arXiv:2303.04408v2 [stat.ME] 7 Dec 2023

具有序列相关的二维函数型数据的主成分分析

Shirun Shen*、Huiya Zhou、Kejun He†和Lan Zhou

摘要

本文提出了一种新模型，用于分析在可能并非矩形的区域上稀疏且不规则观测到的序列相关二维函数型数据。我们采用混合效应模型，将主成分函数表示为定义在三角形上的双变量样条，并将主成分得分视为服从自回归模型的随机效应。我们使用薄板惩罚对双变量函数估计进行正则化，并结合卡尔曼滤波与平滑器，开发了一种有效的EM算法，用于计算参数的惩罚似然估计。我们将该方法应用于模拟数据集，以及得克萨斯州1915年1月至2014年12月的月平均温度数据。

关键词：双变量样条；EM算法；函数型主成分分析；卡尔曼滤波与平滑器；三角剖分。

1 引言

研究天气模式随时间和地理位置变化，对于研究气候变化至关重要。为了探究美国得克萨斯州的温度变化，我们使用了美国历史气候学网络的一个数据集（Menne等人，2009年，USHCN），该数据集由美国国家海洋和大气管理局（NOAA）收集。数据集包含1915年至2014年间得克萨斯州49个分布稀疏的气象站观测到的月平均气温。

Shirun Shen（邮箱：srshen@ruc.edu.cn）为学生，Huiya Zhou（邮箱：freedom00y@ruc.edu.cn）为学生，Kejun He（邮箱：kejunhe@ruc.edu.cn）为助理教授，任职于中国人民大学统计与大数据研究院应用统计中心，中国北京，邮编100872。Lan Zhou（邮箱：lzhou@stat.tamu.edu）为副教授，任职于得克萨斯农工大学统计系，得克萨斯州大学城，邮编TX 77843-3143。本研究得到中国人民大学公共计算云的支持。

*本文主要是在Shirun Shen于得克萨斯农工大学统计系攻读博士期间完成的，地址为得克萨斯州大学城，邮编TX 77843-3143。

†通讯作者。

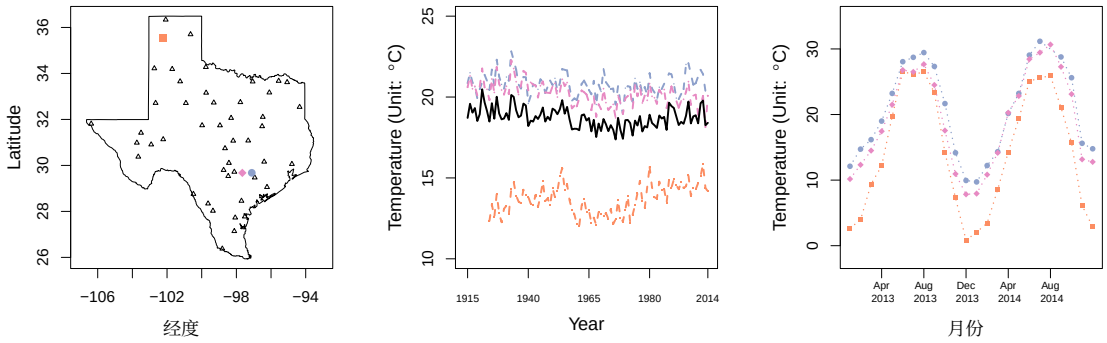


图 1：左：得克萨斯州 49 个气象站的位置。中：3 个气象站（虚线）的年平均温度（1915–2014年）以及所有气象站的平均值（黑色实线）。右：3 个气象站在 2013年和2014年的月平均温度。在中图和右图中，颜色和符号对应左图中的相应气象站。

图1 左侧面板展示了得克萨斯州49个气象站的位置。其中，三个气象站以不同颜色和符号标出。图1 中间和右侧面板分别显示了左侧面板中三个不同颜色的气象站从1915年至2014年的年平均气温，以及它们在2013年和2014年的月平均气温。可以看出，虽然南部相邻的两个气象站随时间呈现相似的趋势，但北部气象站的温度始终较低，且变化趋势不同，尤其是在1980年之前。三个气象站在2013年和2014年的季节变化几乎完全相同；与南部的两个气象站相比，北部气象站的温度在冬夏之间波动更大。

使用该数据集研究得克萨斯州多年间天气模式的变化，会给数据分析带来以下挑战。首先，温度会受到测量位置、气候变化趋势以及季节性的影响。其次，气象站在得克萨斯州这一不规则区域内分布稀疏，而且不同气象站在不同月份存在缺失观测。第三，但同样重要的是，随着时间推移观测到的温度之间通常存在序列相关；例如，参见 Jones等人(1986年) 和Hansen等人(2006年)，尤其是在

温度来自同一气象站时。

为了分析作为纬度和经度所定义位置函数的温度，函数型主成分分析（FPCA，Ramsay and Silverman，2005年）提供了一种富有信息的方式，用于考察函数型数据的协方差结构；否则，这种结构通常难以理解。主成分函数代表函数型数据的主要变异模式，并刻画数据集中典型函数的特征。当函数型数据在一维区域上相互独立时，关于估计少数几个领先函数型主成分函数已有大量研究，包括局部多项式方法（Staniswalis and Lee，1998年；Yao 等人，2005年a、b；Hall 等人，2006年）和样条方法（Rice and Wu，2001年；James 等人，2000年；Huang 等人，2008年；Zhou 等人，2008年；Li 和 Guan，2014年；Liu 等人，2017年）。对于二维独立函数型数据，Zhou 和 Pan（2014a）采用带有三角形上的双变量样条的混合效应模型来估计主成分函数。其他关于二维函数型数据的近期研究包括 Chen 和 Jiang（2017年）、Wang 等人（2020年b）、Ding 等人（2022年）以及 Shi 等人（2022年）。这些论文均未考虑数据中存在的序列相关。

为了分析具有序列相关的函数型数据，例如我们研究中的德克萨斯州温度数据，一种方法是将自回归模型扩展到函数型数据，直接对函数建模，参见Bosq (2000年) 和 Kokoszka and Reimherr (2013年）。然而，当函数仅被稀疏观测时，例如德克萨斯州温度数据，这种方法并不适用。另一种基于降维思想的方法是一个两步过程：首先应用函数主成分分析以获得函数主成分得分，然后用时间序列模型对估计得到的函数主成分得分建模；例如，参见 Hyndmanand Ullah(2007); Shen andHuang (2008); Hyndman and Shang (2009年); Shen(2009年），以及其中的参考文献。 ShangandHyndman (2017年) 将 Hyndmanand Ullah (2007) 的研究扩展到多层结构数据。 Cabrera and Schulz (2017年) 使用惩罚B样条为广义函数分位数构造主成分函数，并再次对函数主成分得分拟合时间序列模型。然而，上述研究仅考虑了一维函数型时间序列，无法简单地扩展到对具有缺失数据、且观测于不规则二维定义域上的二维函数型时间序列建模，例如德克萨斯州温度数据。

在本研究中，我们提出一种统一的方法，用于对序列相关的二维 1

函数型数据进行建模，并使用函数主成分分析模型分析得克萨斯州温度数据。具体而言，为了刻画位置的变化模式，并克服空间域上观测稀疏以及得克萨斯州形状不规则所带来的挑战，我们沿用Zhou and Pan(2014a），使用三角形上的双变量样条对主成分函数进行建模。我们对函数主成分得分采用自回归模型，以捕捉时间上的序列相关。气候变化趋势和季节性分别用单变量样条和傅里叶函数近似。我们将以上各项整合到惩罚完全数据对数似然函数中，并采用EM算法估计未知参数。在EM算法的E步中，我们进一步使用卡尔曼滤波与平滑器，克服计算序列相关潜变量期望时遇到的困难。由于在潜在函数主成分得分上采用了自回归模型，我们提出的序列函数型主成分模型（sFPC）不要求每个时间点的观测数量达到最低要求，因此能够直接处理缺失观测，例如得克萨斯州温度数据中的缺失观测。

本文其余部分安排如下。在第2节中，我们提出一种基于混合效应模型的方法，用于对具有序列相关的二维函数型数据进行主成分分析。我们将主成分函数表示为定义在三角形上的双变量样条，并将主成分得分设为服从自回归模型的随机效应。在第3节中，我们提出一种用于模型拟合的EM算法，并在E步中应用卡尔曼滤波与平滑器程序。第4 节和5节分别通过模拟研究和德克萨斯州温度数据分析，展示所提方法的实证表现。在第6节中，我们总结本文的主要贡献，并给出结论性评述。

2 序列相关二维函数型数据的混合效应模型

设 Ω 是 $\mathbb{R}^2$ 的紧子集，且 $(x,\ y)$ 是 Ω 上的二维指标变量。假设 $Z(x,\ y)$ 是 Ω 上具有有限二阶矩的随机过程，即 $\int_\Omega \mathbb{E}\{Z^2(x,\ y)\}\mathrm{d}x\mathrm{d}y < \infty$。将 $Z(x,\ y)$ 的均值函数记为 $\mu(x,\ y) = \mathbb{E}\{Z(x,\ y)\}$，将协方差函数

记为 $Z(x,\ y)$,
即

$$\mathcal{K}(x_1, y_1; x_2, y_2) = \mathbb{E}\big[\{Z(x_1, y_1) - \mu(x_1, y_1)\}\{Z(x_2, y_2) - \mu(x_2, y_2)\}\big].$$

在温和条件下，Mercer 引理（Mercer，1909）表明，存在一个正交归一序列 $\{\phi_j\}_j$ 属于 $L_2(\Omega)$，其元素为特征函数；还存在一个非负递减序列 $\{\zeta_j\}_j$作为特征值，使得协方差函数可展开为

$$\mathcal{K}(x_1, y_1; x_2, y_2) = \sum_{j=1}^{\infty} \zeta_j \phi_j(x_1, y_1)\phi_j(x_2, y_2).$$

特征函数 ϕ_j的正交归一性意味着$\int_\Omega \phi_j\phi_{j'}\,\mathrm{d}x\mathrm{d}y = \delta_{jj'}$，其中 $\delta_{jj'}$ 是克罗内克 delta。应用 Karhunen-Loève 定理（Karhunen，1946；Lo`eve，1946），随机曲面 $Z(x,\ y)$可展开如下

$$Z(x, y) = \mu(x, y) + \sum_{j=1}^{\infty} \alpha_j \phi_j(x, y), \tag{1}$$

其中 α_j是均值为零、方差为 $\{\zeta_j\}_j$的不相关随机变量。根据Ramsay and Silverman（2005年），随机变量 α_j 和特征函数 $\phi_j(x,\ y)$分别称为第 j个函数主成分得分和主成分函数。

假设 $Z(x,\ y)$ 可由其在前 J 个特征函数张成的空间上的投影很好地近似，并将其余项视为噪声，我们得到以下模型

$$Z(x, y) = \mu(x, y) + \sum_{j=1}^{J} \alpha_j \phi_j(x, y) + \epsilon(x, y), \tag{2}$$

其中，$\epsilon(x,\ y)$ 是均值为 0、方差为 σ^2 的白噪声。

当有 n 个 $Z(x, y)$ 的独立副本，记为 $Z_1(x, y)$、. . . 、 $Z_n(x, y)$ 时，Zhou 和 Pan（2014a）提出了一种基于混合效应模型的方法，并将函数主成分得分建模为随机效应。当随机曲面 $Z_t(x, y)$、 $t = 1$、. . . 、 n 具有依赖于时间和位置的均值函数 $\mu_t(x, y) = \mathbb{E}\{Z_t(x, y)\}$，且 $Z_t(x, y)$ 序列相关时，我们需要同时考虑位置和时间点之间的依赖关系。

我们首先假设位置效应和时间效应在均值函数上是可分离的，因此 $\mu_t(x, y) = \mu_1(x, y)\mu_2(t)$。注意，如果将 $\mu_1(x, y)$ 乘以一个非零常数c，并将 $\mu_2(t)$ 除以 c，则 $\mu_t(x, y)$ 的值不会改变。为了保证可识别性，

我们要求 $\mu_1(x,y)$ 的 L_2-范数为单位，即

$$\|\mu_1(x,y)\|_2 = 1. \tag{3}$$

接下来，借鉴（2）中函数主成分分析的思路，我们提出如下模型

$$Z_t(x,y) = \mu_1(x,y)\mu_2(t) + \sum_{j=1}^{J} \alpha_{j,t}\phi_j(x,y) + \epsilon_t(x,y), \quad t = 1,\ldots,n, \tag{4}$$

其中 $\alpha_{j,t}$ 表示时刻 t 的第 j 个函数主成分得分，$\phi_j(x,y)$ 表示第 j 个主成分函数，且这些函数正交归一；$\epsilon_t(x,y)$ 是均值为 0、方差为 σ^2 的白噪声过程。此外，$\{\alpha_{j,t}\}_{t=1}^n$, $j = 1,\ldots,J$ 是相互独立的平稳时间序列。对于每个 j，时间序列 $\{\alpha_{j,t}\}_{t=1}^n$ 遵循 p 阶自回归模型（AR(p）。具体而言，

$$\alpha_{j,t} = \sum_{i=1}^{p} k_{i,j}\alpha_{j,t-i} + \eta_{j,t}, \quad \eta_{j,t} \overset{\text{ind}}{\sim} \mathrm{N}(0,\sigma_j^2), \quad j = 1,\ldots,J \quad \text{and} \quad t = 1,\ldots,n, \tag{5}$$

其中 $k_{i,j}$'s 和 $\eta_{j,t}$'s 分别是 AR(p) 模型的自回归系数和白噪声。为保证主成分的可识别性，我们假设$\sigma_1^2 > \cdots > \sigma_J^2$。

注意，当 $k_{i,j} = 0$, $i = 1,\ldots,p$和 $j = 1,\ldots,J$ 时，函数主成分得分 $\alpha_{j,t}$'s 相互独立且服从正态分布。此时，所提出的模型（4）退化为 Zhou and Pan (2014a) 中的模型。

假设函数 $\mu_1(x,y)$、$\mu_2(t)$ 和 $\phi_j(x,y)$ 是光滑的，我们可以用基展开对它们进行近似。对于二元函数 $\mu_1(x,y)$ 和 $\phi_j(x,y)$ 的近似，由于正交归一二元伯恩斯坦多项式在不规则区域上具有优势，我们采用在三角形上构造的正交归一二元伯恩斯坦多项式（Lai and Schumaker，2007）。Liu 等人（2016）也从理论上研究了伯恩斯坦多项式的近似性质。关于三角形上的二元基函数的详情，请参见附录的 B节。至于一元函数 $\mu_2(t)$ 的基展开，我们可以选择常用的回归样条（de Boor，1978年）、伯恩斯坦样条（Lorentz，1986年）、傅里叶基（Ramsay and Silverman，2005年）等。

令$\mathbf{b}(x,y)$表示由正交归一的二元基函数组成的 n_b维向量，其中

$$\int_\Omega \mathbf{b}(x,y)\mathbf{b}^\mathsf{T}(x,y)\mathrm{d}x\mathrm{d}y = \mathbf{I}_{n_b}, \tag{6}$$

令$\mathbf{c}(t)$表示由单变量基函数组成的 n_c维向量。我们将光滑函数的基展开写为

$$\mu_1(x,y)=\mathbf{b}(x,y)^\mathsf{T}\boldsymbol{\theta}_b,\quad \mu_2(t)=\mathbf{c}(t)^\mathsf{T}\boldsymbol{\theta}_c,\quad \text{and}\quad \phi_j(x,y)=\mathbf{b}(x,y)^\mathsf{T}\boldsymbol{\theta}_j,\quad j=1,\ldots,J,$$

其中基系数 $\boldsymbol{\theta}_b\in\mathbb{R}^{n_b}$with $\|\boldsymbol{\theta}_b\|=1$, $\boldsymbol{\theta_c}\in\mathbb{R}^{n_c}$, and $\boldsymbol{\theta}_j\in\mathbb{R}^{n_b}$, $j=1,\ldots,J$。此外，$\boldsymbol{\theta}_j$, $j=1,\ldots,J$是正交归一的。记 $\boldsymbol{\Theta}=(\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_J)\in\mathbb{R}^{n_b\times J}$, $\boldsymbol{\alpha}_t=(\alpha_{1,t},\ldots,\alpha_{J,t})^\mathsf{T}$，且 $\mathbf{K}_i=\text{diag}(k_{i,1},\ldots,k_{i,J})$，$i=1,\ldots,J$，模型 (4) 可重写为

$$Z_t(x,y)=\mathbf{b}(x,y)^\mathsf{T}\boldsymbol{\theta}_b\boldsymbol{\theta}_c^\mathsf{T}\mathbf{c}(t)+\mathbf{b}(x,y)^\mathsf{T}\boldsymbol{\Theta}\boldsymbol{\alpha}_t+\epsilon_t(x,y),\quad t=1,\ldots,n,\tag{7}$$

以及自回归模型 (5) 变为

$$\boldsymbol{\alpha}_t=\sum_{i=1}^{p}\mathbf{K}_i\boldsymbol{\alpha}_{t-i}+\boldsymbol{\eta}_t,\quad t=1,\ldots,n,$$

其中 $\boldsymbol{\eta}_t\sim\text{N}(\mathbf{0},\mathbf{H}_J)$，协方差矩阵为 $\mathbf{H}_J=\text{diag}(\sigma_1^2,\ldots,\sigma_J^2)$。

假设在时间 $t=1,\ldots,n$ 观测到稀疏采样的二维曲面。在时间点 t，曲面上有 n_t 个随机采样点 $(x_{t1},y_{t1})\ldots,(x_{tn_t},y_{tn_t})$。为简洁起见，记 $\mathbf{z}_t=(Z_t(x_{t1},y_{t1})\ldots,Z_t(x_{tn_t},y_{tn_t}))^\mathsf{T}$、$\mathbf{B}_t=(\mathbf{b}(x_{t1},y_{t1})\ldots,\mathbf{b}(x_{tn_t},y_{tn_t}))^\mathsf{T}$、$\boldsymbol{\epsilon}_t=(\epsilon_t(x_{t1},y_{t1})\ldots,\epsilon_t(x_{tn_t},y_{tn_t}))^\mathsf{T}$，以及 $\mathbf{c}_t=\mathbf{c}(t)$。于是，观测数据和潜变量的模型 (7) 可写为

$$\begin{cases}\mathbf{z}_t=\mathbf{B}_t\boldsymbol{\theta}_b\boldsymbol{\theta}_c^\mathsf{T}\mathbf{c}_t+\mathbf{B}_t\boldsymbol{\Theta}\boldsymbol{\alpha}_t+\boldsymbol{\epsilon}_t,\quad \boldsymbol{\epsilon}_t\sim\text{N}(\mathbf{0},\sigma^2\mathbf{I}_{n_t}),\\ \boldsymbol{\alpha}_t=\sum_{i=1}^{p}\mathbf{K}_i\boldsymbol{\alpha}_{t-i}+\boldsymbol{\eta}_t,\quad \boldsymbol{\eta}_t\sim\text{N}(\mathbf{0},\mathbf{H}_J),\end{cases}\tag{8}$$

对于 $t=1,$ ....., n，可识别性约束与上文所述相同。注意，n_t 表示在时间 t有观测值的位置数量。它可能随 t而变化，也可以为 0。

记 $\mathbf{K}=(k_{i,\ j})_{i=1,\ j=1}^{p\ J}\in\mathbb{R}^{p\times J}$。对于模型（8），待估计的未知参数可统写为 $\Xi=\{\boldsymbol{\theta}_b,\boldsymbol{\theta}_c,\boldsymbol{\Theta},\mathbf{K}\ \sigma^2,\ \{\sigma_j^2\}_{j=1}^J\}$。

3 模型拟合

3.1 惩罚完全数据对数似然

根据模型（8），自然的做法是通过最大化对数似然函数来估计未知参数 Ξ，并加入一些惩罚项以正则化平滑函数的估计

。然而，由于潜在函数主成分得分 $\{\boldsymbol{\alpha}_t\}_{t=1}^n$ 服从 AR（p）模型，因此无法将 $\{\boldsymbol{\alpha}_t\}_{t=1}^n$ 积分消去，从而得到 Ξ的对数似然函数的解析形式。将 $\{\boldsymbol{\alpha}_t\}_{t=1}^n$ 视为缺失数据后，我们可以得到完全数据对数似然的解析形式，进而应用 EM算法（Dempster等人，1977）进行参数估计。

令 $\mathbf{k}_j$ 为 j 阶列 $\mathbf{K}$，即 $\mathbf{k}_j = (k_{1,j}, \ldots, k_{p,j})^\mathsf{T}$, $j = 1, \ldots, J$。由模型（8），可得完全数据对数似然为

$$-2l_c(\Xi; \{\mathbf{z}_t\}_{t=1}^n, \{\boldsymbol{\alpha}_t\}_{t=1}^n) = \frac{1}{\sigma^2}\sum_{t=1}^n (\mathbf{z}_t - \mathbf{B}_t\boldsymbol{\theta}_b\boldsymbol{\theta}_c^\mathsf{T}\mathbf{c}_t - \mathbf{B}_t\boldsymbol{\Theta}\boldsymbol{\alpha}_t)^\mathsf{T}(\mathbf{z}_t - \mathbf{B}_t\boldsymbol{\theta}_b\boldsymbol{\theta}_c^\mathsf{T}\mathbf{c}_t - \mathbf{B}_t\boldsymbol{\Theta}\boldsymbol{\alpha}_t) + \sum_{t=1}^n n_t \log\sigma^2 + \sum_{j=1}^J \left\{ n\log\sigma_j^2 - \log|\mathbf{M}_j| + \frac{1}{\sigma_j^2}S_j(\mathbf{k}_j) \right\}, \quad (9)$$

其中，$\mathbf{M}_j$ 是 $(\alpha_{j,1}/\sigma_j, \ldots, \alpha_{j,p}/\sigma_j)^\mathsf{T}$ 的精度矩阵，而 $S_j(\mathbf{k}_j)$ 的形式为 $(\mathbf{1}, \mathbf{k}_j^\mathsf{T})\mathbf{D}_j(\mathbf{1}, \mathbf{k}_j^\mathsf{T})^\mathsf{T}$，满足

$$\mathbf{D}_j = \begin{pmatrix} D_{1,1,j} & -D_{1,2,j} & -D_{1,3,j} & \ldots & -D_{1,p+1,j} \\ -D_{1,2,j} & D_{2,2,j} & D_{2,3,j} & \ldots & D_{2,p+1,j} \\ \vdots & \vdots & \vdots & & \vdots \\ -D_{p+1,1,j} & D_{p+1,2,j} & D_{p+1,3,j} & \ldots & D_{p+1,p+1,j} \end{pmatrix} \in \mathbb{R}^{(p+1)\times(p+1)} \quad (10)$$

其中 $D_{i,k,j} = D_{k,i,j} = \alpha_{j,i}\alpha_{j,k} + \alpha_{j,i+1}\alpha_{j,k+1} + \cdots + \alpha_{j,n+1-k}\alpha_{j,n+1-i}$。推导（9）和（10）的详细过程见附录的 A 节。

接下来，我们引入粗糙度惩罚，对光滑函数的估计进行正则化。对于一元函数 $\mu_2(t)$，如 Zhou 等人(2008) 所示，我们使用二阶导数平方积分，粗糙度惩罚的形式为

$$\int_T \left\{\frac{\partial^2\mu_2(t)}{\partial t^2}\right\}^2 \mathrm{d}t = \boldsymbol{\theta}_c^\mathsf{T}\left[\int_T \left\{\frac{\partial^2\mathbf{c}(t)}{\partial t^2}\frac{\partial^2\mathbf{c}(t)^\mathsf{T}}{\partial t^2}\right\}\mathrm{d}t\right]\boldsymbol{\theta}_c := \boldsymbol{\theta}_c^\mathsf{T}\mathbf{P}\boldsymbol{\theta}_c.$$

对于一般的二元函数 $f(x, y)$，我们采用薄板惩罚项（Ruppert等人，2003），其定义为

$$\int_\Omega \left[\left\{\frac{\partial^2 f(x,y)}{\partial x^2}\right\}^2 + 2\left\{\frac{\partial^2 f(x,y)}{\partial x\partial y}\right\}^2 + \left\{\frac{\partial^2 f(x,y)}{\partial y^2}\right\}^2\right]\mathrm{d}x\mathrm{d}y.$$

记为

$$\boldsymbol{\Gamma} = \int_\Omega \left\{\frac{\partial^2\mathbf{b}(x,y)}{\partial x^2}\frac{\partial^2\mathbf{b}(x,y)^\mathsf{T}}{\partial x^2} + 2\frac{\partial^2\mathbf{b}(x,y)}{\partial x\partial y}\frac{\partial^2\mathbf{b}(x,y)^\mathsf{T}}{\partial x\partial y} + \frac{\partial^2\mathbf{b}(x,y)}{\partial y^2}\frac{\partial^2\mathbf{b}(x,y)^\mathsf{T}}{\partial y^2}\right\}\mathrm{d}x\mathrm{d}y.$$

采用基展开后，$\mu_1(x,y)$ 和 $\phi_j(x,y)$ 的薄板惩罚分别可写为 $\boldsymbol{\theta}_b\boldsymbol{\Gamma}\boldsymbol{\theta}_b$ 和 $\boldsymbol{\theta}_j\boldsymbol{\Gamma}\boldsymbol{\theta}_j$, $j=1,\ldots,J$。

因此，惩罚完全数据对数似然的表达式为

$$
\begin{aligned}
&-2l_c(\Xi;\{\mathbf{z}_t\}_{t=1}^n,\{\boldsymbol{\alpha}_t\}_{t=1}^n)+\text{Penalty}(\lambda;\Xi)\\
&\quad=\frac{1}{\sigma^2}\sum_{t=1}^n(\mathbf{z}_t-\mathbf{B}_t\boldsymbol{\theta}_b\boldsymbol{\theta}_c^{\mathsf{T}}\mathbf{c}_t-\mathbf{B}_t\boldsymbol{\Theta}\boldsymbol{\alpha}_t)^{\mathsf{T}}(\mathbf{z}_t-\mathbf{B}_t\boldsymbol{\theta}_b\boldsymbol{\theta}_c^{\mathsf{T}}\mathbf{c}_t-\mathbf{B}_t\boldsymbol{\Theta}\boldsymbol{\alpha}_t)+\sum_{t=1}^n n_t\log\sigma^2 \qquad (11)\\
&\qquad+\sum_{j=1}^J\left\{n\log\sigma_j^2-\log|\mathbf{M}_j|+\frac{1}{\sigma_j^2}S_j(\mathbf{k}_j)\right\}+\lambda_{\mu_s}\boldsymbol{\theta}_b^{\mathsf{T}}\boldsymbol{\Gamma}\boldsymbol{\theta}_b+\lambda_{\mu_t}\boldsymbol{\theta}_c^{\mathsf{T}}\mathbf{P}\boldsymbol{\theta}_c+\lambda_{pc}\sum_{j=1}^J\boldsymbol{\theta}_j^{\mathsf{T}}\boldsymbol{\Gamma}\boldsymbol{\theta}_j,
\end{aligned}
$$

其中，$\lambda=(\lambda_{\mu_s},\ \lambda_{\mu_t},\ \lambda_{pc})$ 为正则化参数。

3.2 EM算法

为估计参数，我们不再最小化（11），而是通过 EM算法迭代地最小化

$$
Q(\Xi|\Xi^{(0)})=\mathbb{E}\big[-2l_c(\Xi;\{\mathbf{z}_t\}_{t=1}^n,\{\boldsymbol{\alpha}_t\}_{t=1}^n)|\{\mathbf{z}_t\}_{t=1}^n,\Xi^{(0)}\big]+\text{Penalty}(\lambda;\Xi),
$$

其中，$\Xi^{(0)}$ 是参数值的当前估计（即上一轮迭代得到的参数值）。

E步。在 E步中，我们计算 $Q(\Xi|\Xi^{(0)})$。记 $\widehat{\boldsymbol{\alpha}}_t=\mathbb{E}(\boldsymbol{\alpha}_t|\{\mathbf{z}_t\}_{t=1}^n,\ \Xi^{(0)})$、$\widehat{\boldsymbol{\Sigma}}_t=\text{Var}(\boldsymbol{\alpha}_t|\{\mathbf{z}_t\}_{t=1}^n,\ \Xi^{(0)})$ 和 $\widehat{\mathbf{D}}_j=\mathbb{E}(\mathbf{D}_j|\{\mathbf{z}_t\}_{t=1}^n,\ \Xi^{(0)})$，$j=1,\ \ldots,\ J$。我们得到

$$
\widehat{S}_j(\mathbf{k}_j):=\mathbb{E}[S_j(\mathbf{k}_j)|\{\mathbf{z}_t\}_{t=1}^n,\Xi^{(0)}]=(1,\mathbf{k}_j)^{\mathsf{T}}\widehat{\mathbf{D}}_j(1,\mathbf{k}_j^{\mathsf{T}})^{\mathsf{T}}. \qquad (12)
$$

根据（9）–（12），可得

$$
\begin{aligned}
Q(\Xi|\Xi^{(0)})=\frac{1}{\sigma^2}\sum_{t=1}^n&\left\{(\mathbf{z}_t-\mathbf{B}_t\boldsymbol{\theta}_b\boldsymbol{\theta}_c^{\mathsf{T}}\mathbf{c}_t-\mathbf{B}_t\boldsymbol{\Theta}\widehat{\boldsymbol{\alpha}}_t)^{\mathsf{T}}(\mathbf{z}_t-\mathbf{B}_t\boldsymbol{\theta}_b\boldsymbol{\theta}_c^{\mathsf{T}}\mathbf{c}_t-\mathbf{B}_t\boldsymbol{\Theta}\widehat{\boldsymbol{\alpha}}_t)+\text{tr}(\mathbf{B}_t\boldsymbol{\Theta}\widehat{\boldsymbol{\Sigma}}_t\boldsymbol{\Theta}^{\mathsf{T}}\mathbf{B}_t^{\mathsf{T}})\right\}\\
&+\sum_{t=1}^n n_t\log\sigma^2+\sum_{j=1}^J\left\{n\log\sigma_j^2-\log|\mathbf{M}_j|+\frac{1}{\sigma_j^2}\widehat{S}_j(\mathbf{k}_j)\right\} \qquad (13)\\
&+\lambda_{\mu_s}\boldsymbol{\theta}_b^{\mathsf{T}}\boldsymbol{\Gamma}\boldsymbol{\theta}_b+\lambda_{\mu_t}\boldsymbol{\theta}_c^{\mathsf{T}}\mathbf{P}\boldsymbol{\theta}_c+\lambda_{pc}\sum_{j=1}^J\boldsymbol{\theta}_j^{\mathsf{T}}\boldsymbol{\Gamma}\boldsymbol{\theta}_j.
\end{aligned}
$$

因此，要计算 (13) 的值，我们只需计算 $\widehat{\boldsymbol{\alpha}}_t$, $\widehat{\boldsymbol{\Sigma}}_t$，以及 $\widehat{\mathbf{D}}j$, $j=1,\ldots,J$。

注意，模型 (8) 可视为一个状态空间模型 (Durbin andKoopman，2012)。具体而言，记 $\boldsymbol{\beta}_t = (\boldsymbol{\alpha}_{t+p}^\top, \ldots, \boldsymbol{\alpha}_t^\top)^\top$, $t = 1, \ldots, n$，其中 $\boldsymbol{\alpha}_t = 0$ 当 $t > n$。记 $\mathbf{S} = \big(\mathbf{0}, \cdots \mathbf{0}, \mathbf{I}$ $\big) \in \mathbb{R}^{J \times (p+1)J}$ 和

$$\mathbf{T} = \begin{pmatrix} \mathbf{K}_1 & \mathbf{K}_2 & \ldots & \mathbf{K}_p & \mathbf{0} \\ \mathbf{I}_J & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_J & \ldots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{I}_J & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{(p+1)J \times (p+1)J},$$

模型（8）可改写为

$$\begin{cases} \mathbf{z}_t = \mathbf{B}_t \boldsymbol{\theta}_b \boldsymbol{\theta}_c^\top \mathbf{c}_t + \mathbf{B}_t \boldsymbol{\Theta} \mathbf{S} \boldsymbol{\beta}_t + \boldsymbol{\epsilon}_t, \\ \boldsymbol{\beta}_t = \mathbf{T} \boldsymbol{\beta}_{t-1} + \boldsymbol{\xi}_t, \end{cases} \tag{14}$$

其中 $\boldsymbol{\xi}_t = (\boldsymbol{\eta}_{t+p}^\top, 0, \cdots, 0)^\top \sim \mathrm{N}(0, \widetilde{\mathbf{H}})$，其中 $\widetilde{\mathbf{H}} = \mathrm{diag}(\mathbf{H}_J, 0, \cdots, 0) \in \mathbb{R}^{(p+1)J \times (p+1)J}$。

现在有了状态空间模型（14），可以使用卡尔曼滤波与平滑器（Durbin andKoopman，$\widehat{}$2012）来求得 $\mathbf{b}_t = \mathbb{E}(\boldsymbol{\beta}_t | \mathbf{z}_1, \ldots, \mathbf{z}_n)$ 和 $\mathbf{V}_t = \mathrm{Var}(\boldsymbol{\beta}_t | \mathbf{z}_1, \ldots, \mathbf{z}_n)$，然后我们可以得到

$$\widehat{\boldsymbol{\alpha}}_t = \mathbb{E}(\boldsymbol{\alpha}_t | \mathbf{z}_1, \ldots, \mathbf{z}_n, \Xi^{(0)}) = \mathbf{S}\widehat{\mathbf{b}}_t \quad \text{and} \quad \widehat{\boldsymbol{\Sigma}}_t = \mathrm{Var}(\boldsymbol{\alpha}_t | \mathbf{z}_1, \ldots, \mathbf{z}_n, \Xi^{(0)}) = \mathbf{S}\mathbf{V}_t\mathbf{S}^\top.$$

根据（10），$\widehat{}\mathbf{D}_j$ 可通过计算 $\widehat{D}_{i,\ k,\ j} := \mathbb{E}(D_{i,\ k,\ j} | \mathbf{Z}\ \Xi^{(0)})$ 得到，如下所示：

$$\begin{aligned} \widehat{D}_{i,k,j} &= \mathbb{E}(\alpha_{j,i}\alpha_{j,k} + \cdots + \alpha_{j,n+1-k}\alpha_{j,n+1-i} | \mathbf{Z}, \Xi^{(0)}) \\ &= \mathbb{E}(\alpha_{j,i} | \mathbf{Z}, \Xi^{(0)})\mathbb{E}(\alpha_{j,k} | \mathbf{Z}, \Xi^{(0)}) + \cdots + \mathbb{E}(\alpha_{j,n+1-k} | \mathbf{Z}, \Xi^{(0)})\mathbb{E}(\alpha_{j,n+1-i} | \mathbf{Z}, \Xi^{(0)}) \\ &\quad + \mathrm{Cov}(\alpha_{j,i}, \alpha_{j,k} | \mathbf{Z}, \Xi^{(0)}) + \cdots + \mathrm{Cov}(\alpha_{j,n+1-k}, \alpha_{j,n+1-i} | \mathbf{Z}, \Xi^{(0)}) \\ &= \widehat{\alpha}_{j,i}\widehat{\alpha}_{j,k} + \ldots \widehat{\alpha}_{j,n+1-k}\widehat{\alpha}_{j,n+1-i} + \sum_{t=1}^{n+1-i-k} \mathbf{V}_{(1+p-i)J+j,(1+p-k)J+j,t} \end{aligned}$$

对于 $j = 1, \ldots, J$，其中 $\widehat{\alpha}_{j,t}$ 是 j 在 $\widehat{\boldsymbol{\alpha}}_t$ 中的第 j 个元素，而 $\mathbf{V}_{(1+p-i)J+j,(1+p-k)J+j,t}$ 是 $\mathbf{V}_t$ 中第 $((1+p-i)J+j, (1+p-k)J+j)$ 个元素，$t = 1, \ldots, n$。

$\widehat{}$接下来需要通过卡尔曼滤波与平滑器（**Durbin and Koopman**$_t$ 和$\mathbf{V}_t$（Durbin and Koopman，2012）来求得。卡尔曼滤波与平滑器的过程是先应用卡尔曼滤波算法，再按相反顺序应用卡尔曼平滑器。下面几段将详细说明。

令 $\mathbf{b}_{t|t-1}$ 和 $\mathbf{Q}_{t|t-1}$ 分别表示 $\boldsymbol{\beta}_t$ 的一步预测及其不确定性，即 $\mathbf{b}_{t|t-1} = \mathbb{E}\{\boldsymbol{\beta}_t|\mathbf{z}^{1:(t-1)}, \Xi^{(0)}\}$ 和 $\mathbf{Q}_{t|t-1} = \mathrm{Var}(\boldsymbol{\beta}_t|\mathbf{z}_{t-1}, \ldots, \mathbf{z}1, \Xi^{(0)})$，$t = 1, \ldots, n$。将时刻 t 状态的估计及其不确定性分别记为 $\mathbf{b}_{t|t} = \mathbb{E}(\boldsymbol{\beta}_t|\mathbf{z}_t, \ldots, \mathbf{z}1, \Xi^{(0)})$和 $\mathbf{Q}_{t|t} = \mathrm{Var}(\boldsymbol{\beta}_t|\mathbf{z}_t, \ldots, \mathbf{z}1, \Xi^{(0)})$。卡尔曼滤波采用预测—校正循环。预测步骤更新

$$\begin{cases} \mathbf{b}_{t|t-1} = \mathbf{T}^{(0)}\mathbf{b}_{t-1|t-1} \\ \mathbf{Q}_{t|t-1} = \mathbf{T}^{(0)}\mathbf{Q}_{t-1|t-1}(\mathbf{T}^{(0)})^{\mathsf{T}} + \widetilde{\mathbf{H}}^{(0)}, \end{cases}$$

其中，$\mathbf{T}^{(0)}$ 和 $\widetilde{\mathbf{H}}^{(0)}$ 分别对应代入 $\Xi^{(0)}$当前取值后的 $\mathbf{T}$ 和 $\widetilde{\mathbf{H}}$。在校正步骤中，我们将 $\mathbf{F}_t = \mathbf{Q}^{t|t-1}(\mathbf{B}_t\boldsymbol{\Theta}^{(0)}\mathbf{S})^{\mathsf{T}}\{(\mathbf{B}_t\boldsymbol{\Theta}^{(0)}\mathbf{S})\mathbf{Q}^{t|t-1}(\mathbf{B}_t\boldsymbol{\Theta}^{(0)}\mathbf{S})^{\mathsf{T}} + \sigma^{2(0)}\mathbf{I}_{n_t}\}^{-1}$记为卡尔曼增益矩阵，并更新

$$\begin{cases} \mathbf{b}_{t|t} = \mathbf{b}_{t|t-1} + \mathbf{F}_t\{\mathbf{z}_t - \mathbf{B}_t\boldsymbol{\theta}_b^{(0)}(\boldsymbol{\theta}_c^{(0)})^{\mathsf{T}}\mathbf{c}_t - \mathbf{B}_t\boldsymbol{\Theta}^{(0)}\mathbf{S}\mathbf{b}_{t|t-1}\} \\ \mathbf{Q}_{t|t} = \mathbf{Q}_{t|t-1} - \mathbf{F}_t(\mathbf{B}_t\boldsymbol{\Theta}^{(0)}\mathbf{S})\mathbf{Q}_{t|t-1}, \end{cases}$$

其中，$\boldsymbol{\Theta}^{(0)}$、$\sigma^{2(0)}$、$\boldsymbol{\theta}_b^{(0)}$ 和 $\boldsymbol{\theta}_c^{(0)}$ 分别对应 $\boldsymbol{\Theta}$、σ^2、$\boldsymbol{\theta}_b$和 $\boldsymbol{\theta}_c$的当前取值。对于初始化，我们采用常用的非信息性取值，即 $\mathbf{b}_{0|0} = \mathbf{0}$ 和 $\mathbf{Q}_{0|0} = \mathbf{0}$。

接下来，我们应用卡尔曼平滑递推，通过更新公式得到 $\mathbf{b}_t$和 $\mathbf{V}_t$, $t = n-1, \ldots, 1$,

$$\begin{cases} \widehat{\mathbf{b}}_t = \mathbf{b}_{t|t} + \mathbf{L}_t(\widehat{\mathbf{b}}_{t+1} - \mathbf{b}_{t+1|t}) \\ \mathbf{V}_t = \mathbf{Q}_{t|t} + \mathbf{L}_t(\mathbf{V}_{t+1} - \mathbf{Q}_{t+1|t})\mathbf{L}_t^{\mathsf{T}}, \end{cases}$$

其中 $\mathbf{L}_t = \mathbf{Q}^{t|t}(\mathbf{T}^{(0)})^{\mathsf{T}}\mathbf{Q}_{t+1|t}^{-1}$，其初始值为 $\widehat{\mathbf{b}}_n = \mathbf{b}^{n|n}$ 和 $\mathbf{V}_n = \mathbf{Q}^{n|n}$，这是由它们的定义所致。

我们指出，即使某些时间点没有观测值，E步中的条件期望也可以使用卡尔曼滤波与平滑器计算；例如参见Cipra 和 Romera(1997年)。

M步。 在M步中，我们求解 $Q(\Xi|\Xi^{(0)})$ 在（13）中的极小化解。通常很难推导出极小化解的显式形式。注意，在（13）中，参数 $\boldsymbol{\theta}_b$、$\boldsymbol{\theta}_c$、$\boldsymbol{\Theta}$、$\mathbf{K}$ σ^2、和 $\{\sigma_j^2\}_{j=1}^J$彼此分离。因此，我们采用分块优化，并在其他参数固定为当前值 $\Xi^{(0)}$时，讨论每个参数的更新规则。

关于 $\boldsymbol{\theta}_b \in \mathbb{R}^{n_b}$ 在（13）中的优化问题，等价于在约束 $\boldsymbol{\theta}_b^{\mathsf{T}}\boldsymbol{\theta}_b = 1$下最小化（$\boldsymbol{\theta}_b - \mathbf{m}$）$^{\mathsf{T}}\mathbf{A}$（$\boldsymbol{\theta}_b - \mathbf{m}$），其中

$$\begin{cases} \mathbf{m} = \left\{ \sum_{t=1}^{n} (\boldsymbol{\theta}_c^{(0)\mathsf{T}}\mathbf{c}_t)^2 \mathbf{B}_t^{\mathsf{T}}\mathbf{B}_t + \sigma^{2(0)}\lambda_{\mu_s}\boldsymbol{\Gamma} \right\}^{-1} \sum_{t=1}^{n} (\boldsymbol{\theta}_c^{(0)\mathsf{T}}\mathbf{c}_t)\mathbf{B}_t^{\mathsf{T}}(\mathbf{z}_t - \mathbf{B}_t\boldsymbol{\Theta}^{(0)}\widehat{\boldsymbol{\alpha}}_t) \\ \mathbf{A} = \sum_{t=1}^{n} (\boldsymbol{\theta}_c^{(0)\mathsf{T}}\mathbf{c}_t)^2 \mathbf{B}_t^{\mathsf{T}}\mathbf{B}_t + \sigma^{2(0)}\lambda_{\mu_s}\boldsymbol{\Gamma}. \end{cases}$$

将 $\boldsymbol{\theta}_b$ 视为 n_b维欧几里得空间中的嵌入子流形后，$\boldsymbol{\theta}_b$ 便可使用子流形上的梯度下降算法进行更新（Absil等人，2009年）。附录C.1 第 C.1 节中的算法 S.1 针对在球面子流形上使用梯度下降算法的实现进行了专门说明。

我们提出通过将相应的分块导数设为零，来更新 $\boldsymbol{\theta}_c$、σ^2、σ_j^2、$j = 1$、……、J。具体而言，对式（13）关于 $\boldsymbol{\theta}_c$求导并令其为零，可按如下方式更新 $\boldsymbol{\theta}_c$：

$$\widehat{\boldsymbol{\theta}}_c = \left\{ \sum_{t=1}^{n} (\mathbf{B}_t\widehat{\boldsymbol{\theta}}_b\mathbf{c}_t^{\mathsf{T}})^{\mathsf{T}}(\mathbf{B}_t\widehat{\boldsymbol{\theta}}_b\mathbf{c}_t^{\mathsf{T}}) + \sigma^{2(0)}\lambda_{\mu_t}\mathbf{P} \right\}^{-1} \sum_{t=1}^{n} (\mathbf{B}_t\widehat{\boldsymbol{\theta}}_b\mathbf{c}_t^{\mathsf{T}})^{\mathsf{T}}(\mathbf{z}_t - \mathbf{B}_t\boldsymbol{\Theta}^{(0)}\widehat{\boldsymbol{\alpha}}_t).$$

同理，σ^2 的更新公式为

$$\widehat{\sigma}^2 = \frac{1}{\sum_{t=1}^{n} n_t} \sum_{t=1}^{n} \Big\{ (\mathbf{z}_t - \mathbf{B}_t\widehat{\boldsymbol{\theta}}_b\widehat{\boldsymbol{\theta}}_c^{\mathsf{T}}\mathbf{c}_t - \mathbf{B}_t\boldsymbol{\Theta}^{(0)}\widehat{\boldsymbol{\alpha}}_t)^{\mathsf{T}}(\mathbf{z}_t - \mathbf{B}_t\widehat{\boldsymbol{\theta}}_b\widehat{\boldsymbol{\theta}}_c^{\mathsf{T}}\mathbf{c}_t - \mathbf{B}_t\boldsymbol{\Theta}^{(0)}\widehat{\boldsymbol{\alpha}}_t) + \operatorname{tr}(\mathbf{B}_t\boldsymbol{\Theta}^{(0)}\widehat{\boldsymbol{\Sigma}}_t\boldsymbol{\Theta}^{(0)\mathsf{T}}\mathbf{B}_t^{\mathsf{T}}) \Big\},$$

σ_j^2的更新公式为 $\widehat{\sigma}_j^2 = \widehat{S}_j(\mathbf{k}_j^{(0)})/n$, $j = 1, \ldots, J$，其中 $\widehat{S}_j(\mathbf{k}_j^{(0)})$与将 $\mathbf{k}_j^{(0)}$（j的第 列 (12) $\mathbf{K}^{(0)}$）代入后的形式相同。

对于 $\boldsymbol{\Theta}$，我们首先依次更新 $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1$，……，$\boldsymbol{\theta}_J)$ 的各列。关于13) 对 $\boldsymbol{\theta}_j$ 进行最小化，等价于将其最小化

$$\sum_{t=1}^{n} \left\| \mathbf{z}_t - \mathbf{B}_t\widehat{\boldsymbol{\theta}}_b\widehat{\boldsymbol{\theta}}_c^{\mathsf{T}}\mathbf{c}_t - \sum_{j' \neq j} \mathbf{B}_t\boldsymbol{\theta}_{j'}\widehat{\alpha}_{j',t} - \mathbf{B}_t\boldsymbol{\theta}_j\widehat{\alpha}_{j,t} \right\|^2 + \sum_{t=1}^{n} \operatorname{tr}(\mathbf{B}_t\boldsymbol{\Theta}\widehat{\boldsymbol{\Sigma}}_t\boldsymbol{\Theta}^{\mathsf{T}}\mathbf{B}_t^{\mathsf{T}}) + \widehat{\sigma}^2\lambda_{pc}\boldsymbol{\theta}_j^{\mathsf{T}}\boldsymbol{\Gamma}\boldsymbol{\theta}_j,$$

它有解析形式

$$\widehat{\boldsymbol{\theta}}_j = \left\{ \sum_{t=1}^{n} (\widehat{\alpha}_{j,t}^2 + \widehat{\boldsymbol{\Sigma}}_{t,jj})\mathbf{B}_t^{\mathsf{T}}\mathbf{B}_t + \widehat{\sigma}^2\lambda_{pc}\boldsymbol{\Gamma} \right\}^{-1} \cdot \sum_{t=1}^{n} \mathbf{B}_t^{\mathsf{T}} \left\{ (\mathbf{z}_t - \mathbf{B}_t\widehat{\boldsymbol{\theta}}_b\widehat{\boldsymbol{\theta}}_c^{\mathsf{T}}\mathbf{c}_t)\widehat{\alpha}_{j,t} - \sum_{j' \neq j} (\widehat{\alpha}_{j',t}\widehat{\alpha}_{j,t} + \widehat{\boldsymbol{\Sigma}}_{t,j'j})\mathbf{B}_t\widehat{\boldsymbol{\theta}}_{j'} \right\}.$$

为保证 $\widehat{\boldsymbol{\Theta}}$的正交归一性，我们利用 $\widehat{\boldsymbol{\Theta}}\,\widehat{\mathbf{H}}_J\widehat{\boldsymbol{\Theta}}^{\mathsf{T}}$的谱分解，其中 $\widehat{\mathbf{H}}_J = \operatorname{diag}(\widehat{\sigma}_1^2, \ldots, \widehat{\sigma}_J^2)$。具体而言，令 $\widehat{\boldsymbol{\Theta}}\,\widehat{\mathbf{H}}_J\widehat{\boldsymbol{\Theta}}^{\mathsf{T}} = \widetilde{\mathbf{Q}}\,\widetilde{\mathbf{D}}\,\widetilde{\mathbf{Q}}^{\mathsf{T}}$，其中 $\widetilde{\mathbf{Q}}$是正交归一的

~ ^ ^且 $\mathbf{D}$ 是一个对角元素递减的对角矩阵。随后，我们分别用$\tilde{}\mathbf{Q}$ 和 $\tilde{}\mathbf{D}$替换 $\boldsymbol{\Theta}$ 和$\mathbf{H}_J$。此外，我们用 $\tilde{}\mathbf{Q}^\mathsf{T}\widehat{\boldsymbol{\Theta}}\widehat{\boldsymbol{\alpha}}_t$替换 $\widehat{\boldsymbol{\alpha}}_t$，这种变换$\hat{}$保持了 $\boldsymbol{\Theta}\widehat{\boldsymbol{\alpha}}_t$的方差。

最后，我们的目标是对13进行最小化，关于$\mathbf{K}$，这等价于最小化

$$\sum_{j=1}^{J}\left\{-\log|\mathbf{M}_j|+\frac{1}{\widehat{\sigma}_j^2}\widehat{S}_j(\mathbf{k}_j)\right\}, \tag{15}$$

其中，$\mathbf{M}_j$是$(\alpha_{j,1}/\sigma_j,\ldots,\alpha_{j,p}/\sigma_j)^\mathsf{T}$的精度矩阵，而 $\widehat{S}_j(\mathbf{k}_j)$由(12)给出。注意，$\log|\mathbf{M}_j|$ 的值不随样本量 n的变化而变化，而(15)中的第二项则依赖于 n。为简化计算，我们使用第二项$\sum_{j=1}^{J}\{(1/\widehat{\sigma}_j^2)\widehat{S}_j(\mathbf{k}_j)\}$ 来近似(15)。由于 $\widehat{S}_j(\mathbf{k}_j)=(1,\mathbf{k}_j^\mathsf{T})\,\hat{}\mathbf{D}_j(1,\mathbf{k}_j^\mathsf{T})^\mathsf{T}$ 关于$\mathbf{k}_j$是一个二次型，因此利用加权最小二乘，我们可以按列更新$\mathbf{K}$，如下所示：$\hat{}\mathbf{k}_j=\left(\hat{}\mathbf{D}_{pj}\right)^{-1}\hat{}\mathbf{d}_j$, $j=1,\ldots,J$，其中 $\hat{}\mathbf{d}_j=(\widehat{D}_{1,2,j},\ldots,\widehat{D}_{1,p+1,j})^\mathsf{T}$，且 $\hat{}\mathbf{D}_{pj}$是$\mathbf{D}_j$右下角的$\sim p\times p$阶主子矩阵。

3.3 模型选择

构建三角剖分的一般准则是，应避免出现内角非常小的三角形，并且将所有时间点的观测位置合并后，不应存在不包含任何数据点的三角形。关于三角剖分的详细讨论，请参阅 Lai and Schumaker (2007)第4章。采用惩罚样条法时，在许多应用中，基函数的数量并非关键，只要数量适中且足够大即可，因为粗糙度惩罚有助于正则化估计并防止过拟合（Ruppert等人，2003）。此外，二维基函数$\mathbf{b}(x,\ y)$的光滑度由多项式的阶数 d 以及三角形相连边上光滑度参数的阶数 r 决定。实际应用中，用户也可以根据对数据的先验知识设定这两个阶数。主成分的数量由时间FPC得分方差的经验比例决定。潜变量自回归模型的阶数 p可以通过赤池信息准则（AIC，赤池弘次，1974年）或贝叶斯信息准则（BIC,

Schwarz，1978年）等数据驱动准则进行选择，使得 p 最小化

$$\sum_{j=1}^{J}\left\{n\log\widehat{\sigma}_j^2+\frac{1}{\widehat{\sigma}_j^2}\widehat{S}_j(\widehat{\mathbf{k}}_j)\right\}+2p \quad \text{or} \quad \sum_{j=1}^{J}\left\{n\log\widehat{\sigma}_j^2+\frac{1}{\widehat{\sigma}_j^2}\widehat{S}_j(\widehat{\mathbf{k}}_j)\right\}+\log(n)p.$$

正则化参数 λ_{μ_s}、 λ_{μ_t}和 λ_{pc} 可通过最小化以下值来确定： K折留一位置交叉验证（CV）。具体而言，每个时间点的样本位置会被随机划分为 K 个大小相等的子集。然而，由于有三个正则化参数，经典的全网格搜索因计算成本过高而不切实际。作为替代方案，我们建议采用单纯形法（Nelder 和 Mead，1965）来寻找局部最优解。正则化参数（λ_{μ_s}、 λ_{μ_t}、λ_{pc}）的整体选择过程分为两步。第一步，我们设置少量网格点，使其分布足够稀疏，以覆盖较宽的正则化参数范围，并应用 K折交叉验证，根据交叉预测误差确定最佳候选值。随后，我们将选定的点作为初始值，并使用单纯形法获得局部最优的最终正则化参数。

4 模拟研究

在本节中，我们通过模拟研究比较所提出的序列相关函数主成分（sFPC）模型与Zhou and Pan (2014a)中的多变量模型（mFPC）的性能，其中未考虑序列相关性。

在模拟实验中，不规则区域 Ω 设为一个 2×2 正方形，中央有一个 1×1 正方形孔洞，如图 2所示。我们根据模型（ $Z_t(x,y)$）和（4）以及（5），在 Ω 上生成了包含 $J=2$ 个主成分、阶数为$p=2$的自回归模型的数据。均值函数 $\mu_t(x,y)=\mu_1(x,y)\mu_2(t)$ 及主成分函数如下所示：

$$\begin{aligned}
\mu_1(x,y) &= 5\big\{\exp\big(\sqrt{0.1x^2+0.2y}\big)+\exp\big(-\sqrt{0.1x^2+0.2y}\big)\big\},\\
\mu_2(t) &= \cos(2\pi t/12)+t/n \quad \text{or} \quad \mu_2(t)=1,\\
\phi_1(x,y) &= 0.8578\sin(x^2+0.5y^2),\\
\phi_2(x,y) &= 0.8721\sin(0.3x^2+0.6y^2)-0.2988\sin(x^2+0.5y^2).
\end{aligned}$$

注意，主成分函数是正交归一的，因此 $\int_\Omega \phi_1^2(x,y)\mathrm{d}x\mathrm{d}y = 1$,$\int_\Omega \phi_2^2(x,y)\mathrm{d}x\mathrm{d}y = 1$，且 $\int_\Omega \phi_1(x,y)\phi_2(x,y)\mathrm{d}x\mathrm{d}y = 0$。

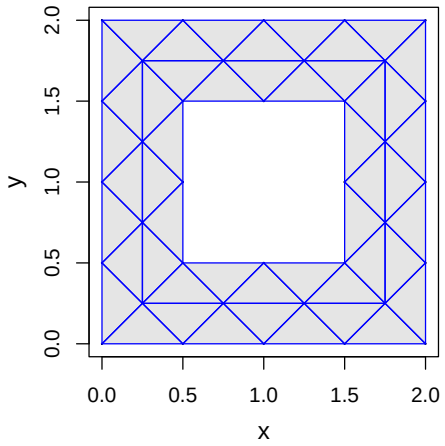


图 2：区域 Ω（灰色部分）及其三角剖分。

在模拟研究中，我们考虑了四种设置：i) $\mu_2(t) = \cos(2\pi t/12) + t/n$，AR(2)系数为 $k_{1,1} = k_{1,2} = 0.8$ 和 $k_{2,1} = k_{2,2} = 0.1$；ii) $\mu_2(t) = \cos(2\pi t/12) + t/n$，AR(2)系数为 $k_{1,1} = k_{1,2} = k_{2,1} = k_{2,2} = 0$；iii) $\mu_2(t) = 1$，AR(2)系数为$k_{1,1} = k_{1,2} = 0.8$ 和 $k_{2,1} = k_{2,2} = 0.1$；iv) $\mu_2(t) = 1$，AR(2)系数为 $k_{1,1} = k_{1,2} = k_{2,1} = k_{2,2} = 0$。注意，在设置 (ii) 和 (iv) 中，AR(2)模型退化为白噪声模型，因此函数主成分得分相互独立。在每种设置中，我们采用两个方差水平：$\sigma^2 = 1$，$(\sigma_1^2, \sigma_2^2) = (1, 0.1)$；或 $\sigma^2 = 0.1$，$(\sigma_1^2, \sigma_2^2) = (0.1, 0.01)$。为模拟一个数据集，我们将时间点数设为 $n = 500$。在每个时间 t, $t = 1, \cdots, 500$，观测位置的数量从 $\{50, 51, \ldots, 60\}$ 均匀抽取，且每个位置均从不规则区域上的均匀分布中随机抽取。对于四种设置与两个方差水平的每种组合，我们进行了 100 次模拟。对每个模拟数据集，均应用了所提出的 sFPC 模型和 mFPC 模型。

对于 sFPC和mFPC模型，在空间域 Ω 上，我们使用了图 2所示的相同三角剖分。二元基函数由伯恩斯坦基多项式构造，其中 $d = 3$ （三次样条）和 $r = 1$ （在相连边上具有连续一阶导数），与 Zhou andPan（2014年b）中的方法相同。有关基函数构造的详情，请参见附录B 节。

在时间域上，为使用 sFPC模型估计 $\mu_2(t)$，我们使用了 14 个基函数，

设置	σ^2, (σ_1^2, σ_2^2)	模型	PA	MIAE $\mu_t(x, y)$	MIAE $Z_t(x, y)$
i)	1.0, (1.0, 0.1)	sFPC	4.6283 (0.0759)	0.1001 (0.0052)	0.1388 (0.0003)
		mFPC	6.6644 (0.2103)	0.2223 (0.0068)	0.1833 (0.0010)
	0.1, (0.1, 0.01)	sFPC	4.6259 (0.0756)	0.0324 (0.0017)	0.0436 (0.0001)
		mFPC	6.9333 (0.2605)	0.0771 (0.0027)	0.0585 (0.0003)
ii)	1.0, (1.0, 0.1)	sFPC	9.6592 (0.1411)	0.0376 (0.0009)	0.1384 (0.0003)
		mFPC	9.7682 (0.1659)	0.0713 (0.0009)	0.1498 (0.0003)
	0.1, (0.1, 0.01)	sFPC	9.7135 (0.1419)	0.0126 (0.0003)	0.0439 (0.0001)
		mFPC	11.908 (0.5291)	0.0264 (0.0003)	0.0487 (0.0002)
iii)	1.0, (1.0, 0.1)	sFPC	6.9663 (0.0872)	0.1052 (0.0045)	0.1418 (0.0003)
		mFPC	7.0707 (0.0884)	0.0851 (0.0045)	0.1538 (0.0003)
	0.1, (0.1, 0.01)	sFPC	4.5682 (0.0751)	0.0414 (0.0024)	0.0430 (0.0001)
		mFPC	4.6599 (0.0771)	0.0502 (0.0030)	0.0473 (0.0001)
iv)	1.0, (1.0, 0.1)	sFPC	13.593 (0.1100)	0.0316 (0.0012)	0.1430 (0.0003)
		mFPC	13.582 (0.1092)	0.0302 (0.0010)	0.1425 (0.0003)
	0.1, (0.1, 0.01)	sFPC	8.9945 (0.1497)	0.0110 (0.0004)	0.0434 (0.0001)
		mFPC	8.9911 (0.1494)	0.0110 (0.0004)	0.0433 (0.0001)

表1：均值函数 $\mu_t(x, y)$ 和随机曲面 $Z_t(x, y)$ 的 PA 和 MIAE 均值及标准误。结果基于 100 次模拟运行。

其中包括用于建模趋势的三次多项式，以及用于建模季节性的傅里叶基函数 $\sin(2k\pi t/12)$、$\cos(2k\pi t/12)$、$k = 1, \cdots, 5$。

由于 mFPC 模型假设均值函数在 t上为常数，因此，为了在设置 i) 和 ii) 中应用 mFPC 模型，我们采用了两步法：1) 通过基展开估计均值函数$\mu_t(x, y)$，以及 2) 使用 mFPC 模型拟合残差。为了构造第 1) 步中的基函数，我们首先忽略位置效应，并通过将数据对 sFPC 模型所用的 14 个时间区域基函数进行回归，构造总体时间效应的估计，记为 $\tilde{\nu}(t)$。接下来，我们将 $\tilde{\nu}(t)$与空间区域 Ω 上的二元基函数相乘，以生成基函数；然后将数据对这些基函数进行回归，从而估计 $\mu_1(x, y)\mu_2(t)$。在设置 iii) 和 iv) 中，由于真实均值函数

在 t上为常数，符合 mFPC 模型的假设，因此直接将 mFPC 模型应用于模拟数据。

为简化起见，在模拟研究中拟合模型时，主成分的数量和自回归的阶数均设为与真实值相同。我们通过单纯形法最小化第λ_{μ_s}*3.3* λ_{μ_t}节 λ_{pc}所述的五折留位置交叉验证值，来选取三个惩罚参数（λ_{μ_s}，λ_{μ_t}，λ_{pc}）。

为了定量衡量均值函数和随机曲面的估计性能，我们采用平均积分绝对误差（MIAE）定义如下：

$$\frac{1}{n}\sum_{t=1}^{n}\int_{\Omega}|\mu_t(x,y)-\widehat{\mu}_t(x,y)|\mathrm{d}x\mathrm{d}y \quad \text{and} \quad \frac{1}{n}\sum_{t=1}^{n}\int_{\Omega}|Z_t(x,y)-\widehat{Z}_t(x,y)|\mathrm{d}x\mathrm{d}y,$$

其中，每个积分均通过空间域内均匀分布的 1976 个网格点上的缩放求和来计算（网格点为在矩形上均匀分布的 51×51 个点，并去除了洞内的点）。

由于主成分函数被限制为单位范数，且其符号无法唯一确定（±1），我们采用主角度（PA）评估主成分函数的估计性能。主角度（PA）的计算方法如下：首先，在区域内均匀分布的 1976 个网格$\widehat{}$点上计算主成分函数，得到两个矩阵 $\mathbf{V}$ 和 $\mathbf{V}$，分别对应估计的和真实的主成分函数。然后，计算主角度为角度 $=\cos^{-1}(\rho)\times180/\pi$，其中 ρ 是矩阵 $\mathbf{Q}_{\widehat{\mathbf{V}}}^{\mathsf{T}}\mathbf{Q}_{\mathbf{V}}$ 的最小奇异值，其中 $\mathbf{Q}_{\widehat{\mathbf{V}}}$ 和 $\mathbf{Q}_{\mathbf{V}}$ 分别是对 V 和 V进行 QR 分解得到的正交归一矩阵（Golub和Van Loan，2013年）。

100 次模拟运行中，均值函数的 MIAE 均值和标准误 $\mu_t(x,\ y)$ 与随机曲面 $z_t(x,\ y)$ 的 MIAE 均值和标准误，以及主成分函数的预测准确度（ PA ）汇总于表 1。从表中可以清楚地看出，在设置 i) 中，与 mFPC 方法相比，sFPC 方法使均值函数和随机曲面的 MIAE 分别降低了 55%–58% 和 24%–26%；在主成分函数估计准确度方面，sFPC 模型的 PA 提高了约 30%–33%。在设置 ii) 中，sFPC 在均值函数和主成分函数估计方面均有所改善，MIAE 分别降低了 47%–52%，PA 则降低了 1%–18%。在随机曲面估计方面，sFPC 模型略优于 mFPC 模型，MIAE 降低了 8%–10%。这并不令人意外

因为设置 ii) 假设函数主成分得分之间不存在序列相关，这有利于 mFPC。在设置 iii) 中，在两个方差水平下，sFPC 方法估计均值函数的 MIAE 分别低 17% 和高 24%；这是因为我们在此设置中令 $\mu_2(t) = 1$，符合 mFPC 模型关于时间效应的假设，而 sFPC 模型则使用非参数方法估计这个常数 $\mu_2(t)$。然而，与 mFPC 方法相比，sFPC 方法在随机曲面估计方面仍有 8%–9% 的改进，PA 也降低了约 2%。在设置 iv) 中，真实模型有利于 mFPC，两种模型在均值函数、主成分和随机曲面的估计方面结果相近。

5 得克萨斯州温度数据分析

在本节中，我们通过分析从美国历史气候学网络第二版（USHCN v2.5, https://cdiac.ess-dive.lbl.gov/epubs/ndp/ushcn/ushcn.html）下载的德克萨斯州温度数据，应用所提出的模型来研究得克萨斯州的气候变化。该数据集包含 1915年1月至 2014 年 12 月期间得克萨斯州 49 个气象站记录的月平均气温。这些气象站的位置如图 3 所示。

得克萨斯州面积为 696,200 km^2, ，拥有多样的地形和气候类型。总体而言，得克萨斯州东半部的墨西哥湾沿岸平原和中北部平原属于湿润亚热带气候；西半部的沙漠和高耸山脉则属于半干旱气候。由于各种原因，6.82% 的数据缺失。具体而言，只有 3 个站点的记录完整，而 13 个站点缺失的数据超过 120 个月。数据缺失没有明显规律。

令 $W_t(x, y)$ 表示位置（x, y）和时间 t 处的原始数据。在应用所提出的 sFPC 模型之前，我们首先通过对原始数据拟合非参数回归，估计位置和时间主效应，即 $W_t(x, y) = \mu(x, y) + \nu(t) + \epsilon_t(x, y) = \mathrm{b}(x, y)^\top \boldsymbol{\theta}_\mu + \mathrm{c}(t)^\top \boldsymbol{\theta}_\nu + \epsilon_t(x, y)$，其中 $\mathrm{b}(x, y)$ 和 $\mathrm{c}(t)$ 分别是定义在位置域和时间域上的基函数。具体而言，二元基函数 $\mathrm{b}(x, y)$ 由伯恩斯坦多项式构造，$d = 3$ and $r = 1$ 位于图 3 所示的三角剖分上。该三角剖分也用于 Zhou and Pan (2014b)，覆盖得克萨斯州的不规则区域

且每个三角形内至少有一个气象站。在时间域上，考虑到季节效应以及气候随时间的变化，我们构造了时间基函数 $\mathbf{c}(t) = \{\mathbf{c}_1(t)^\top, \mathbf{c}_2(t)^\top\}^\top$，其中 $\mathbf{c}_1(t)$ 是三次样条，在 1940 年 12 月、1965 年和 1990 年各设一个内部节点，用于拟合趋势；$\mathbf{c}_2(t)$ 是傅里叶函数 $\sin(2k\pi t/12)$、$\cos(2k\pi t/12)$、$k = 1, \cdots, 5$，用于拟合季节效应。我们采用惩罚最小二乘法估计参数，并通过 5 折留位置交叉验证选择惩罚参数；在每个时间点，将样本位置随机划分为 ^ ^个大小相等的子集。令 $\boldsymbol{\theta}_\mu$ 和 $\boldsymbol{\theta}_\nu$ 分别为 $\boldsymbol{\theta}_\mu$ 和 $\boldsymbol{\theta}_\nu$的估计值。估计得到的位置和时间主效应分别为 $\widehat{\mu}(x, y) = \mathbf{b}(x, y)^\top \widehat{\boldsymbol{\theta}}_\mu$ 和$\widehat{\nu}(t) = \mathbf{c}(t)^\top \widehat{\boldsymbol{\theta}}_\nu$。接下来，我们从原始数据中减去估计的主效应，得到去均值数据 $Z_t(x, y) = W_t(x, y) - \widehat{\mu}(x, y) - \widehat{\nu}$。

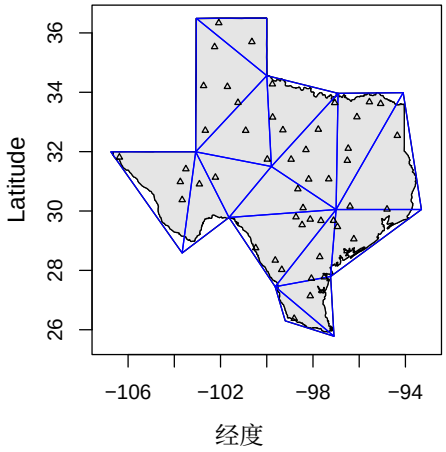


图 3：德克萨斯州温度数据分析应用中使用的三角剖分。

按照 sFPC 方法，使用模型（8）拟合了 $Z_t(x, y)$，所用基函数 $\mathbf{b}(x, y)$ 和 $\mathbf{c}(t)$ 与上一段所述相同。如第3.3节所述，在检查时间FPC得分的方差经验占比后，我们使用了 $J = 3$ 个主成分，并将自回归模型的阶数设为使赤池信息准则最小的 $p = 4$。三个惩罚参数 λ_{μ_s}、λ_{μ_t} 和 λ_{pc} 按照第3.3节所述，通过单纯形法进行5折留位置交叉验证选定。我们还使用 Zhou 和 Pan（2014a）中的 mFPC 方法拟合了 $Z_t(x, y)$，以作比较。表2列出了两种方法对每个月份的（样本内）残差计算所得的平均绝对误差（MAEs），两者相近，sFPC方法的平均绝对误差为 mFPC 方法的94%至104%

。然而，sFPC 和 mFPC 的交叉验证误差分别为 $\mathrm{CV}_{\mathrm{sFPC}} = 0.5679$ 和 $\mathrm{CV}_{\mathrm{mFPC}} = 1.999$。所提出的 sFPC 方法交叉验证误差更小，表明其在该真实数据上的预测性能更优。

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
sFPC	0.551	0.563	0.584	0.550	0.522	0.533	0.512	0.522	0.517	0.499	0.510	0.523
mFPC	0.571	0.568	0.560	0.570	0.536	0.548	0.520	0.525	0.517	0.509	0.526	0.556

表2：sFPC和mFPC方法对每个月残差的平均绝对误差。

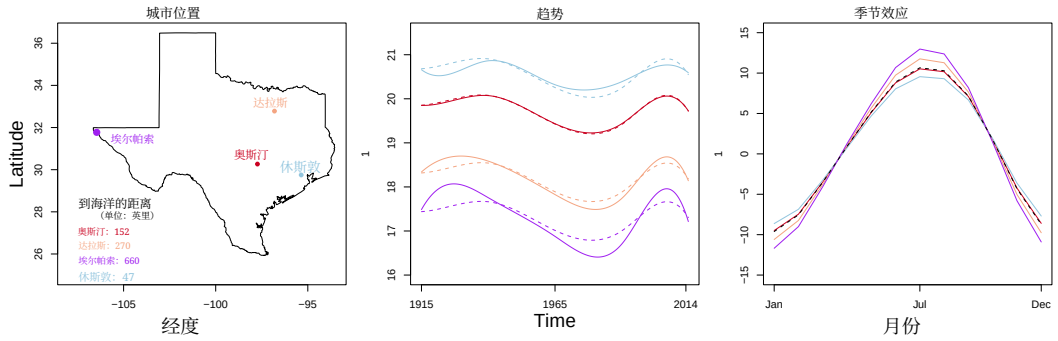


图4：左：得克萨斯州奥斯汀、达拉斯、埃尔帕索和休斯敦的位置及其与海洋的距离。中：两种方法估计的1915年1月至2014年12月期间的趋势。右：两种方法估计的季节效应。在中图和右图中，实线和虚线分别表示使用sFPC和mFPC方法得到的估计结果，颜色对应左图中的城市。

记$\boldsymbol{\theta}_\nu = \{\boldsymbol{\theta}_{\nu1}^\mathsf{T}, \boldsymbol{\theta}_{\nu2}^\mathsf{T}\}^\mathsf{T}$和$\boldsymbol{\theta}_c = \{\boldsymbol{\theta}_{c1}^\mathsf{T}, \boldsymbol{\theta}_{c2}^\mathsf{T}\}^\mathsf{T}$，其中 $\boldsymbol{\theta}_{\nu1}$和 $\boldsymbol{\theta}_{c1}$对应于基函数$\mathsf{c}_1(t)$，而 $\boldsymbol{\theta}_{\nu2}$和 $\boldsymbol{\theta}_{c2}$对应于基函数$\mathsf{c}_2(t)$。在得克萨斯州位置（x, y）和1915年1月至2014年12月之间的时间 t，sFPC方法估计的均值函数 $\widehat{\mu}(x,y)+\widehat{\nu}(t)+\widehat{\mu}_1(x,y)\widehat{\mu}_2(t)$可进一步分解为估计趋势 $\widehat{\mu}(x,y)+\mathsf{c}_1(t)^\mathsf{T}\widehat{\boldsymbol{\theta}}_{\nu1} + \widehat{\mu}_1(x,y)\mathsf{c}_1(t)^\mathsf{T}\widehat{\boldsymbol{\theta}}_{c1}$和估计季节效应$\mathsf{c}_2(t)^\mathsf{T}\widehat{\boldsymbol{\theta}}_{\nu2} + \widehat{\mu}_1(x,y)\mathsf{c}_2(t)^\mathsf{T}\widehat{\boldsymbol{\theta}}_{c2}$。图4以实线展示了sFPC方法估计的奥斯汀、达拉斯、埃尔帕索和休斯敦在1915年1月至2014年12月之间的趋势和季节效应。可以看出，这四座城市的趋势和季节性模式相似。温度与纬度和海拔均有关，而季节变化随着距

海洋距离的增加而增大。类似地，我们用mFPC方法将估计的均值函数分解为趋势和季节效应，并在图4 中以虚线展示。

为展示通过 sFPC 和 mFPC 方法估计的位置变化的主要模式，我们在图 5中绘制了估计的二维主成分函数（曲面）热图及其等高线。根据这两种方法，估计得到的第一个主成分曲面在得克萨斯州的大部分区域变化很小，但经度 $-104°$以西的区域除外，这可能是该地区气象站数量稀少造成的假象。对于 sFPC 方法，估计得到的第二个主成分曲面的等高线大致与纬度平行，而第三个估计主成分曲面的等高线大致与经度平行，这可能与海拔变化有关。对于 mFPC 方法，估计得到的第二个主成分曲面的等高线可能受到墨西哥湾和中北部平原的影响，而第三个估计主成分曲面与 sFPC 方法的结果相似。我们进一步采用半参数自助法来量化估计的主成分函数的不确定性。详细步骤及自助法标准差曲面图见附录第 D 节。

我们还比较了 sFPC 和 mFPC 模型在短期预测方面的表现：利用 1915 年至 2013 年的数据拟合模型，并预测 2014 年的月平均气温。对于 sFPC 和 mFPC，我们均按照前文所述的相同流程拟合数据，并使用相同数量的主成分（$J = 3$）和相同阶数的自回归模型（$p = 4$）。使用 sFPC 模型预测 2014 年的月气温时，我们将均值函数的基函数外推至时间域，并根据拟合的自回归模型预测主成分得分。使用 mFPC 模型时，我们将估计的平均气温作为预测值。表3展示了 49 个气象站在 2014 年 12 个月的平均绝对预测误差（平均绝对预测误差）。与 mFPC方法相比，sFPC方法使预测误差的平均绝对预测误差降低了 2%–39%；但在十月，sFPC方法的平均绝对预测误差略微增加了 3%。

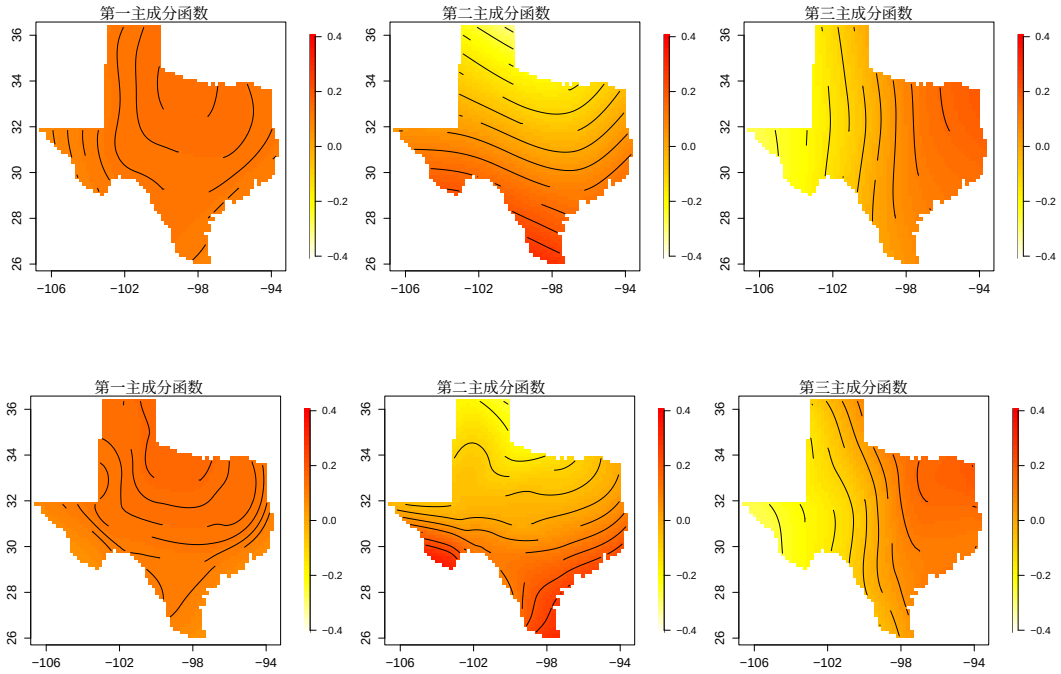


图5：真实数据分析中使用两种方法估计得到的前三个主成分函数。第一行和第二行分别对应 sFPC 和 mFPC 模型。从左到右依次为估计得到的第一、第二和第三个主成分函数。

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
sFPC	1.43	1.50	1.75	0.64	1.02	1.01	0.78	0.80	0.75	1.80	2.50	1.53
mFPC	1.74	1.80	1.88	0.74	1.25	1.37	1.27	1.11	0.82	1.74	2.55	1.60

表3：两种方法的 2014年1月至12月两种方法的平均绝对预测误差 .

6 总结与讨论

本文提出了一种名为 sFPC 的新方法，用于分析具有序列相关的二维函数型数据。为刻画数据的位置效应和时间效应，我们联合使用三角形上的双变量样条以及针对潜在函数主成分得分的自回归模型。我们结合卡尔曼滤波与平滑器开发了一种 EM算法，以计算 E步中的期望值。借助潜在自回归模型和 EM算法，所提出的 sFPC方法能够处理曲面定义域上的稀疏性和不规则性，同时也

作为时间域上的缺失观测。模拟研究和真实数据分析表明，在假设潜在得分相互独立的情况下，sFPC方法优于mFPC方法。

当前模型的一个局限是，假设所有噪声均服从高斯分布。然而，在许多实际问题中，这一假设往往不成立。因此，未来的一个研究方向是将高斯分布替换为其他分布（例如指数分布族），以推广我们的方法。此外，假设每个主成分对应的潜在函数主成分得分都服从平稳的AR(p)过程，这一假设也过于严格。值得进一步研究更一般的模型，以扩展我们的方法；在该模型中，潜在函数主成分得分服从可能非平稳的ARIMA过程。

伦理声明

作者声明不存在任何利益冲突。

附录

一个导数 完整数据对数似然的推导 d

将随机效应 $\boldsymbol{\alpha}_i$ 视为潜变量，负二倍对数似然为

$$-2l_c(\Xi; \{\mathbf{z}_t\}_{t=1}^n, \{\boldsymbol{\alpha}_t\}_{t=1}^n) = -2\log p(\mathbf{z}_1, \ldots, \mathbf{z}_n, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n).$$

联合概率 $p(\mathbf{z}_1, \ldots\ldots, \mathbf{z}_n, \boldsymbol{\alpha}_1, \ldots\ldots, \boldsymbol{\alpha}_n)$ 可分解为概率密度 $p(\boldsymbol{\alpha}_1, \ldots\ldots, \boldsymbol{\alpha}_n)$ 与给定潜变量时观测值的似然 $p(\mathbf{z}_1, \ldots\ldots, \mathbf{z}_n|\boldsymbol{\alpha}_1, \ldots\ldots, \boldsymbol{\alpha}_n)$ 的乘积。因此，负二倍对数似然可写为

$$\begin{aligned} &-2\log p(\mathbf{z}_1, \ldots, \mathbf{z}_n, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n) \\ &\quad = -2\log p(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n) - 2\log p(\mathbf{z}_1, \ldots, \mathbf{z}_n|\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n), \end{aligned} \tag{A.1}$$

自回归（p）时间序列 构表明，式（A.1）中的第一部分可分解为以下各项之和 t

初始状态 $\boldsymbol{\alpha}_1$、……、 $\boldsymbol{\alpha}_p$和后续状态 $\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1}$、……、 $\boldsymbol{\alpha}_{t-p}$，其中 $t=p+1$、……、 n

$$\begin{aligned}-2\log p(\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_n) &= -2\log p(\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_p) - 2\log p(\boldsymbol{\alpha}_{p+1},\ldots,\boldsymbol{\alpha}_n|\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_p)\\ &= -2\log p(\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_p) - 2\sum_{t=p+1}^{n}\log p(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1},\ldots,\boldsymbol{\alpha}_{t-p}).\end{aligned} \tag{A.2}$$

(A.2的第二项具有如下显式表达式

$$\begin{aligned}&-2\sum_{t=p+1}^{n}\log p(\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1},\ldots,\boldsymbol{\alpha}_{t-p})\\ &\quad=\sum_{j=1}^{J}\sum_{t=p+1}^{n}\left\{\log\sigma_j^2+\frac{1}{\sigma_j^2}(\alpha_{j,t}-k_{1,j}\alpha_{j,t-1}-\cdots-k_{p,j}\alpha_{j,t-p})^2\right\},\end{aligned}$$

这来自条件概率 $\boldsymbol{\alpha}_t|\boldsymbol{\alpha}_{t-1},\ldots,\boldsymbol{\alpha}_{t-p}\sim \mathrm{N}(\sum_{i=1}^{p}\mathbf{K}_i\boldsymbol{\alpha}_{t-i},\mathbf{H}_J)$，其中 $\mathbf{K}_i=\mathrm{diag}(k_{i,1},\ldots,k_{i,J})$，$i=1,\ldots,p$。根据 Box 等人 (2015 年)，(A.2 的第一项可推导为

$$\begin{aligned}-2\log p(\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_p) &= -2\sum_{j=1}^{J}\log p(\alpha_{j,1},\ldots,\alpha_{j,p})\\ &=\sum_{j=1}^{J}\left\{p\log\sigma_j^2-\log|\mathbf{M}_j|+\frac{1}{\sigma_j^2}S_{pj}(\mathbf{k}_j)\right\},\end{aligned}$$

其中 $\mathbf{M}_j$是 $(\alpha_{j,1}/\sigma_j,\ldots,\alpha_{j,p}/\sigma_j)^{\mathsf{T}}$的精度矩阵，使得

$$\mathbf{M}_j=\begin{pmatrix}\gamma_{0,j} & \gamma_{1,j} & \cdots & \gamma_{p-1,j}\\ \gamma_{1,j} & \gamma_{0,j} & \cdots & \gamma_{p-2,j}\\ \vdots & \vdots & \ddots & \vdots\\ \gamma_{p-1,j} & \gamma_{p-2,j} & \cdots & \gamma_{0,j}\end{pmatrix}^{-1}\in\mathbb{R}^{p\times p} \tag{A.3}$$

其中 $\gamma_{i,j}=\mathbb{E}(\alpha_{j,1+i}\alpha_{j,1})/\sigma_j^2=\mathbb{E}(\alpha_{j,1}\alpha_{j,1+i})/\sigma_j^2$, $i=0,\ldots,p-1$和 $j=1,\ldots,J$（取决于 $\mathbf{k}_j=(k_{1,j},\ldots,k_{p,j})^{\mathsf{T}}$）；且 $S_{pj}(\mathbf{k}_j)=\tilde{\boldsymbol{\alpha}}_j^{\mathsf{T}}\mathbf{M}_j\tilde{\boldsymbol{\alpha}}_j$是残差平方和，其中有 $\tilde{\boldsymbol{\alpha}}_j=(\alpha_{j,1},\ldots,\alpha_{j,p})^{\mathsf{T}}$。因此，(A.1) 的第一部分为

$$-2\log p(\boldsymbol{\alpha}_1,\ldots,\boldsymbol{\alpha}_n)=\sum_{j=1}^{J}\left\{n\log\sigma_j^2-\log|\mathbf{M}_j|+\frac{1}{\sigma_j^2}S_j(\mathbf{k}_j)\right\},$$

其中残差平方和总和为

$$S_j(\mathbf{k}_j)=S_{pj}(\mathbf{k}_j)+\sum_{t=p+1}^{n}(\alpha_{j,t}-k_{1,j}\alpha_{j,t-1}-\cdots-k_{p,j}\alpha_{j,t-p})^2=(1,\mathbf{k}_j^{\mathsf{T}})^{\mathsf{T}}\mathbf{D}_j(1,\mathbf{k}_j^{\mathsf{T}}),$$

with

$$\mathbf{D}_j = \begin{pmatrix} D_{1,1,j} & -D_{1,2,j} & -D_{1,3,j} & \dots & -D_{1,p+1,j} \\ -D_{1,2,j} & D_{2,2,j} & D_{2,3,j} & \dots & D_{2,p+1,j} \\ \vdots & \vdots & \vdots & & \vdots \\ -D_{1,p+1,j} & D_{2,p+1,j} & D_{3,p+1,j} & \dots & D_{p+1,p+1,j} \end{pmatrix},$$

以及第（i， k）个元素 $D_{i,\ k,\ j} = D_{k,\ i,\ j} = \alpha_{j,\ i}\alpha_{j,\ k} + \alpha_{j,\ i+1}\alpha_{j,\ k+1} + \cdots + \alpha_{j,\ n+1-k}\alpha_{j,\ n+1-i}$。更多详情请参阅Box 等人（2015年）中附录 A7.4。根据观测变量 $\mathbf{z}_t$ 与潜变量 $\boldsymbol{\alpha}_t$ 之间的关系，完全数据对数似然（A.1）的第二部分，即给定潜变量时观测数据的似然，可写为

$$\begin{aligned} &-2\log\{p(\mathbf{z}_1,\dots,\mathbf{z}_n|\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_n)\} \\ &\quad= -2\sum_{t=1}^{n}\log p(\mathbf{z}_t|\boldsymbol{\alpha}_t) \\ &\quad= \sum_{t=1}^{n} n_t \log\sigma^2 + \frac{1}{\sigma^2}\sum_{t=1}^{n}(\mathbf{z}_t - \mathbf{B}_t\boldsymbol{\theta}_b\boldsymbol{\theta}_c^{\mathsf{T}}\mathbf{c}_t - \mathbf{B}_t\boldsymbol{\Theta}\boldsymbol{\alpha}_t)^{\mathsf{T}}(\mathbf{z}_t - \mathbf{B}_t\boldsymbol{\theta}_b\boldsymbol{\theta}_c^{\mathsf{T}}\mathbf{c}_t - \mathbf{B}_t\boldsymbol{\Theta}\boldsymbol{\alpha}_t). \end{aligned}$$

因此，完全数据对数似然可写成主论文中的（9）。

B 三角剖分上的双变量样条基函数

本节讨论二维正交归一基函数 $\mathsf{b}(x, y)$ 的构造。一种简单的选择是张量积 B 样条基函数，即 $\mathsf{b}(x, y) = \mathsf{b}_1(x) \otimes \mathsf{b}_2(y)$。然而，张量积 B 样条基函数会带来两个问题：(1) 由于张量积基函数数量众多，计算成本通常很高；(2) 该基函数只能用于矩形等规则区域。为解决这些问题，我们转而引入三角剖分上的双变量伯恩斯坦多项式。Lai and Schumaker（2007）阐述了双变量样条的数学性质。Zhou 和 Pan（2014a）将这类双变量样条应用于其 mFPC 模型。由于伯恩斯坦多项式具有出色的逼近性质（Liu 等人，2016年），它们在空间统计模型中得到了广泛应用（Yu 等人，2020年；Wang 等人，2020a年）。主论文的图 2 和图 3 展示了模拟研究和德克萨斯州温度数据分析中将使用的三角剖分示例。正如我们所见，与

常用的单变量基函数张量积不同，这种二元基函数可以轻松处理 $\mathbb{R}^2$ 上的不规则形状。

记 δ 为一个三角形，其顶点按逆时针方向排列为（$\mathbf{v}_1$,$\mathbf{v}_2$,$\mathbf{v}_3$）。那么，对于任意点$\mathbf{v}$ $\in \mathbb{R}^2$，都存在唯一的表示形式$\mathbf{v} = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + b_3\mathbf{v}_3$。这三个系数（$b_1$，$b_2$，$b_3$）称为$\mathbf{v}$ 相对于三角形δ的重心坐标。给定一个非负整数 d，对于任意满足 $i + j + k = d$的 i，j，k，相对于三角形 δ 定义的 d 次伯恩斯坦多项式为

$$B^d_{ijk}(\mathbf{v}) = \frac{d!}{i!j!k!} b_1^i b_2^j b_3^k.$$

设 $P_d(\delta)$为定义在三角形 δ 上的 d次多项式空间。那么，伯恩斯坦多项式 B^d_{ijk}，$i + j + k = d$构成 $P_d(\delta)$的一组基。也就是说，对于任意函数$s \in P_d(\delta)$，有

$$s(\mathbf{v}) = \sum_{i+j+k=d} \gamma_{ijk} B^d_{ijk}(\mathbf{v}).$$

对于不规则区域，我们可以构造一个三角剖分 $\Delta = \{\delta_1, \ldots, \delta_M\}$，其并集覆盖不规则区域 Ω（例如，参见 Lai and Schumaker，2007）。我们针对每个 δ_i构造伯恩斯坦多项式基函数，所有此类多项式构成 $P_d(\Delta)$的一组基；该空间由定义在 d 上的 Δ次连续分片多项式构成。若进一步要求直到 r 阶的导数连续，则可相应地构造二元基函数。关于施加此类光滑性条件以及应用格拉姆-施密特过程以确保基函数$\mathbf{b}(x, y)$满足正交归一条件的详细内容，请参见 Zhou and Pan (2014b)。

C M步的详细内容

在 M步 中，我们求出主论文中式 ($Q(\Xi|\Xi^{(0)})$ 的极小化解 (13)。请注意，在式 (13) 中，参数 $\boldsymbol{\theta}_b$、$\boldsymbol{\theta}_c$、$\boldsymbol{\Theta}$、$\mathbf{K}$ σ^2 和 $\{\sigma_j^2\}_{j=1}^J$彼此分离。因此，我们提出采用分块优化，即在每次迭代中，在其他参数保持当前值不变的情况下更新每个参数。

C.1 更新 $\boldsymbol{\theta}_b$

回想一下，更新 $\boldsymbol{\theta}_b$ 等价于求解以下目标函数的极小化解：

$$f(\boldsymbol{\theta}_b) = (\boldsymbol{\theta}_b - \mathbf{m})^{\mathsf{T}} \mathbf{A} (\boldsymbol{\theta}_b - \mathbf{m}), \tag{C.1}$$

满足约束 $\boldsymbol{\theta}_b^{\mathsf{T}} \boldsymbol{\theta}_b = 1$，其中

$$\mathbf{m} = \left\{ \sum_{t=1}^{n} (\boldsymbol{\theta}_c^{(0)\mathsf{T}} \mathbf{c}_t)^2 \mathbf{B}_t^{\mathsf{T}} \mathbf{B}_t + \sigma^{2(0)} \lambda_{\mu_s} \boldsymbol{\Gamma} \right\}^{-1} \sum_{t=1}^{n} (\boldsymbol{\theta}_c^{(0)\mathsf{T}} \mathbf{c}_t) \mathbf{B}_t^{\mathsf{T}} (\mathbf{z}_t - \mathbf{B}_t \boldsymbol{\Theta}^{(0)} \widehat{\boldsymbol{\alpha}}_t)$$

and

$$\mathbf{A} = \sum_{t=1}^{n} (\boldsymbol{\theta}_c^{(0)\mathsf{T}} \mathbf{c}_t)^2 \mathbf{B}_t^{\mathsf{T}} \mathbf{B}_t + \sigma^{2(0)} \lambda_{\mu_s} \boldsymbol{\Gamma}.$$

我们提出在球面子流形上使用梯度下降算法，以求得（C.1）的极小化解。子流形上的梯度下降算法包含四个迭代步骤：

i 在欧几里得空间中不施加任何约束，计算目标函数的负梯度；

ii 将得到的负梯度函数投影到流形的切空间上；

iii 沿投影负梯度的方向评估更新值，在
步骤 ii：给定一个步长；

iv 将步骤 iii 中计算出的值映射回流形结构。

上述步骤 ii 中的步长可以使用 Armijo 回溯法确定（例如，参见Absil等人，2009年第 4.2 章）。

从给定的初始值 $\widehat{\boldsymbol{\theta}}_b^{(0)}$ 开始，我们在第 $\widehat{\boldsymbol{\theta}}_b^{(k-1)}$ 次 k迭代中更新 $\widehat{\boldsymbol{\theta}}_b^{(k)}$，$k = 1$、2、.....，直到满足收敛条件。具体而言，在第 k次迭代中，我们首先计算（C.1）中目标函数的负梯度，其显式表达式为

$$-\frac{\partial f(\boldsymbol{\theta}_b)}{\partial \boldsymbol{\theta}_b} = -2\mathbf{A}(\boldsymbol{\theta}_b - \mathbf{m}).$$

令 $\boldsymbol{\eta}_k$ 表示球面切空间上的投影负梯度函数。根据Absil等人(2009年)，可以证明

$$\boldsymbol{\eta}_k = \mathrm{Proj}\left\{ -\frac{\partial f(\boldsymbol{\theta}_b)}{\partial \boldsymbol{\theta}_b} \right\} = -2\left\{ \mathbf{A}(\widehat{\boldsymbol{\theta}}_b^{(k-1)} - \mathbf{m}) - \widehat{\boldsymbol{\theta}}_b^{(k-1)} \left(\widehat{\boldsymbol{\theta}}_b^{(k-1)}\right)^{\mathsf{T}} \mathbf{A}(\widehat{\boldsymbol{\theta}}_b^{(k-1)} - \mathbf{m}) \right\}.$$

算法 C.1 用于更新 $\boldsymbol{\theta}_b$ 的梯度下降算法。

要求： 标量 β， $\gamma \in (0, 1)$ 和初始化 $\widehat{\boldsymbol{\theta}}_b^{(0)}$。

for $k = 1, 2, \ldots$ **do**

i) Compute $\boldsymbol{\eta}_k = -2\{\mathbf{A}(\widehat{\boldsymbol{\theta}}_b^{(k-1)} - \mathbf{m}) - \widehat{\boldsymbol{\theta}}_b^{(k-1)}(\widehat{\boldsymbol{\theta}}_b^{(k-1)})^{\mathsf{T}}\mathbf{A}(\widehat{\boldsymbol{\theta}}_b^{(k-1)} - \mathbf{m})\}$.

ii) Find the smallest integer $n \geq 0$ such that $f\{R_{\widehat{\boldsymbol{\theta}}_b^{(k-1)}}(\beta^n \boldsymbol{\eta}_k)\} \leq f(\widehat{\boldsymbol{\theta}}_b^{(k-1)}) - \gamma\beta^n \boldsymbol{\eta}_k^{\mathsf{T}}\boldsymbol{\eta}_k$, where $R_{\widehat{\boldsymbol{\theta}}_b^{(k-1)}}(\beta^n \boldsymbol{\eta}_k) = (\widehat{\boldsymbol{\theta}}_b^{(k-1)} + \beta^n \boldsymbol{\eta}_k)/\|\widehat{\boldsymbol{\theta}}_b^{(k-1)} + \beta^n \boldsymbol{\eta}_k\|$.

iii) $\widehat{\boldsymbol{\theta}}_b^{(k)} = R_{\widehat{\boldsymbol{\theta}}_b^{(k-1)}}(\beta^n \boldsymbol{\eta}_k)$.

iv) Repeat until $\|\widehat{\boldsymbol{\theta}}_b^{(k)} - \widehat{\boldsymbol{\theta}}_b^{(k-1)}\|$ is small enough.

end for

return $\widehat{\boldsymbol{\theta}}_b = \widehat{\boldsymbol{\theta}}_b^{(k)}$.

随后，我们沿投影负梯度方向评估更新值 $\boldsymbol{\eta}_k$，步长由 Armijo回溯法确定（例如参见 Absil等人， 2009年的第 4.2 章）。具体而言，设 β， $\gamma \in (0, 1)$ 为给定的预设参数，Armijo回溯法将步长参数化为 β^n 其中$n \geq 0$。指数 $n \geq 0$ 是满足以下条件的最小整数

$$f\{R_{\widehat{\boldsymbol{\theta}}_b^{(k-1)}}(\beta^n \boldsymbol{\eta}_k)\} \leq f(\widehat{\boldsymbol{\theta}}_b^{(k-1)}) - \gamma\beta^n \|\boldsymbol{\eta}_k\|_2^2$$

成立，其中 $R_{\widehat{\boldsymbol{\theta}}_b^{(k-1)}}(\cdot)$ 是回缩函数。对于球面，回缩函数可以选为归一化（Absil等人， 2009年），即

$$R_{\widehat{\boldsymbol{\theta}}_b^{(k-1)}}(\beta^n \boldsymbol{\eta}_k) = \frac{\widehat{\boldsymbol{\theta}}_b^{(k-1)} + \beta^n \boldsymbol{\eta}_k}{\|\widehat{\boldsymbol{\theta}}_b^{(k-1)} + \beta^n \boldsymbol{\eta}_k\|}.$$

最后，在确定 n 后，我们将 $\widehat{\boldsymbol{\theta}}_b^{(k-1)} + \beta^n \boldsymbol{\eta}_k$ 回缩到球面上，作为更新后的第 k 次迭代，即

$$\widehat{\boldsymbol{\theta}}_b^{(k)} = R_{\widehat{\boldsymbol{\theta}}_b^{(k-1)}}(\beta^n \boldsymbol{\eta}_k).$$

上述讨论引出了算法C.1，用于总结球面子流形上的梯度下降法。

C.2 更新 $\boldsymbol{\theta}_c$

对于 $\boldsymbol{\theta}_c$，我们需要最小化的相应目标函数为

$$\frac{1}{\sigma^{2(0)}}\sum_{t=1}^{n}(\mathbf{z}_t-\mathbf{B}_t\widehat{\boldsymbol{\theta}}_b\boldsymbol{\theta}_c^{\mathsf{T}}\mathbf{c}_t-\mathbf{B}_t\boldsymbol{\Theta}^{(0)}\widehat{\boldsymbol{\alpha}}_t)^{\mathsf{T}}(\mathbf{z}_t-\mathbf{B}_t\widehat{\boldsymbol{\theta}}_b\boldsymbol{\theta}_c^{\mathsf{T}}\mathbf{c}_t-\mathbf{B}_t\boldsymbol{\Theta}^{(0)}\widehat{\boldsymbol{\alpha}}_t)+\lambda_{\mu_t}\boldsymbol{\theta}_c^{\mathsf{T}}\mathbf{P}\boldsymbol{\theta}_c.$$

我们取 firs上述目标函数关于 $\boldsymbol{\theta}_c$ 的 t 阶导数，得到

$$\frac{2}{\sigma^{2(0)}}\sum_{t=1}^{n}(\mathbf{B}_t\widehat{\boldsymbol{\theta}}_b\mathbf{c}_t^{\mathsf{T}})^{\mathsf{T}}(\mathbf{z}_t-\mathbf{B}_t\widehat{\boldsymbol{\theta}}_b\boldsymbol{\theta}_c^{\mathsf{T}}\mathbf{c}_t-\mathbf{B}_t\boldsymbol{\Theta}^{(0)}\widehat{\boldsymbol{\alpha}}_t)+2\lambda\mu_t\mathbf{P}\boldsymbol{\theta}_c.$$

令上述表达式为零，即可将 $\boldsymbol{\theta}_c$ 更新为

$$\widehat{\boldsymbol{\theta}}_c=\Big\{\sum_{t=1}^{n}(\mathbf{B}_t\widehat{\boldsymbol{\theta}}_b\mathbf{c}_t^{\mathsf{T}})^{\mathsf{T}}(\mathbf{B}_t\widehat{\boldsymbol{\theta}}_b\mathbf{c}_t^{\mathsf{T}})+\sigma^{2(0)}\lambda_{\mu_t}\mathbf{P}\Big\}^{-1}\sum_{t=1}^{n}(\mathbf{B}_t\widehat{\boldsymbol{\theta}}_b\mathbf{c}_t^{\mathsf{T}})^{\mathsf{T}}(\mathbf{z}_t-\mathbf{B}_t\boldsymbol{\Theta}^{(0)}\widehat{\boldsymbol{\alpha}}_t).$$

C.3 更新 σ^2

对于 σ^2，我们需要最小化的相应目标函数为 is

$$\frac{1}{\sigma^2}\sum_{t=1}^{n}\big\{(\mathbf{z}_t-\mathbf{B}_t\widehat{\boldsymbol{\theta}}_b\widehat{\boldsymbol{\theta}}_c^{\mathsf{T}}\mathbf{c}_t-\mathbf{B}_t\boldsymbol{\Theta}^{(0)}\widehat{\boldsymbol{\alpha}}_t)^{\mathsf{T}}(\mathbf{z}_t-\mathbf{B}_t\widehat{\boldsymbol{\theta}}_b\widehat{\boldsymbol{\theta}}_c^{\mathsf{T}}\mathbf{c}_t-\mathbf{B}_t\boldsymbol{\Theta}^{(0)}\widehat{\boldsymbol{\alpha}}_t)$$
$$+\operatorname{tr}(\mathbf{B}_t\boldsymbol{\Theta}^{(0)}\widehat{\boldsymbol{\Sigma}}_t{\boldsymbol{\Theta}^{(0)}}^{\mathsf{T}}\mathbf{B}_t^{\mathsf{T}})\big\}+\sum_{t=1}^{n}n_t\log\sigma^2.$$

我们对上述目标函数关于 σ^2 求一阶导数，得到

$$-\frac{1}{(\sigma^2)^2}\Big\{(\mathbf{z}_t-\mathbf{B}_t\widehat{\boldsymbol{\theta}}_b\widehat{\boldsymbol{\theta}}_c^{\mathsf{T}}\mathbf{c}_t-\mathbf{B}_t\boldsymbol{\Theta}^{(0)}\widehat{\boldsymbol{\alpha}}_t)^{\mathsf{T}}(\mathbf{z}_t-\mathbf{B}_t\widehat{\boldsymbol{\theta}}_b\widehat{\boldsymbol{\theta}}_c^{\mathsf{T}}\mathbf{c}_t-\mathbf{B}_t\boldsymbol{\Theta}^{(0)}\widehat{\boldsymbol{\alpha}}_t)$$
$$+\operatorname{tr}(\mathbf{B}_t\boldsymbol{\Theta}^{(0)}\widehat{\boldsymbol{\Sigma}}_t{\boldsymbol{\Theta}^{(0)}}^{\mathsf{T}}\mathbf{B}_t^{\mathsf{T}})\Big\}+\frac{1}{\sigma^2}\sum_{t=1}^{n}n_t,$$

通过将上述表达式设为零，我们即可更新 σ^2 a s

$$\widehat{\sigma}^2=\frac{1}{\sum_{t=1}^{n}n_t}\sum_{t=1}^{n}\Big\{(\mathbf{z}_t-\mathbf{B}_t\widehat{\boldsymbol{\theta}}_b\widehat{\boldsymbol{\theta}}_c^{\mathsf{T}}\mathbf{c}_t-\mathbf{B}_t\boldsymbol{\Theta}^{(0)}\widehat{\boldsymbol{\alpha}}_t)^{\mathsf{T}}(\mathbf{z}_t-\mathbf{B}_t\widehat{\boldsymbol{\theta}}_b\widehat{\boldsymbol{\theta}}_c^{\mathsf{T}}\mathbf{c}_t-\mathbf{B}_t\boldsymbol{\Theta}^{(0)}\widehat{\boldsymbol{\alpha}}_t)$$
$$+\operatorname{tr}(\mathbf{B}_t\boldsymbol{\Theta}^{(0)}\widehat{\boldsymbol{\Sigma}}_t{\boldsymbol{\Theta}^{(0)}}^{\mathsf{T}}\mathbf{B}_t^{\mathsf{T}})\Big\}.$$

C.4 更新 σ_j^2

对于 σ_j^2，我们需要最小化的相应目标函数为

$$n\log\sigma_j^2+\frac{1}{\sigma_j^2}\widehat{S}_j(\mathbf{k}_j^{(0)}).$$

我们对上述目标函数关于 σ_j^2 求一阶导数，并令其为零，以更新 σ_j^2，得到 $\widehat{\sigma}_j^2 = \widehat{S}_j(\mathbf{k}_j^{(0)})/n$，$j = 1, \ldots, J$，其中 $\widehat{S}_j(\mathbf{k}_j^{(0)})$ 是将 12 中的 $\mathbf{k}_j^{(0)}$ (即 jK 的第$\mathbf{K}^{(0)}$ 列) 代入主论文中的式 (

C.5 更新 $\boldsymbol{\Theta}$

对于 $\boldsymbol{\Theta}$，我们首先依次更新 $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1$、……、$\boldsymbol{\theta}_J)$的各列。更新 $\boldsymbol{\theta}_j$ 等价于求解以下目标函数的极小化解:

$$\sum_{t=1}^{n} \left\| \mathbf{z}_t - \mathbf{B}_t \widehat{\boldsymbol{\theta}}_b \widehat{\boldsymbol{\theta}}_c^{\mathsf{T}} \mathbf{c}_t - \sum_{j' \neq j} \mathbf{B}_t \boldsymbol{\theta}_{j'} \widehat{\alpha}_{j',t} - \mathbf{B}_t \boldsymbol{\theta}_j \widehat{\alpha}_{j,t} \right\|^2 + \sum_{t=1}^{n} \operatorname{tr}(\mathbf{B}_t \boldsymbol{\Theta} \widehat{\boldsymbol{\Sigma}}_t \boldsymbol{\Theta}^{\mathsf{T}} \mathbf{B}_t^{\mathsf{T}}) + \widehat{\sigma}^2 \lambda_{pc} \boldsymbol{\theta}_j^{\mathsf{T}} \boldsymbol{\Gamma} \boldsymbol{\theta}_j$$

对上述目标函数关于 $\boldsymbol{\theta}_j$ 求一阶导数，得到

$$-2 \sum_{t=1}^{n} \widehat{\alpha}_{j,t} \mathbf{B}_t^{\mathsf{T}} \left(\mathbf{z}_t - \mathbf{B}_t \widehat{\boldsymbol{\theta}}_b \widehat{\boldsymbol{\theta}}_c^{\mathsf{T}} \mathbf{c}_t - \sum_{j' \neq j} \mathbf{B}_t \boldsymbol{\theta}_{j'} \widehat{\alpha}_{j',t} - \mathbf{B}_t \boldsymbol{\theta}_j \widehat{\alpha}_{j,t} \right)$$
$$+ 2 \sum_{t=1}^{n} \left(\widehat{\boldsymbol{\Sigma}}_{t,jj} \mathbf{B}_t^{\mathsf{T}} \mathbf{B}_t \boldsymbol{\theta}_j + \widehat{\boldsymbol{\Sigma}}_{t,j'j} \mathbf{B}_t^{\mathsf{T}} \mathbf{B}_t \boldsymbol{\theta}_{j'} \right) + 2 \widehat{\sigma}^2 \lambda_{pc} \boldsymbol{\Gamma} \boldsymbol{\theta}_j .$$

将上述表达式设为零，即可更新 $\boldsymbol{\theta}_j$ a s

$$\widehat{\boldsymbol{\theta}}_j = \left\{ \sum_{t=1}^{n} (\widehat{\alpha}_{j,t}^2 + \widehat{\boldsymbol{\Sigma}}_{t,jj}) \mathbf{B}_t^{\mathsf{T}} \mathbf{B}_t + \widehat{\sigma}^2 \lambda_{pc} \boldsymbol{\Gamma} \right\}^{-1} \sum_{t=1}^{n} \mathbf{B}_t^{\mathsf{T}} \left\{ (\mathbf{z}_t - \mathbf{B}_t \widehat{\boldsymbol{\theta}}_b \widehat{\boldsymbol{\theta}}_c^{\mathsf{T}} \mathbf{c}_t) \widehat{\alpha}_{j,t} - \sum_{j' \neq j} (\widehat{\alpha}_{j',t} \widehat{\alpha}_{j,t} + \widehat{\boldsymbol{\Sigma}}_{t,j'j}) \mathbf{B}_t \widehat{\boldsymbol{\theta}}_{j'} \right\}.$$

为保证 $\widehat{\boldsymbol{\Theta}}$的正交归一性，我们对 $\widehat{\boldsymbol{\Theta}} \widehat{\mathbf{H}}_J \widehat{\boldsymbol{\Theta}}^{\mathsf{T}}$进行谱分解，其中 $\widehat{\mathbf{H}}_J = \operatorname{diag}(\widehat{\sigma}_1^2$、……、$\widehat{\sigma}_J^2)$。具体而言，令 $\widehat{\boldsymbol{\Theta}} \widehat{\mathbf{H}}_J \widehat{\boldsymbol{\Theta}}^{\mathsf{T}} = \widetilde{\mathbf{Q}} \widetilde{\mathbf{D}} \widetilde{\mathbf{Q}}^{\mathsf{T}}$，其中 $\widetilde{\mathbf{Q}}$ 是正交归一的 $\widehat{\quad}$ $\widehat{\quad}$，且$\mathbf{D}$ 是对角元素按递减顺序排列的对角矩阵。接着，我们分别用$\widetilde{\mathbf{Q}}$ 和 $\widetilde{\mathbf{D}}$替换 $\boldsymbol{\Theta}$ 和$\mathbf{H}_J$。此外，我们用 $\widetilde{\mathbf{Q}}^{\mathsf{T}} \widehat{\boldsymbol{\Theta}} \widehat{\boldsymbol{\alpha}}_t$替换 $\widehat{\boldsymbol{\alpha}}_t$，这种变换$\widehat{\ }$保持了 $\boldsymbol{\Theta} \widehat{\boldsymbol{\alpha}}_t$的方差。

C.6 更新 K

对于 $\mathbf{K} = (\mathbf{k}_1 \ \ldots, \mathbf{k}_J)$，我们需要最小化的相应目标函数 is

$$\sum_{j=1}^{J} \left\{ -\log |\mathbf{M}_j| + \frac{1}{\widehat{\sigma}_j^2} \widehat{S}_j(\mathbf{k}_j) \right\}, \tag{C.2}$$

$\widehat{\ }$其中 $\mathbf{M}j$ 的形式见（A.3），而 S_j（$\mathbf{k}j$）见主论文中的（12）。注意，$\log |\mathbf{M}_j|$ 的值不随样本量 n的变化而改变，而

（C.2）中的第二项依赖于 n。为简化计算，我们使用第二项$\sum_{j=1}^{J}\{(1/\widehat{\sigma}_j^2)\widehat{S}_j(\mathbf{k}j)\}$来近似（C.2），该项对 K 的每一列分别计算。由于$\widehat{S}_j(\mathbf{k}j) = (1, \mathbf{k}_j^{\mathsf{T}})\widehat{\mathbf{D}}j(1, \mathbf{k}_j^{\mathsf{T}})^{\mathsf{T}}$关于 $\mathbf{k}j$具有二次型，对 $\widehat{S}_j(\mathbf{k}j)$ 关于 $\mathbf{k}j$ 求一阶导数并令其为零，可得 $\widehat{\mathbf{k}}j = (\widehat{\mathbf{D}}pj)^{-1}\widehat{\mathbf{d}}j$, $j = 1, \ldots, J$，其中 $\widehat{\mathbf{d}}j = (\widehat{D}_{1,2,j}, \ldots, \widehat{D}_{1,p+1,j})^{\mathsf{T}}$，且 $\widehat{\mathbf{D}}pj$是 $\widehat{\mathbf{D}}j$右下角的 $p \times p$ 阶主子矩阵。

D 不确定性量化

我们进一步采用半参数自助法，获得 sFPC 模型主成分函数的自助法标准差（SD）曲面。半参数自助法的步骤总结如下。设 $\widehat{\Xi} = \{\widehat{\boldsymbol{\theta}}_b, \widehat{\boldsymbol{\theta}}_c, \widehat{\boldsymbol{\Theta}}, \widehat{\mathbf{K}}\ \widehat{\sigma}^2, \{\widehat{\sigma}_j^2\}_{j=1}^{J}\}$为将所提出的 sFPC 模型拟合到数据集后得到的估计参数。记 sFPC 的拟合残差为 $\{\widehat{\epsilon}_t(x_{t1}, y_{t1}) \ldots, \widehat{\epsilon}_t(x_{tn_t}, y_{tn_t})\}$, $t = 1, \ldots, n$。对于自助法的第 b 个重复样本，我们首先通过自回归模型生成序列相关的FPC得分 $\widehat{\alpha}_{jt}^{(b)},, j = 1, \ldots, J\widehat{}$和 $t = 1, \ldots, n$，其中模型系数采用估计值 $k_{i,j}$，方差采用估计值 $\widehat{\sigma}_j^2$。随后，对于每个时刻 t，我们利用基展开的估计系数生成个体函数，即 $\widehat{f}_t^{(b)} = \mathbf{b}_t(x, y)^{\mathsf{T}}\widehat{\boldsymbol{\theta}}_b\widehat{\boldsymbol{\theta}}_c^{\mathsf{T}}\mathbf{c}_t + \sum_{j=1}^{J}\widehat{\alpha}_{jt}^{(b)}\mathbf{b}j(x, y)\widehat{\boldsymbol{\theta}}_j$。之后，在位置 $\{(x_{t1}, y_{t1}) \ldots, (x_{tn_t}, y_{tn_t})\}$ 处对个体函数求值，并加入从残差 $\{\widehat{\epsilon}_t(x_{t1}, y_{t1}) \ldots, \widehat{\epsilon}_t(x_{tn_t}, y_{tn_t})\}$ 中有放回随机抽取的测量误差，从而生成自助法数据，适用于 $t = 1, \ldots, n$。我们将所提出的 sFPC 模型拟合到第 b 个自助法数据重复样本，并获得估计的主成分函数 $\widehat{\phi}_j(x, y)^{(b)}$，适用于 $j = 1, \ldots, J$。我们将上述步骤重复 B 次，即令 $b = 1, \ldots, B$，并收集估计的第 j 个主成分函数 $\widehat{\phi}_j(x, y)^{(b)}$, $b = 1, \ldots, B$，从而获得每个 j 对应的自助法标准差曲面。

我们真实数据按照上述程序得到的相应结果 $B = 100$ 如图D.1所示。我们观察到，主成分自助法标准差曲面的所有形状均表现为中心区域的数值较低，而边界区域的数值较高。这一现象可能是因为边界区域的站点较少，而中心区域的站点较多。具体到各个成分，第一主成分的自助法标准差值最高可达0.02。与拟合的第一主成分函数的幅度相比，相对较小的标准差值表明

sFPC在这一主导成分上的表现稳定。另一方面，第二和第三主成分的自助法标准差值分别最高可达0.14和0.20。这两个成分较大的标准差值表明，第二和第三主成分函数的估计存在更多不确定性。

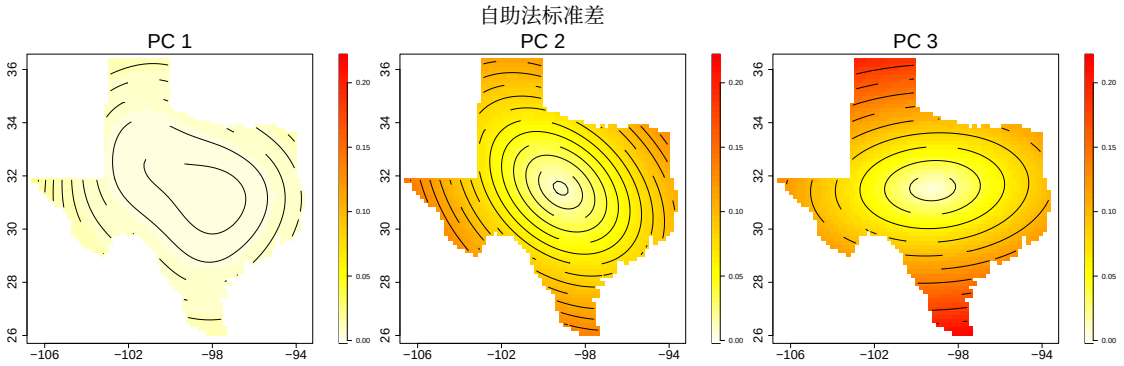


图 D.1：从左到右依次为真实数据分析中 sFPC模型的第一、第二和第三个主成分函数的自助法标准差曲面。

参考文献

Absil, P. A., Mahony, R., and Sepulchre, R. 矩阵流形上的优化算法. 新泽西州普林斯顿：普林斯顿大学出版社（2009年）。Akaike, H. “统计模型识别的新视角。” *IEEE*自动控制汇刊, 19(6):716–723（1974年）。Bosq, D. 函数空间中的线性过程：理论与应用. 纽约：施普林格（2000年）。Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. 时间序列分析：预测与控制. 霍博肯：约翰·威立父子公司（2015年）。Cabrera, B. L. and Schulz, F. “电力需求广义分位数预测：一种函数型数据方法。”《美国统计学会会刊》, 112(517):127–136（2017年）。Chen, L.-H. and Jiang, C.-R. “多维函数型主成分分析。” 统计与计算, 27(5):1181–1192（2017年）。Cipra, T. and Romera, R. “含离群值和缺失观测的卡尔曼滤波器。” *Test*, 6:379–395（1997年）。de Boor, C. 样条实用指南. 纽约：施普林格出版社（1978年）。

Dempster, A. P., Laird, N. M., and Rubin, D. B. “通过EM算法从不完整数据中进行最大似然估计。” 英国皇家统计学会杂志：*B*辑（方法论）, 39(1):1–22（1977年）。Ding, F., He, S., Jones, D. E., and Huang, J. Z. “具有协变量依赖均值和协方差结构的函数型主成分分析。” *Technometrics*, 64(3):335–345（2022年）。Durbin, J. and Koopman, S. J. 基于状态空间方法的时间序列分析. 牛津：牛津大学出版社，第2版（2012年）。Golub, G. H. and Van Loan, C. F. 矩阵计算. 马里兰州巴尔的摩：约翰·霍普金斯大学出版社（2013年）。Hall, P., M¨uller, H.-G., and Wang, J.-L. “函数型和纵向数据分析中主成分方法的性质。” 《统计年刊》, 34(3):1493–1517（2006年）。Hansen, J., Sato, M., Ruedy, R., Lo, K., Lea, D. W., and Medina-Elizade, M. “全球气温变化。” 美国国家科学院院刊, 103(39):14288–14293（2006年）。Huang, J. Z., Shen, H., and Buja, A. “通过惩罚秩一近似进行函数型主成分分析。” 电子统计学杂志, 2:678–695（2008年）。Hyndman, R. J. and Shang, H. L. “函数型时间序列预测。” 韩国统计学会杂志, 38(3):199–211（2009年）。Hyndman, R. J. and Ullah, M. S. “死亡率和生育率的稳健预测：一种函数型数据方法。” 计算统计学与数据分析, 51(10):4942–4956（2007年）。James, G. M., Hastie, T. J., and Sugar, C. A. “稀疏函数型数据的主成分模型。” *Biometrika*, 87(3):587–602（2000年）。Jones, P. D., Wigley, T. M., and Wright, P. B. “1861年至1984年间的全球气温变化。” 自然, 322(6078):430–434（1986年）。Karhunen, K. “关于随机过程的谱理论。” 芬兰科学院年刊。*A*辑 *I*：数学, 34:1–7（1946年）。Kokoszka, P. and Reimherr, M. “确定函数型自回归模型的阶数。” 时间序列分析杂志, 34(1):116–129（2013年）。Lai, M.-J. and Schumaker, L. L. 三角剖分上的样条函数. 剑桥：剑桥大学出版社（2007年）。Li, Y. and Guan, Y. “时空点过程的函数型主成分分析及其在疾病监测中的应用。” 《美国统计学会会刊》, 109(507):1205–1215（2014年）。Liu, C., Ray, S., and Hooker, G. “空间相关数据的函数型主成分分析。” 统计与计算, 27(6):1639–1654（2017年）。

Liu, X., Guillas, S., and Lai, M.-J. “使用SPDE方法与双变量样条进行高效空间建模。”《计算与图形统计学杂志》，25(4):1176– 1194（2016年）。Lo`eve, M. “具有指数正交分解的随机 `a函数。”《科学评论》，84:159–162（1946）。Lorentz, G. G. 伯恩斯坦多项式。纽约：Chelsea出版公司，第2版（1986年）。Menne, M. J., Williams Jr, C. N., and Vose, R. S. “美国历史气候学网络月度温度数据，版本2。”《美国气象学会公报》，90(7):993–1008（2009年）。Mercer, J. “正型与负型函数及其与积分方程理论的联系。”《伦敦皇家学会哲学汇刊 . *A* 辑：数学或物理特征论文集》，209:415–446 （ 1909 ）。Nelder, J. A. and Mead, R. “一种用于函数最小化的单纯形法。”《计算机学报》，7(4):308–313（1965）。Ramsay, J. O. and Silverman, B. W. 函数型数据分析。纽约：Springer科学与商业媒体，第2版（2005年）。Rice, J. A. and Wu, C. O. “不等间隔采样的含噪曲线的非参数混合效应模型。”《生物统计学》，57(1):253–259（2001）。Ruppert, D., Wand, M. P., and Carroll, R. J. 半参数回归。剑桥：剑桥大学出版社（2003）。Schwarz, G. “估计模型的维数。”《统计年刊》，6(2):461–464（1978年）。Shang, H. L. and Hyndman, R. J. “分组函数型时间序列预测：年龄别死亡率的应用。”《计算与图形统计学杂志》，26(2):330–343（2017年）。Shen, H. “光滑曲线时间序列的建模与预测。” *Technometrics*，51(3):227–238（2009年）。Shen, H. and Huang, J. Z. “呼叫中心来电的日间预测与日内更新。”《制造与服务运营管理》，10(3):391–410（2008）。Shi, H., Yang, Y., Wang, L., Ma, D., Beg, M. F., Pei, J., and Cao, J. “用于图像特征提取的二维函数型主成分分析。”《计算与图形统计学杂志》，31(4):1127–1140（2022）。Staniswalis, J. G. and Lee, J. J. “纵向数据的非参数回归分析。”《美国统计学会会刊》，93(444):1403–1418（1998）。Wang, L., Wang, G., Lai, M.-J., and Gao, L. “通过三角剖分上的二元惩罚样条，高效估计复杂区域数据的部分线性模型。”《中华统计学志》，30:347–369（2020a）。

Wang, Y., Wang, G., Wang, L., and Ogden, R. T. “成像数据函数型数据分析中均值函数的同时置信带。”《生物统计学》，76(2):427–437（2020b）。Yao, F., M¨uller, H.-G., and Wang, J.-L. “稀疏纵向数据的函数型数据分析。”《美国统计学会会刊》，100:577–590（2005a）。—. “纵向数据的函数型线性回归分析。”《统计年刊》，33(6):2873–2903（2005b）。Yu, S., Wang, G., Wang, L., Liu, C., and Yang, L. “广义地理加性模型的估计与推断。”《美国统计学会会刊》，115(530):761–774（2020）。Zhou, L., Huang, J. Z., and Carroll, R. “使用主成分联合建模配对稀疏函数型数据。”*Biometrika*，95(3):601–619（2008）。Zhou, L. and Pan, H. “二维函数型数据的主成分分析。”《计算与图形统计学杂志》，23(3):779–801（2014a）。—. “使用三角剖分上的惩罚双变量样条对不规则区域的含噪数据进行平滑。”《计算统计学》，29(1):263–281（2014b）。