的最近邻加权复合似然，适用于（非）高斯大型空间数据，并应用于 Tukey-hh随机场估计。《计算统计与数据分析》**191**, 107887 (2024)

Cressie, N., Johannesson, G.：针对超大型空间数据集的固定秩克里金法。《英国皇家统计学会杂志：B辑（统计方法学）》**70**(1), 209–226 (2008)

Cressie, N.A.C.：《空间数据统计学》，修订版。威利出版社，纽约 (1993)

Cressie, N.A.C.：《空间数据统计学》，修订版，第 900 页。威利出版社，纽约（1993）

Emery, X.：克里金更新方程及其在邻近数据选择中的应用。《计算地球科学》**13**，269–280（2009）

Furrer, R., Genton, M.G., Nychka, D.：用于高空间数据集插值的协方差渐消。《计算与图形统计学杂志》**15**(3), 502–523

（2006）

Friedman, J., Hastie, T., Tibshirani, R.：广义线性模型的坐标下降正则化路径。《统计软件杂志》**33**(1), 1（2010）Gramacy, R.B., Apley, D.W.：面向大型计算机实验的局部高斯过程近似。《计算与图形统计学杂志》**24**(2), 561–578（2015）Guinness, J., Fuentes, M.：球面上的各向同性协方差函数：若干性质与建模考虑。《多元分析杂志》**143**，143–152（2016）Gaetan, C., Guyon, X., 等：空间统计与建模。Springer，纽约，纽约州（2010）Gaines, B.R., Kim, J., Zhou, H.：约束 LASSO 的拟合算法。《计算与图形统计学杂志》**27**(4), 861–871（2018）Goovaerts, P.：《自然资源评价地统计学》，第 483 页。牛津大学出版社，纽约（1997）。附录 C 介绍了汝拉数据集。Heaton, M.J., Datta, A., Finley, A.O., Furrer, R., Guinness, J., Guhaniyogi, R., Gerber, F., Gramacy, R.B., Hammerling, D., Katzfuss, M., 等：大型空间数据分析方法的案例研究竞赛。《农业、生物与环境统计学杂志》**24**(3), 398–425（2019）Heuvelink, G.B., Pebesma, E.J.,等：普通克里金方差是否是插值误差的恰当度量？载于：第五届自然资源与环境科学空间精度评估国际研讨会。皇家墨尔本理工大学，墨尔本，第 179–186 页（2002）Hastie, T., Tibshirani, R., Friedman, J.H., Friedman, J.H.：《统计学习基础：数据挖掘、推断与预测》。Springer，纽约，纽约州（2009）Hung, Y.：计算机实验中的惩罚盲克里金。《中华统计学志》**21**(3), 1171–1190（2011）https://doi.org/10.5705/ss.2009.226Ishii, M., Shouji, A., Sugimoto, S., Matsumoto, T.：利用 ICOADS 和神户资料集对 20 世纪海表温度及海洋气象变量进行客观分析。《国际气候学杂志：皇家气象学会杂志》**25**(7), 865–879（2005）Park, I.：通过 LASSO 克里金高效选择全局趋势模型。《结构与多学科优化》**64**, 1527–1543（2021）https://doi.org/10.1007/s00158-021-02939-7Schabenberger, O., Gotway, C.A.：《空间数据分析的统计方法》。Chapman and Hall/CRC，博卡拉顿，佛罗里达州（2005）

Stein, M.L.：克里金中的屏蔽效应。《统计年鉴》**30**(1), 298–323（2002）
Tibshirani, R.：通过 LASSO 实现回归收缩与变量选择。《英国皇家统计学会杂志 B 辑：统计方法学》**58**(1), 267–288（1996）van den Boogaart, K.G., Tolosana-Delgado, R., Bren, M.：《组成：成分数据分析》（2025）。R 软件包版本 2.0-9。https://CRAN.R-project.org/package=compositionsWebster, R., Atteia, O., Dubois, J.-P.：瑞士汝拉土壤中痕量金属的协区域化。《欧洲土壤科学杂志》**45**(2), 205–218（1994）https://doi.org/10.1111/j.1365-2389.1994.tb00502.x

Zou, H., Hastie, T.：通过弹性网络进行正则化和变量选择。《英国皇家统计学会杂志B辑：统计方法》**67**(2), 301–320 (2005)

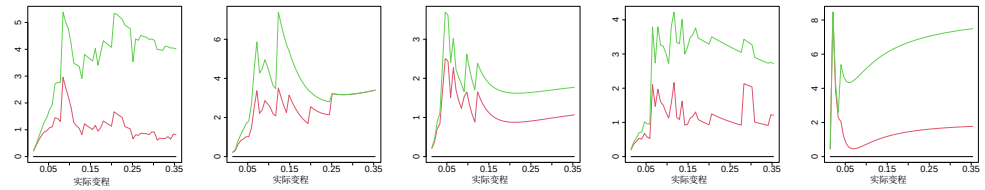
Zou, H.：自适应LASSO及其预言机性质。《美国统计学会杂志》**101**(476), 1418–1429 (2006)

完整的方差比较结果

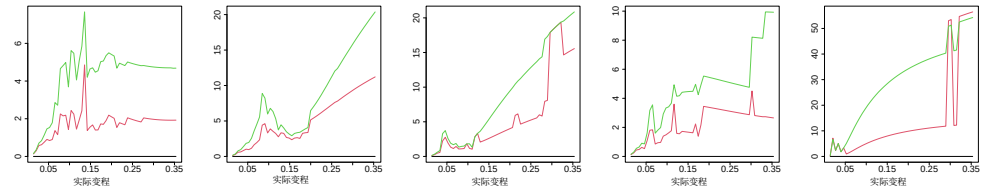
本附录展示了所有 15 种协方差模型（指数型、Mat´ern $\nu = 1.5$、球状）与预测位置（最远邻居、平均邻居、最密邻居、角点邻居、侧边邻居）组合的完整方差比较结果。每个面板均展示了相对于使用全部 500 个观测值的普通克里金，随实际变程变化的相对方差增幅（%）。红线表示惩罚克里金，绿线表示使用与K近邻相同数量邻居的局部克里金，黑线表示普通克里金基准线（0%）。这些图补充了正文图5 中呈现的部分结果。

B 克里金系数的空间分布

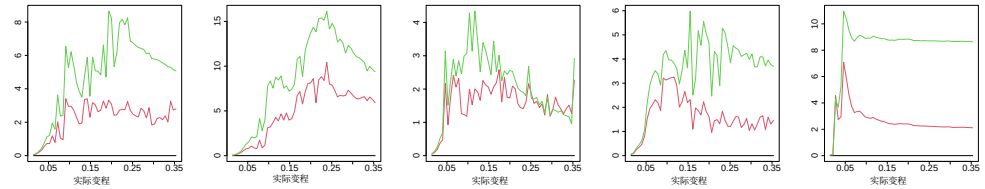
本节展示了所有协方差模型（指数型、Mat´ern $\nu = 1.5$、球状）、实用变程水平（低、中、高）和预测位置（最远、平均、最密集、角落、侧视图）组合对应的完整空间系数分布。灰色点表示系数为零的邻居（未被自适应LASSO选中），彩色点表示选定邻居，颜色深浅表示系数大小，红色点标出预测位置。所有图中的色标均相同，以便于比较。



(a) 指数型 – FN (b) 指数型 – AN (c) 指数型 – DN (d) 指数型 – SN (e) 指数型 – CN



(f) Mat´ern – FN (g) Mat´ern – AN (h) Mat´ern – DN (i) Mat´ern – SN (j) Mat´ern – CN



(k) 球状 – FN (l) 球状 – AN (m) 球状 – DN (n) 球状 – SN (o) 球状 – CN

图13 惩罚克里金（红色）和使用相同数量最近邻的局部克里金（绿色）的相对方差增幅（%），相较于普通克里金（黑色基线为0%）。各行对应协方差模型（指数型、Mat´ern $\nu = 1.5$、球状），各列对应预测位置（FN：最远、AN：平均、DN：最密集、SN：侧视图、CN：角落）。

C 梯度下降算法

我们使用近端梯度下降求解惩罚克里金问题（7）。约束条件 $\boldsymbol{X}\boldsymbol{\lambda} = \boldsymbol{x}_0$ 使我们能够表示 $\boldsymbol{\lambda}_p = \boldsymbol{X}_p^{-1}(\boldsymbol{x}_0 - \boldsymbol{X}_{-p}\boldsymbol{\lambda}_{-p})$，从而将问题简化为关于 $\boldsymbol{\lambda}_{-p} \in \mathbb{R}^{N-p}$ 的无约束优化。

令 $Q(\boldsymbol{\lambda}_{-p}) = \boldsymbol{\lambda}^\top \Sigma \boldsymbol{\lambda} - 2\boldsymbol{\lambda}^\top \boldsymbol{c}_0$。对 $\Sigma = [\Sigma_{p,p}, \Sigma_{p,-p}$ 进行分块；$\Sigma_{-p,\ p}$，$\Sigma_{-p,\ -p}]$ 和 $\boldsymbol{c}_0 = [\boldsymbol{c}_{0p}, \boldsymbol{c}_{0-p}]^\top$ 相容地，梯度为：

$$\begin{aligned}\nabla Q(\boldsymbol{\lambda}_{-p}) = &-2\boldsymbol{X}_{-p}^\top(\boldsymbol{X}_p^{-\top}\Sigma_{p,p}\boldsymbol{X}_p^{-1})(\boldsymbol{x}_0 - \boldsymbol{X}_{-p}\boldsymbol{\lambda}_{-p}) \\ &-2\boldsymbol{X}_{-p}^\top\boldsymbol{X}_p^{-\top}\Sigma_{p,-p}\boldsymbol{\lambda}_{-p} + 2\Sigma_{-p,p}\boldsymbol{X}_p^{-1}(\boldsymbol{x}_0 - \boldsymbol{X}_{-p}\boldsymbol{\lambda}_{-p}) \\ &+2\Sigma_{-p,-p}\boldsymbol{\lambda}_{-p} + 2\boldsymbol{X}_{-p}^\top\boldsymbol{X}_p^{-\top}\boldsymbol{c}_{0p} - 2\boldsymbol{c}_{0-p}\end{aligned}$$

该算法在梯度下降和软阈值处理之间交替进行：

算法 2 惩罚克里金的梯度下降

需要： Σ, $\boldsymbol{c}_0$, $\eta \geq 0$, $\boldsymbol{\lambda}_{-p}^{(0)}$, $\epsilon > 0$, $\alpha > 0$

$k \leftarrow 0$

当 $\|\boldsymbol{\lambda}_{-p}^{(k+1)} - \boldsymbol{\lambda}_{-p}^{(k)}\|_2 > \epsilon$ 且 $k < 1000$ **时执行**

  $\boldsymbol{g} \leftarrow \nabla Q(\boldsymbol{\lambda}_{-p}^{(k)})$

  $\boldsymbol{\lambda}_{-p}^{(k+1)} \leftarrow \boldsymbol{\lambda}_{-p}^{(k)} - \alpha \boldsymbol{g}$

  $\boldsymbol{\lambda}_{-p}^{(k+1)} \leftarrow$ 符号$(\boldsymbol{\lambda}_{-p}^{(k+1)}) \odot (|\boldsymbol{\lambda}_{-p}^{(k+1)}| - \eta\alpha)+$

  $k \leftarrow k + 1$

结束循环

返回 $\boldsymbol{\lambda} = [\boldsymbol{X}_p^{-1}(\boldsymbol{x}_0 - \boldsymbol{X}_{-p}\boldsymbol{\lambda}_{-p}^{(k)}) \ \boldsymbol{\lambda}_{-p}^{(k)}]^\top$

其中，$(a)_+ = \max(a,0)$，$\odot$ 表示逐元素乘法。对于自适应~LASSO，对每个分量 i，将 $\eta\alpha$ 替换为 $\eta\alpha w_i$，其中 $w_i = 1/|\lambda_i|$ 为自适应权重。我们使用 $\alpha = 0.01$，并在可用时以普通克里金权重作为热启动。

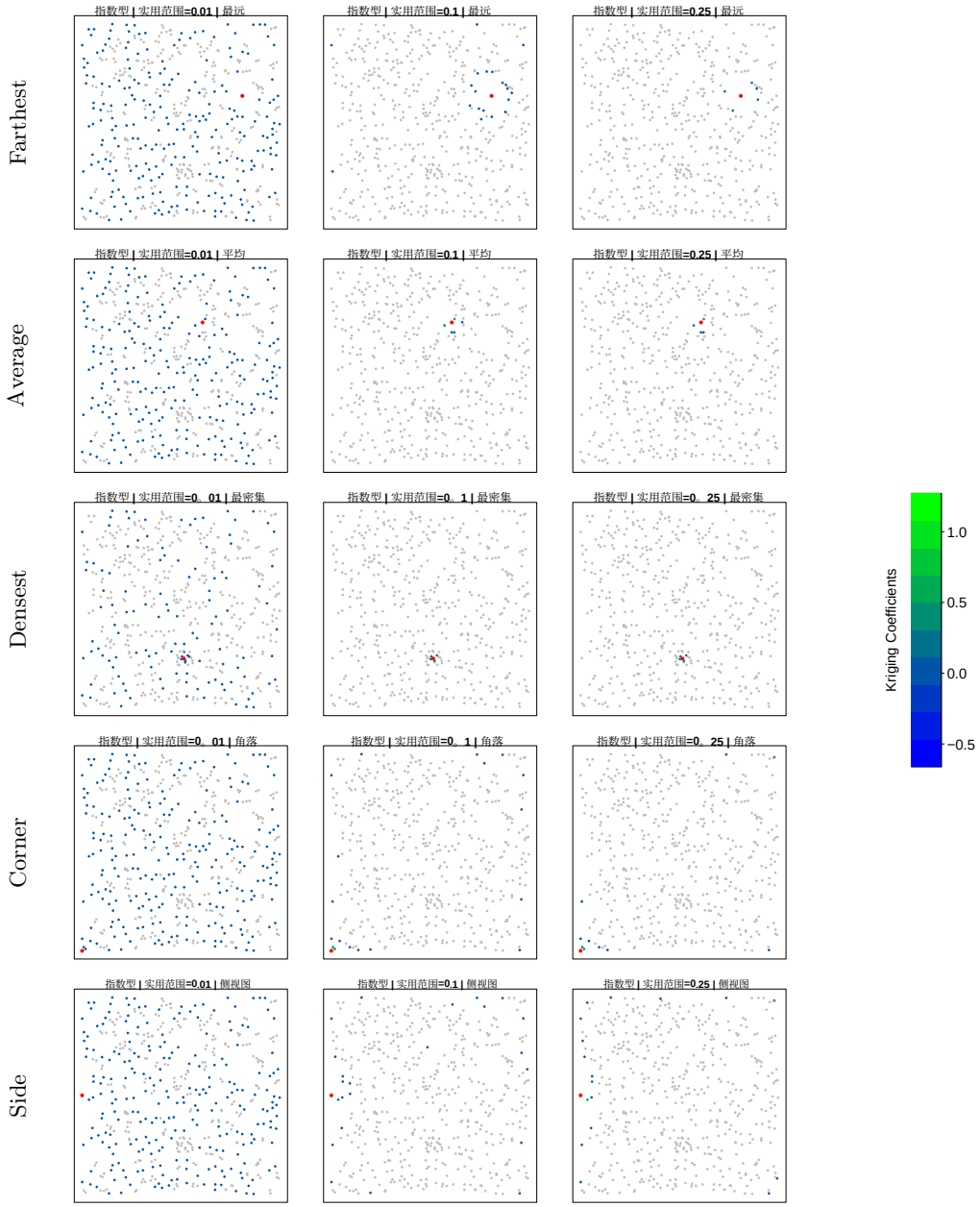


图14 五个预测位置（行）和三个实用范围级别（列）下，指数型协方差对应的自适应 LASSO 克里金系数的空间分布。灰色点表示系数为零的邻居（未选定）。彩色点表示选定邻居，颜色深浅表示系数大小（蓝色 =较小，绿色 = 较大）。红点标记预测位置 $\boldsymbol{s}_0$。

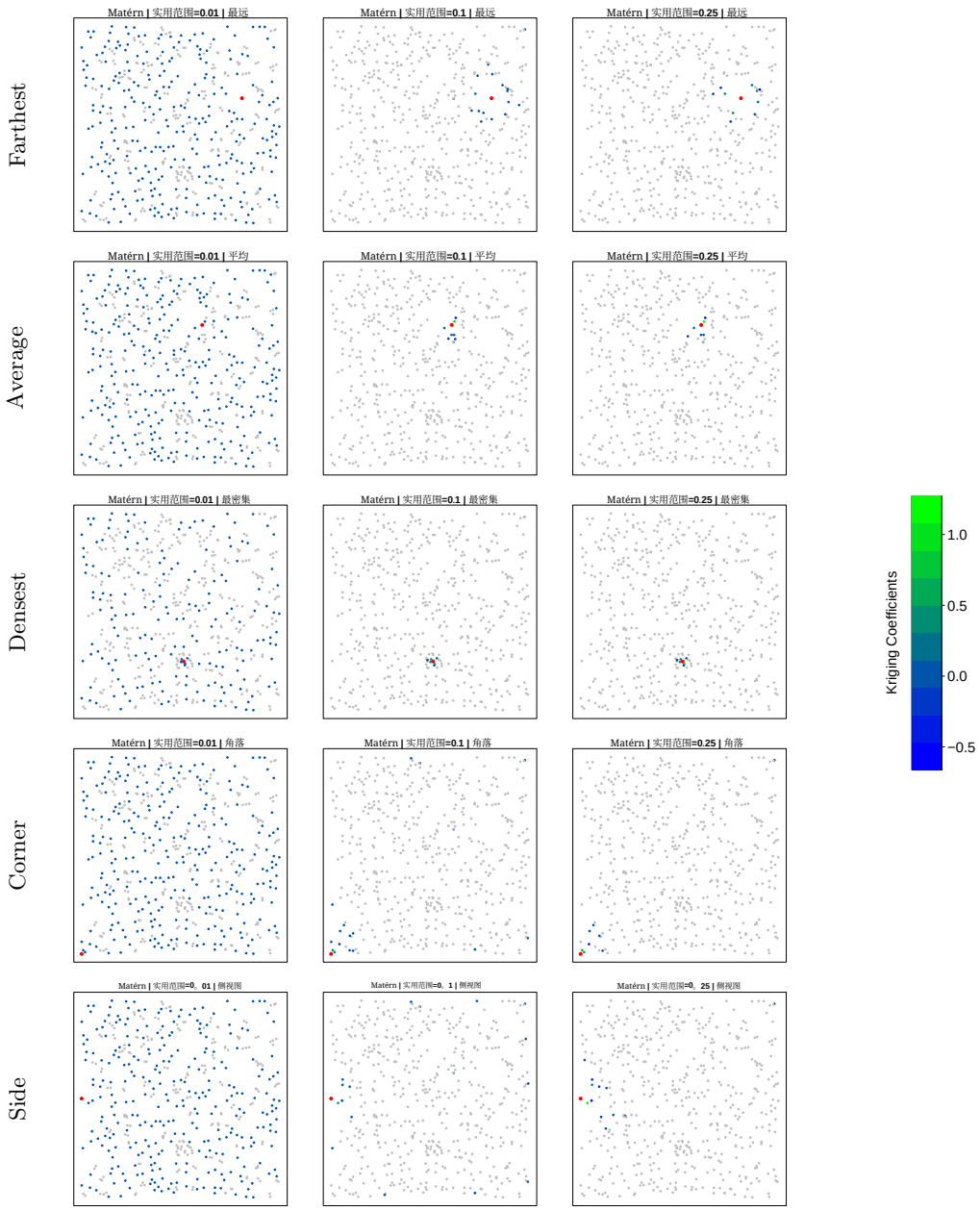


图 15 Mat´ern 协方差（$\nu = 1.5$）下自适应 LASSO 克里金系数在五个预测位置（行）和三个实用范围级别（列）上的空间分布。灰色点表示系数为零的邻居（未被选定）。彩色点表示选定邻居，颜色深浅表示系数大小（蓝色 = 小，绿色 = 大）。红点标记预测位置 $\boldsymbol{s}_0$。

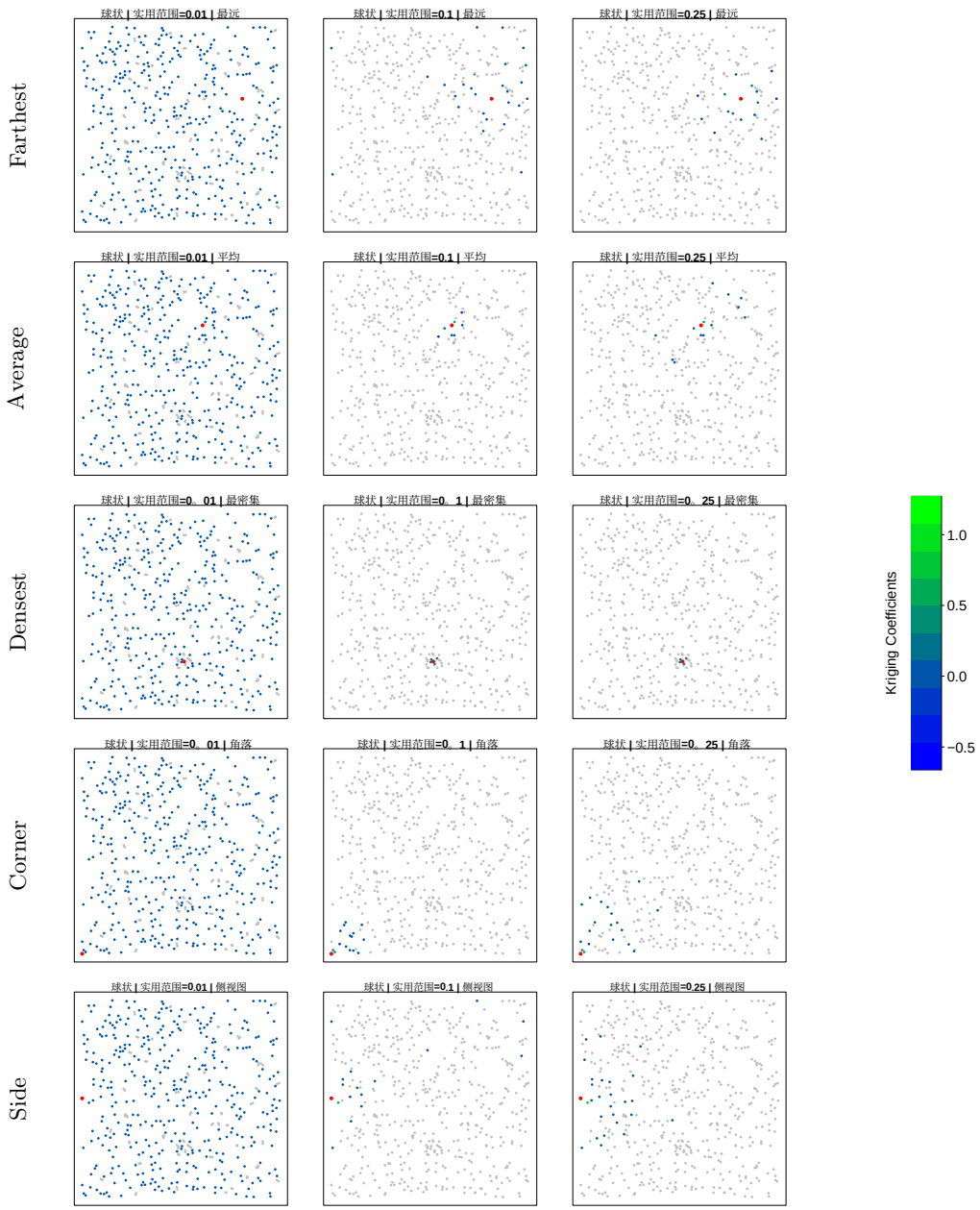


图 16 球状协方差下，自适应 LASSO 克里金系数在五个预测位置（行）和三个实用范围级别（列）上的空间分布。灰色点表示系数为零的邻居（未被选中）。彩色点表示被选中的邻居，颜色深浅表示系数大小（蓝色 = 较小，绿色 = 较大）。红色点表示预测位置 s_0。