arXiv:2302.04134v1 [stat.ME] 8 Feb 2023

用于估计时空平流-扩散过程初始条件的逆模型

刘晓[1] 和 Kyongmin Yeo[2]

[1]阿肯色大学工业工程系[2]IBM T. J. Watson 研究中心

摘要

反问题是利用观测数据推断物理过程未知参数的问题。本文研究一类重要的反问题——利用空间稀疏数据流估计时空平流-扩散过程的初始条件。本文考虑三种空间采样方案，包括不规则采样、非均匀采样和平移均匀采样。不规则采样方案是一般情形，而对于非均匀采样和平移均匀采样，则可在谱域中高效求解。针对每种采样方案，本文将反问题表述为一个正则化凸优化问题，通过最小化正向模型输出与观测之间的距离来求解。该优化问题采用交替方向乘子法求解；该算法也能处理对模型输出施加线性不等式约束（例如非负性）的情况。本文给出数值示例，在 GitHub 上提供代码，并通过讨论总结所提出反演建模方法的一些有益见解。

关键词： 逆模型、时空过程、对流扩散过程、交替方向乘子法

1 引言

1.1 例证

反问题是利用观测数据推断物理过程未知参数的方法，在科学与工程应用中十分常见。例如，在城市空气质量和环境监测领域，反问题旨在迅速确定导致公共卫生问题的气态污染物瞬时排放源（Eckhardt 等人，2008；Martinez-Camara 等人，2014；Hwang 等人，2019），或检测工业运营意外释放造成的逸散排放（Hosseini 和 Stockie，2016；Klein 等人，2016）。在医疗应用中，逆模型已被用于根据体表测量获取心脏表面电位，这一问题称为心电图反问题（Yao 和 Yang，2021年）。在地震学中，反问题旨在根据地球表面的地震波，获取有关地震震源处作用力结构的信息（Apostol，2019）。反演建模也被用于利用海洋漂浮物（Miron 等人，2019）或声重力波（Kadri，2019）检测失踪的马来西亚航空MH370航班的坠落地点。

本文研究一类重要的统计反问题——利用空间稀疏数据流估计时空对流扩散过程的初始条件。考虑检测工业运营中逸散排放的意外泄漏（Hosseini 和 Stockie，2016）。图1展示了一个 2×2 千米^2 的区域，其中包括位于加拿大不列颠哥伦比亚省特雷尔的一家大型铅锌冶炼厂。四个大型红色圆圈表示硫酸锌（$ZnSO_4$）的潜在排放源，小型蓝色圆圈表示部署用于检测意外 $ZnSO_4$ 泄漏的九个受体（即传感器）的位置。$ZnSO_4$ 的输运受形式为偏微分方程（PDE）的平流-扩散方程支配。发生意外 $ZnSO_4$ 排放时，传感器监测数据用于估计可能的排放位置。这一逆问

题需要一个统计模型，该模型能够（i）通过整合底层的平流-扩散物理，建立观测、排放源和过程参数（例如风、扩散率和衰减）之间明确且易于解释的联系；（ii）纳入传感数据流，以估计释放 $ZnSO_4$ 时的初始条件；以及（iii）处理来自不同传感器网络布局的数据，例如不规则、均匀、非均匀、嵌套等布局。

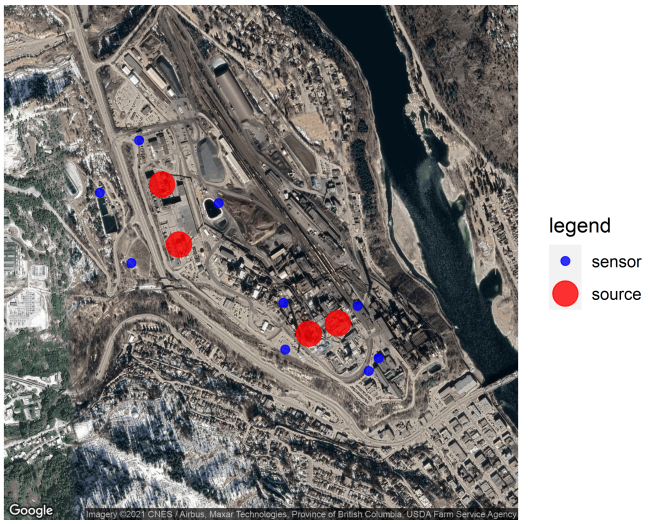

图1：一个包含四个潜在 $ZnSO_4$ 排放源（大红圆圈）和九个地面测量受体（小蓝圆圈）的空间区域

1.2 统计逆模型与文献综述

逆模型通常涉及构建一个优化问题，使正向模型输出与观测之间的距离最小化（Constantinescu 等人，2019）。考虑一个由方程 $F(\xi, \Theta) = 0$ 控制的物理过程，其中 ξ 和 Θ 分别为状态和参数（未知量）。由于过程的状态 ξ 必须依照控制方程取决于参数 Θ，我们可以定义从 Θ 到 ξ 的映射，即 $\mathcal{F}(\Theta) = \xi$，称为参数到可观测量映射。一旦获得过程的观测 $\boldsymbol{Y}$，就可以将逆问题概念化地表述为 $\hat{\Theta} = \text{argmin}_{\Theta} \mathcal{L}(\mathcal{F}(\Theta)\ \boldsymbol{Y})$，其中 $\mathcal{L}$ 是某个预先定义的损失函数。例如，Hwang 等人（2019）提出了一种贝叶斯逆模型，通过利用伴随平流-扩散算子来估计二维源函数。作者采用有限

差分法求解正向和反向物理模型，并根据观测构建了排放率的似然函数。Oates 等人（2019）提出了一种逆模型，用于估计工业水力旋流器设备电势模型中的时变参数。他们采用贝叶斯方法，将数值解中物理方程误差的统计模型纳入考量。Yeo 等人（2019）提出了一种用于对流扩散过程源检测的谱方法。作者使用高斯径向基函数来近似空间上的平滑排放函数，并用广义多项式混沌对谱系数进行建模。

注意，物理模型 F 通常通过将偏微分方程转换为一个大型常微分方程（ODE）系统来求解，其中物理域采用有限差分离散化。当离散化的维度较高时，直接求解控制方程 F 以获得正向模型输出通常计算成本高昂。因此，采用统计代理建模来构建参数到可观测量映射$\mathcal{F}$ （Mak 等人，2018年；Qian 等人，2019年；Gul 等人，2018年）。例如，高斯过程（GP）已被广泛用于构建统计代理模型（Hung 等人，2015年；Deng 等人，2017年；Gramacy，2020年；Zhang 等人，2021年；Sauer 等人，2021年）。特别是在对流扩散过程中，Sigrist 等人（2015年）通过求解具有空间和时间上不变的对流扩散算子的偏微分方程，获得了一个高斯过程；Liu 等人（2022年）则通过考虑空间变化的平流-扩散扩展了这一方法。近年来，物理信息机器学习正在迅速兴起，用于数据驱动发现具有物理意义的控制物理规律，以及进行状态/参数/算子推断。例如，Raissi 等人（2019年）提出了一个深度学习框架，用于求解非线性偏微分方程的正问题和逆问题。Chen 等人（2021年）提出了一种主动学习方法，用于估计未知的微分方程。结合 D最优性和最大最小空间填充准则的自适应设计准则被用于减少实验数据量，其中 D最优性考虑未知的

微分方程及其导数的解。

1.3 问题陈述、贡献与概述

本文研究一种统计反问题模型，旨在根据空间上稀疏的传感器测量，估计对流扩散过程在整个空间域上的初始条件。该问题可正式表述如下：

问题陈述。 设 $\xi(t, \boldsymbol{s})$ 为一个对流扩散过程，在 M 个空间位置进行监测，持续 L 个离散时间段。本文关注一个逆问题，即利用空间上稀疏的传感器数据流，估计整个空间域上的 $\xi(0, \boldsymbol{s})$ 。

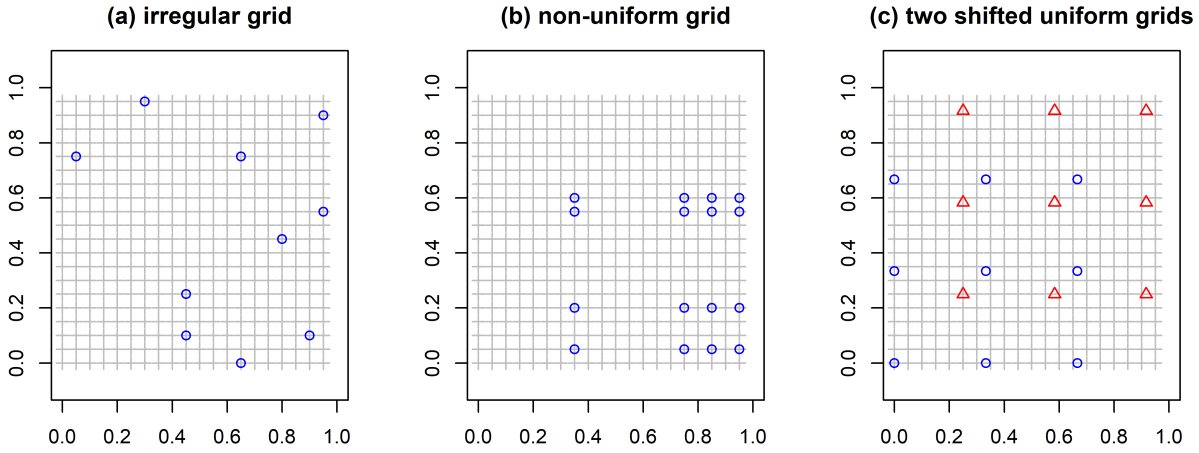

图2：本文考虑的三种空间采样方案示意图：(a) 不规则网格，(b) 非均匀网格，(c) 平移均匀网格（嵌套）。

具体而言，本文考察了三种重要的空间采样方案（即网络布局）：不规则采样、非均匀采样和平移均匀采样。请注意：（i）不规则采样（图2a）是一般情形，包含非均匀采样、平移均匀采样和均匀采样作为其特殊情形；（ii）本文还研究了两个特殊情形，即非均匀采样和平移均匀采样（图2b 和图2c），因为针对这两种特殊方案，谱域中存在计算效率较高的解法。在实践中，非均匀采样常用于缩短采集时间、降低传感器安装成本和功耗，尤其适用于监测低活动信号

（Venkataramani 和 Bresler，2001年；Beyrouthy 等，2015年）。平移均匀采样（又称嵌套阵列或差分共阵）包含两个相互嵌套的均匀传感网络，可显著增加线性阵列的自由度。通过嵌套两个或更多均匀线性阵列，平移均匀采样仅使用 $\mathcal{O}(M^2)$ 个自由度，即可使用 M 个物理传感器，从而缓解谱分析中的谱混叠问题（Pal 和 Vaidyanathan，2010年；Qin 和 Amin，2021年）。

本文的贡献总结如下：(i) 本文首次基于 Liu 等人（2022年）提出的平流-扩散过程正向时空模型，构建了逆模型。该正向模型为我们的逆模型提供参数到观测量映射 $\mathcal{F}$，通过空间基的线性组合和谱系数多元随机过程分解物理时空过程。谱系数的时间动态由平流-扩散方程决定，从而将支配物理规律融入统计模型中；详见第 2.1 节。本文遵循谱分解的思想，通过估计决定 $\xi(0,\boldsymbol{s})$ 的零时刻谱系数，即可在整个空间域内估计 $\xi(0,\boldsymbol{s})$。(ii) 由于在给定传感器网络布局的情况下，估计零时刻谱系数需要时空范围内的充分观测，第 2.2 节进行了理论研究，并获得了唯一估计零时刻谱系数的充分必要条件。当这些条件不满足时，谱系数无法被唯一确定，这在信号处理中称为谱混叠。(iii) 在第 2.3 节中，我们进一步指出，并非总能唯一估计所有零时刻谱系数。传感器网络布局以及时空观测数量通常受到实际条件的限制。因此，我们往往需要在充分条件和必要条件均不满足的情形下估计谱系数。此外还应注意，如果 $\xi(0,\boldsymbol{s})$ 具有平滑性，高频模态下的谱系数可能会快速衰减。为解决这一问题，第 2.3 节提出

了一个正则化逆问题，用于在任意不规则传感器网络布局下估计零时刻的谱系数。该正则化同时引入了谱系数的稀疏性以及相邻谱系数的平滑性。第 3 节介绍了在非均匀采样和平移均匀采样下获得数据时的特殊结果，在这些情况下可以在谱域中高效计算。(iv) 最后，第 4 节开发了交替方向乘子法（ADMM）算法，以高效求解所提出的正则化逆问题。随后，我们扩展了所提出的逆模型和 ADMM 算法，以处理 $\xi(0,\boldsymbol{s})$ 上的非负约束，即 $\xi(0,\boldsymbol{s}) \geq 0$。第 5 节进行了全面的数值研究。通过敏感性分析展示了所提方法的稳健性。

2 一般不规则采样下的逆向建模

2.1 预备知识

考虑由偏微分方程描述的物理时空对流扩散过程 $\xi(t,\ \boldsymbol{s})$:

$$\mathcal{A}\xi(t,\boldsymbol{s}) = \phi(t,\boldsymbol{s}), \qquad \boldsymbol{s} \in \mathbb{S}, t \geq 0 \tag{1}$$

其中，$\mathbb{S}$ 为空间域，$\phi(t,\boldsymbol{s})$为源项，而对流扩散算子$\mathcal{A}$ 由 $\mathcal{A}\xi(t,\boldsymbol{s}) = \dot{\xi}(t,\boldsymbol{s}) + \vec{\boldsymbol{v}}^T\nabla\xi(t,\boldsymbol{s}) - \nabla\cdot[\boldsymbol{D}\nabla\xi(t,\boldsymbol{s})] + \zeta\xi(t,\boldsymbol{s})$ 给出，其中 $\vec{\boldsymbol{v}}$、$\boldsymbol{D}$、ζ、∇和 $\nabla\cdot$ 分别表示速度场、扩散张量、衰减、梯度和散度算子。偏微分方程 (1) 是一大类物理现象背后的控制方程，在这些现象中，粒子和能量会在系统内部传递。

在本文中，过程 $\xi(t,\boldsymbol{s})$ 只能通过空间分布式传感器在离散时刻进行观测。因此，逆问题旨在估计整个空间域上的初始条件$\xi(0,\boldsymbol{s})$。根据 Liu 等人 (2022年)，$\xi(0,\boldsymbol{s})$ 假设由有限个正交空间傅里叶基函数张成，

$$\xi(0,\boldsymbol{s}) = \sum_{\boldsymbol{k}\in\mathcal{K}} \eta(\boldsymbol{k}) f_{\boldsymbol{k}}(\boldsymbol{s}) \tag{2}$$

其中，$\boldsymbol{k}=(k_1,k_2)^T\in\mathcal{K}$ 表示波数，$f_{\boldsymbol{k}}(\boldsymbol{s})=e^{\imath 2\pi \boldsymbol{s}^T\boldsymbol{k}}$ 表示傅里叶基函数，$\eta(\boldsymbol{k})$ 是决定各傅里叶模态权重的系数，并且

$$\mathcal{K}=\left\{(k_1,k_2)^T;k_1=-\frac{N_1}{2}+1,-\frac{N_1}{2}+2,\cdots,\frac{N_1}{2},k_2=-\frac{N_2}{2}+1,-\frac{N_2}{2}+2,\cdots,\frac{N_2}{2}\right\}. \quad (3)$$

请注意，当初始条件为带限，且其傅里叶展开中的高频部分快速衰减至恰好为零时，式（2）中的等式成立。根据式（2），已证明过程 $\xi(t,\boldsymbol{s})$ 在 $\mathbb{S}$ 中保持 $t\geq 0$，并且具有如下谱表示：$\xi(t,\boldsymbol{s})=\sum_{\boldsymbol{k}\in\mathcal{K}}\alpha(t,\boldsymbol{k})f_{\boldsymbol{k}}(\boldsymbol{s})$，其中 $\alpha(t,\boldsymbol{k})$ 是随时间演化的傅里叶系数，而 $\alpha(0,\boldsymbol{k})=\eta(\boldsymbol{k})$（Sigrist 等人，2015年；Liu 等人（2022年））。

接下来，考虑一个传感器网络，其中有 M 个传感器，分别位于空间位置 $\boldsymbol{s}_1,\boldsymbol{s}_2,...,\boldsymbol{s}_M$。令列向量 $\boldsymbol{Y}(l)=(Y(l,\boldsymbol{s}_1)\ Y(l,\boldsymbol{s}_2)...,Y(l,\boldsymbol{s}_M))^T$ 包含时间 $l(l=1,2,...,L)$由对流扩散过程（1）产生的观测，并令 $M\times L$ 矩阵 $\boldsymbol{Y}$为来自 L 个时间段的观测集合：$\boldsymbol{Y}=[\boldsymbol{Y}(1)\ \boldsymbol{Y}(2)\cdots,\boldsymbol{Y}(L)]$。随后，Liu 等人（2022年）提出了一种基于偏微分方程（1）的时空模型：

$$\boldsymbol{Y}=\boldsymbol{F}\boldsymbol{E}\boldsymbol{G}+\boldsymbol{V}. \quad (4)$$

这里，

- $\boldsymbol{F}$是一个由傅里叶基函数组成的 $M\times N$矩阵（$N=N_1\times N_2$），$\boldsymbol{F}=(\boldsymbol{f}_{\boldsymbol{k}_1},\ \boldsymbol{f}_{\boldsymbol{k}_2},\ \cdots,\ \boldsymbol{f}_{\boldsymbol{k}_N})$，其中 $\boldsymbol{f}_{\boldsymbol{k}_i}=(e^{\imath 2\pi \boldsymbol{s}_1^T\boldsymbol{k}_i},\ e^{\imath 2\pi \boldsymbol{s}_2^T\boldsymbol{k}_i},\ \cdots,\ e^{\imath 2\pi \boldsymbol{s}_M^T\boldsymbol{k}_i})^T$，而 $\imath$ 是虚数单位。

- $\boldsymbol{E}=\text{diag}(\boldsymbol{\alpha}(0,\ \boldsymbol{k}))=\text{diag}(\boldsymbol{\eta})$ 是一个 $N\times N$ 矩阵，表示时间 0 时的谱系数；$\boldsymbol{\eta}$ 是一个包含所有 $\boldsymbol{k}\in\mathcal{K}$ 对应的 $\eta(\boldsymbol{k})$ 的向量。

- $\boldsymbol{G}$ 是一个 $N\times L$ 矩阵，$\boldsymbol{G}=(\boldsymbol{g}(1)\ \boldsymbol{g}(2)\cdots,\ \boldsymbol{g}(L))$，它刻画了 $\boldsymbol{E}$ 中各元素的时间演化。这里，$\boldsymbol{g}(l)=(g_1(l).....,\ g_N(l))^T$是一个列向量，其中 $g_j(l)=e^{\gamma_j(l-1)\Delta}$，$\Delta$是时间采样间隔，而 $\gamma_j=-\boldsymbol{k}_j^T\boldsymbol{D}\boldsymbol{k}_j-\zeta-\imath\vec{\boldsymbol{v}}^T\boldsymbol{k}_j$适用于 $j=1,\ 2,\ \cdots,\ N$。

- $\boldsymbol{V}=(\boldsymbol{v}(1)\ \boldsymbol{v}(2)...,\ \boldsymbol{v}(L))$ 是一个刻画测量误差的 $M\times L$ 矩阵，而$\boldsymbol{v}(l)$ 服从多元高斯分布 $N_M(0,\ \boldsymbol{\Sigma}_v)$，其中 $l=1,\ 2,\ ...,\ L$。

读者可参考Sigrist 等人 (2015年）和Liu 等人 (2022年）以了解（4）的详细信息。该

模型（4）基于利用谱理论和特征函数展开求解非线性动力系统的经典解，并作为本文建立逆模型的基础。

2.2 逆问题及其基本性质

在本文所考虑的反演建模问题中，$\boldsymbol{Y}$ 为传感器观测，$\boldsymbol{F}$ 和 $\boldsymbol{G}$ 均为预先计算得到，目标是估计系数向量 $\boldsymbol{\eta}$，该向量决定初始条件 $\xi(0, \boldsymbol{s})$。请注意，在空间和时间上的观测不足且传感器网络布局特殊时，谱系数 $\boldsymbol{\eta}$ 可能无法被唯一确定（这称为谱混叠）。命题1首先为唯一估计 $\boldsymbol{\eta}$ 中的所有分量确立了必要条件。

命题1。若传感器数量为 M，且对该过程（1）进行 L 个离散时间段的观测，则当至少满足以下两个条件之一时，可以根据模型（$\boldsymbol{\eta}$ 4）唯一估计出其中的所有谱系数：

条件A：给定速度 $\vec{\boldsymbol{v}}$ 和扩散率 $\boldsymbol{D}$，不存在$\boldsymbol{k}_{j_1}$ 和$\boldsymbol{k}_{j_2}$ （$\boldsymbol{k}_{j_1}$，$\boldsymbol{k}_{j_2} \in \mathcal{K}$ 和 $j_1 \neq j_2$）使得

$$\vec{\boldsymbol{v}}^T(\boldsymbol{k}_{j_1} - \boldsymbol{k}_{j_2}) = 0 \quad and \quad \boldsymbol{k}_{j_1}^T \boldsymbol{D} \boldsymbol{k}_{j_1} = \boldsymbol{k}_{j_2}^T \boldsymbol{D} \boldsymbol{k}_{j_2}. \tag{5}$$

条件B：存在至少两个采样位置 $\boldsymbol{s}$ 和 $\boldsymbol{s}'$，使得以下条件均不成立：

$$2\boldsymbol{k}_{j_1}(\boldsymbol{s} - \boldsymbol{s}') \in \mathbb{Z}^{(odd)}, \quad 2\boldsymbol{k}_{j_2}(\boldsymbol{s} - \boldsymbol{s}') \in \mathbb{Z}^{(odd)} \tag{6a}$$

$$2\boldsymbol{k}_{j_1}(\boldsymbol{s} - \boldsymbol{s}') \in \mathbb{Z}^{(even)}, \quad 2\boldsymbol{k}_{j_2}(\boldsymbol{s} - \boldsymbol{s}') \in \mathbb{Z}^{(even)}. \tag{6b}$$

所有证明均见补充材料。正如该命题所示，能否唯一估计 $\boldsymbol{\eta}$ 中的所有谱系数，取决于底层过程的关键物理参数，例如速度 $\vec{\boldsymbol{v}}$ 和扩散率 $\boldsymbol{D}$，这正如人们自然会预期的那样。

接下来，我们研究使 $\boldsymbol{\eta}$ 中所有分量都能被唯一

估计的充分条件。一般而言，充分条件要求空间观测（即较大的 M）足够多，或时间观测（即较大的 L）足够多，或者两者兼具。这在直觉上是成立的，下面将对此进行详细讨论。

（**当空间观测数量较大时**）。若 $M \geq N$，则 $\boldsymbol{F}$ 的左逆存在。令 $\boldsymbol{F}_L^{-1}$ 为 $\boldsymbol{F}$的左逆，则（4）可改写为 $[\boldsymbol{F}_L^{-1}\boldsymbol{Y}]^T = \boldsymbol{G}^T\boldsymbol{E}^T + [\boldsymbol{F}_L^{-1}\boldsymbol{V}]^T$。因此，令 $\boldsymbol{G}_{:,\ j}^T$ 为 $\boldsymbol{G}^T$ 的第 j 列，$j = 1$，2，……，N，我们有

$$\mathrm{vec}([\boldsymbol{F}_L^{-1}\boldsymbol{Y}]^T) = \mathrm{diag}(\{\boldsymbol{G}_{:,j}^T\}_{j=1}^N)\boldsymbol{\eta} + \mathrm{vec}([\boldsymbol{F}_L^{-1}\tilde{\boldsymbol{V}}]^T) \tag{7}$$

其中 vec(·) 表示矩阵向量化，而 diag($\{\boldsymbol{G}_{:,j}^T\}_{j=1}^N$) 是一个分块对角矩阵，其列向量 $\boldsymbol{G}_{:,\ j}^T$ 是第 j 个对角块。$g_{i,\ j}$ 在 $\boldsymbol{G}$ 中的指数结构保证了 $LN \times N$ 矩阵 diag($\{\boldsymbol{G}_{:,j}^T\}_{j=1}^N$) 满列秩，因此 $\boldsymbol{\eta}$ 中的所有元素均可唯一确定。请注意，当 $M \geq N$ 时，空间采样频率超过奈奎斯特频率——即在不发生混叠的情况下可采样的最大带宽。

（**当时间观测数量较大时**）。的较大取值对应于另一种情况，即时间样本充足。通过考察（4）中 γ_j 的表达式，可以找到 j 和 j' （$j \neq j'$），使得 $\gamma_j = \gamma_{j'}$。换言之，$\boldsymbol{G}$ 可能行秩亏，且各行相同。

设 $\{1, 2, ..., N\} \overset{\mathcal{M}}{\to} \Psi = \{\Psi_1, \Psi_2, ..., \Psi_{\tilde{N}}\}$ 为一个映射，其中 Ψ_i($i = 1$、2、……、$\tilde{N}$) 是一个集合，满足 $\gamma_j = \gamma_{j'}$ 对于 j，$j' \in \Psi_i$。换言之，映射 $\mathcal{M}$ 将 $\tilde{\boldsymbol{G}}$ 划分为若干部分，每个部分都包含相同的行，而 $\boldsymbol{G}$ 的行秩为 N。

命题2。如果时间样本数 L 大于 $\boldsymbol{G}$ 的（行）秩，即，$L > N$，那么使 $\boldsymbol{\eta}$ 中所有分量都能被唯一确定的充分条件是

$$\mathrm{rank}(\tilde{\boldsymbol{F}}_i) = |\Psi_i|, \qquad \forall i = 1, ..., \tilde{N} \tag{8}$$

其中 $\boldsymbol{F}_i = \{f_{m,j}\}_{m=1,...,M, j\in\Psi_i}$ 是一个 $M \times |\Psi_i|$ 矩阵，$|\Psi_i|$ 表示集合 Ψ_i 对于 $i = 1$，…，N的基数。

2.3 正则化逆问题

上述命题1和命题2为从空间分布式传感器数据流中唯一确定 $\boldsymbol{\eta}$ 中的所有分量建立了充分且必要的条件。尽管这两个命题具有理论价值，但实际应用可能并不总是要求唯一估计 $\boldsymbol{\eta}$ 中的所有分量。例如，空间/时间样本的数量以及传感器可部署的位置始终受到实际约束。因此，我们通常要求在既不满足充分条件也不满足必要条件的场景下估计 $\boldsymbol{\eta}$ 。

我们首先重写（4），并定义 $\vec{\mathscr{Y}}$、$\mathscr{X}^{(\text{P-I})}$，以及 $(\mathbf{0},\ \boldsymbol{\Sigma})$，如下：

$$\underbrace{\begin{pmatrix}\boldsymbol{Y}(1)\\ \vdots \\ \boldsymbol{Y}(L)\end{pmatrix}}_{\vec{\mathscr{Y}}} = \underbrace{\begin{pmatrix}\boldsymbol{F}\mathrm{diag}(\{\boldsymbol{g}_j(1)\}_{j=1}^N)\\ \vdots \\ \boldsymbol{F}\mathrm{diag}(\{\boldsymbol{g}_j(L)\}_{j=1}^N)\end{pmatrix}}_{\mathscr{X}^{(\text{P-I})}} \boldsymbol{\eta} + \underbrace{\begin{pmatrix}\boldsymbol{v}(1)\\ \vdots \\ \boldsymbol{v}(L)\end{pmatrix}}_{\vec{\mathscr{V}}} \tag{9}$$

并考虑一个正则化逆问题：

$$\text{Problem P-I:} \qquad \min \ \frac{1}{2}(\vec{\mathscr{Y}} - \mathscr{X}^{(\text{P-I})}\boldsymbol{\eta})^T\boldsymbol{\Sigma}^{-1}(\vec{\mathscr{Y}} - \mathscr{X}^{(\text{P-I})}\boldsymbol{\eta}) + \mathcal{R}(\boldsymbol{\eta}). \tag{10}$$

这里，选择正则化 $\mathcal{R}(\boldsymbol{\eta})$ 是基于以下两点考虑：**(i)** 对于平滑或带限初始条件（见 (2)），高频模态会迅速衰减，即能量集中在低频区域。这促使我们施加 L_1-正则化，以 $\boldsymbol{\eta}$ （即稀疏性） $\boldsymbol{\eta}$。**(ii)** 对于平滑的初始条件，预计 $\boldsymbol{\eta}$ 中的分量会逐渐衰减（不一定单调）。这促使我们对相邻的相邻分量之间 $\boldsymbol{\eta}$的平滑性施加一定要求，以防止估计的特殊系数出现突变。具体而言，对 L_2中相连分量之间的差值施加 $\boldsymbol{\eta}$ -正则化，并同时沿水平方向和垂直方向进行：

$$\|\boldsymbol{J}_1\boldsymbol{\eta}\|_2^2 = \sum_{i,j\in\mathcal{K}^{(h)}} (\eta(\boldsymbol{k}_i), \eta(\boldsymbol{k}_j))^2, \quad \|\boldsymbol{J}_2\boldsymbol{\eta}\|_2^2 = \sum_{i,j\in\mathcal{K}^{(v)}} (\eta(\boldsymbol{k}_i), \eta(\boldsymbol{k}_j))^2 \tag{11}$$

其中， $\boldsymbol{J}_1$ 和 $\boldsymbol{J}_2$ 分别是水平方向和垂直方向上的矩阵差分算子，集合 $\mathcal{K}^{(h)}$ 和 $\mathcal{K}^{(v)}$ 由所有在水平方向和垂直方向上相连的频率 $\boldsymbol{k} \in \mathcal{K}$ 组成，即 $\mathcal{K}^{(h)} = \{\boldsymbol{k}_i, \boldsymbol{k}_j; k_{j,1} - k_{i,1} = 1, i < j\}$ 和 $\mathcal{K}^{(v)} = \{\boldsymbol{k}_i, \boldsymbol{k}_j; k_{j,2} - k_{i,2} = 1, i < j\}$。注意，式 (11) 修改了 Fussed Lasso（Tibshirani 等，2005）的思想。两者的区别在于，Fussed Lasso 对系数之间的差值施加 L_1-正则化，从而得到稀疏分段常数解；而我们适合采用 L_2-正则化，使 $\boldsymbol{\eta}$ 中的分量能够在高频区域和低频区域之间平滑变化。

最后，式（ $\mathcal{R}(\boldsymbol{\eta})$（10）中的正则化项为

$$\mathcal{R}(\boldsymbol{\eta}) = \lambda_1 \|\boldsymbol{\eta}\|_1 + \lambda_2 \|\boldsymbol{J}\boldsymbol{\eta}\|_2^2 \tag{12}$$

其中， $\boldsymbol{J} = (\boldsymbol{J}_1^T, \boldsymbol{J}_2^T)^T$是一个二维差分算子， λ_1和 λ_2分别控制 $\boldsymbol{\eta}$ 中的稀疏性，以及 $\boldsymbol{\eta}$中相邻分量之间的平滑性。逆问题（10）可通过交替方向乘子法（ADMM）（ZouandHastie，2005；Ramdas andTibshirani，2016）求解。我们的问题所用的 ADMM 算法详见第4节。

3 两个特殊情形

本节将进一步研究两种特殊的采样方案，即第2.3节讨论的非均匀采样和平移均匀采样，并表明在这两种特殊方案下，可以在谱域中获得计算高效的解。

3.1 非均匀采样

考虑一个矩形网格系统，它由两个一维配置点集合的张量积构成， $\mathcal{M} = \mathcal{M}_1 \otimes \mathcal{M}_2$，其中 $\mathcal{M}_1 = \{m_1; m_1 = 0, 1, ..., M_1 - 1\}$ 和 $\mathcal{M}_2 = \{m_2; m_2 = 0, 1, ..., M_2 - 1\}$ 是配置点集合。这里， $\mathcal{M}$ 是一个网格系统

由传感器可能部署的候选位置构成。令 $\mathcal{M}_1 \subseteq$ ~$\mathcal{M}_1$ 和 $\mathcal{M}_2 \subseteq \mathcal{M}_2$，非均匀采样网格由网格系统 $\mathcal{M} = \mathcal{M}_1 \otimes \mathcal{M}_2$给出，其中 $|\mathcal{M}_1| = M_1$、$|\mathcal{M}_1| = M_2$，且 $|\cdot|$ 表示集合的基数。令 $y(l, \boldsymbol{m})$ 表示时间 l 、位置 $\boldsymbol{m} \in \mathcal{M}$处的观测。那么，$y(l, \boldsymbol{m})$ 的II型非均匀离散傅里叶变换（NUDFT-II）为:

$$\beta(l, \boldsymbol{q}) = \frac{1}{|\mathcal{M}|} \sum_{\boldsymbol{m} \in \mathcal{M}} y(l, \boldsymbol{m}) e^{-\imath 2\pi \boldsymbol{m}^T \boldsymbol{q}}, \qquad \text{for } \boldsymbol{q} \in \mathcal{Q}, \tag{13}$$

其中 $\mathcal{Q} = \{(q_1, q_2)^T; q_1 = -\frac{M_1}{2} + 1, -\frac{M_1}{2} + 2, \cdots, \frac{M_1}{2}, q_2 = -\frac{M_2}{2} + 1, -\frac{M_2}{2} + 2, \cdots, \frac{M_2}{2}\}$.

将 $y(l, \boldsymbol{m})$ 在 (13) 中替换为其在区域 $\mathcal{K}$上的离散傅里叶变换，可得

$$\begin{aligned} \beta(l, \boldsymbol{q}) &= \frac{1}{|\mathcal{M}|} \sum_{\boldsymbol{m}} \left\{ \sum_{\boldsymbol{k} \in \mathcal{K}} \left[\alpha(l, \boldsymbol{k}) + \varepsilon(l, \boldsymbol{k}) \right] e^{\imath 2\pi \boldsymbol{m}^T \boldsymbol{k}} \right\} e^{-\imath 2\pi \boldsymbol{m}^T \boldsymbol{q}} \\ &= \sum_{i \in \mathcal{I}_{\boldsymbol{q}}} \sum_{i \in \mathcal{J}_{\boldsymbol{q}}} \left[\alpha(l, \boldsymbol{q} + (iM_1, jM_2)^T) + \varepsilon(l, \boldsymbol{q} + (iM_1, jM_2)^T) \right] \end{aligned} \tag{14}$$

其中 ε 源于观测误差，集合 $\mathcal{I}_{\boldsymbol{q}}$ 和 $\mathcal{J}_{\boldsymbol{q}}$ 分别定义为

$$\begin{aligned} \mathcal{I}_{\boldsymbol{q}} &= \left\{ i; -\frac{N_1}{2} + 1 \le (q_1 + iM_1) \le \frac{N_1}{2}, i \in \mathbb{Z} \right\} \\ \mathcal{J}_{\boldsymbol{q}} &= \left\{ j; -\frac{N_2}{2} + 1 \le (q_2 + jM_2) \le \frac{N_2}{2}, j \in \mathbb{Z} \right\}. \end{aligned} \tag{15}$$

这里，(14) 的第一行是将$y(l, \boldsymbol{m})$的傅里叶变换直接代入 (18) 得到的。(14) 的第二行则利用了傅里叶基众所周知的正交性质，即仅当 $\boldsymbol{k} = \boldsymbol{q} + (iM_1, jM_2)^T$时，$\sum_{\boldsymbol{k} \in \mathcal{K}} [\alpha(l, \boldsymbol{k}) + \varepsilon(l, \boldsymbol{k})] e^{\imath 2\pi \boldsymbol{m}^T (\boldsymbol{k} - \boldsymbol{q})} = 1$；其中 $i \in \mathcal{I}_{\boldsymbol{q}}$且 $j \in \mathcal{J}_{\boldsymbol{q}}$；否则，$\sum_{\boldsymbol{k} \in \mathcal{K}} [\alpha(l, \boldsymbol{k}) + \varepsilon(l, \boldsymbol{k})] e^{\imath 2\pi \boldsymbol{m}^T (\boldsymbol{k} - \boldsymbol{q})} = 0$。因此，给定网格系统 $\mathcal{M}$ （即采集数据的空间位置），(14) 表明，$\beta(l, \boldsymbol{q})$可表示为多个傅里叶系数 $\alpha(l, \boldsymbol{q} + (iM_1, jM_2)^T)$之和，其中 $i \in \mathcal{I}_{\boldsymbol{q}}$ 且 $j \in \mathcal{J}_{\boldsymbol{q}}$。换言之，无法根据 $\beta(l, \boldsymbol{q})$唯一估计所有 $i \in \mathcal{I}_{\boldsymbol{q}}$和 $j \in \mathcal{J}_{\boldsymbol{q}}$对应的 $\alpha(l, \boldsymbol{q} + (iM_1, jM_2)^T)$。这称为谱混叠。

为了清晰地揭示频谱混叠结构，我们引入一个集合 $\mathcal{K}_{\boldsymbol{q}}$

$$\mathcal{K}_{\boldsymbol{q}} = \left\{(k_1, k_2)^T; k_1 = q_1 + iM_1, k_2 = q_2 + jM_2, i \in \mathcal{I}_{\boldsymbol{q}}, j \in \mathcal{J}_{\boldsymbol{q}}\right\} \tag{16}$$

该集合包含 $\mathcal{K}$ 中与 $\boldsymbol{q} \in \mathcal{Q}$ 对应的所有波数。显然，与 $\mathcal{K}_{\boldsymbol{q}}$ 中波数对应的傅里叶系数都相互混淆，无法唯一确定，除非时间观测的数量 L 足够多（见命题2）。考虑一个简单的示例，其中 $N_1 = N_2 = 4$ 和 $M_1 = M_2 = 2$，并定义四个集合： $\mathcal{K}_{(0,0)^T} = \{(0,0)^T、(0,2)^T、(2,0)^T、(2,2)^T\}$， $\mathcal{K}_{(0,1)^T} = \{(0,1)^T、(0,-1)^T、(2,1)^T、(2,-1)^T\}$， $\mathcal{K}_{(1,0)^T} = \{(1,0)^T、(-1,0)^T、(1,2)^T、(-1,2)^T\}$，以及 $\mathcal{K}_{(1,1)^T} = \{(1,1)^T、(-1,-1)^T、(-1,1)^T、(1,-1)^T\}$，使得每个集合都包含 $\mathcal{K}$ 中的波数，当时间样本数量不足时，这些波数对应的傅里叶系数会相互混淆。注意，对于 $\boldsymbol{q} \neq \boldsymbol{q}'$， $\bigcup \mathcal{K}_{\boldsymbol{q}} = \mathcal{K}$ 和 $\mathcal{K}_{\boldsymbol{q}} \bigcap \mathcal{K}_{\boldsymbol{q}'} = 0$，也就是说， $\mathcal{K}_{\boldsymbol{q}}$，互斥且完备。

将 $\boldsymbol{\alpha}$ 在（4）中的时间动态代入（14），可得

$$\beta(l, \boldsymbol{q}) = \mathbf{1}^T \mathrm{diag}(\boldsymbol{\eta}_{\boldsymbol{q}}) \boldsymbol{g}_{\boldsymbol{q}}(l) + \mathbf{1}^T \boldsymbol{\varepsilon}_{\boldsymbol{q}}(l), \qquad l = 1, 2, ..., L, \tag{17}$$

其中$\mathbf{1}$ 是全 1 列向量， $\boldsymbol{g}_{\boldsymbol{q}}(l)$、 $\boldsymbol{\varepsilon}_{\boldsymbol{q}}(l)$和 $\boldsymbol{\eta}_{\boldsymbol{q}}$ 分别是从 $\boldsymbol{g}(l)$、 $\boldsymbol{\varepsilon}(l)$和 $\boldsymbol{\eta}$ 中仅保留与 $\boldsymbol{k} \in \mathcal{K}_{\boldsymbol{q}}$对应的分量后得到的列向量。

将所有 L 个采样时刻的 β（l、 $\boldsymbol{q}$）合并起来，由（17）可得

$$\underbrace{\begin{pmatrix} \beta(1, \boldsymbol{q}) \\ \vdots \\ \beta(L, \boldsymbol{q}) \end{pmatrix}}_{\boldsymbol{\beta}_{\boldsymbol{q}}} = \underbrace{\begin{pmatrix} \mathbf{1}^T & & \\ & \ddots & \\ & & \mathbf{1}^T \end{pmatrix}}_{\boldsymbol{B}_{\boldsymbol{q}}} \underbrace{\begin{pmatrix} \mathrm{diag}(\boldsymbol{g}_{\boldsymbol{q}}(1)) \\ \vdots \\ \mathrm{diag}(\boldsymbol{g}_{\boldsymbol{q}}(L)) \end{pmatrix}}_{\boldsymbol{G}_{\boldsymbol{q}}} \boldsymbol{\eta}_{\boldsymbol{q}} + \underbrace{\begin{pmatrix} \mathbf{1}^T & & \\ & \ddots & \\ & & \mathbf{1}^T \end{pmatrix}}_{\boldsymbol{B}_{\boldsymbol{q}}} \underbrace{\begin{pmatrix} \boldsymbol{\varepsilon}_{\boldsymbol{q}}(1) \\ \vdots \\ \boldsymbol{\varepsilon}_{\boldsymbol{q}}(L) \end{pmatrix}}_{\boldsymbol{W}_{\boldsymbol{q}}} \tag{18}$$

其中 $\boldsymbol{\beta}_{\boldsymbol{q}}$ 是一个 $L \times 1$ 列向量， $\boldsymbol{B}_{\boldsymbol{q}} = \mathrm{diag}\{\mathbf{1}^T\}$ 是一个 $L \times$ （$L \times d_{\boldsymbol{q}}$）块对角矩阵，其对角块为 $d_{\boldsymbol{q}} = |\mathcal{K}_q|$，而 $\boldsymbol{G}_{\boldsymbol{q}}$ 是一个（$L \times d_{\boldsymbol{q}}$） $\times d_{\boldsymbol{q}}$ 矩阵。因此，对任意 $\boldsymbol{q} \in \mathcal{Q}$，由（17）可得线性模型

$$\boldsymbol{\beta}_{\boldsymbol{q}} = \boldsymbol{B}_{\boldsymbol{q}} \boldsymbol{G}_{\boldsymbol{q}} \boldsymbol{\eta}_{\boldsymbol{q}} + \boldsymbol{B}_{\boldsymbol{q}} \boldsymbol{W}_{\boldsymbol{q}}, \qquad \boldsymbol{B}_{\boldsymbol{q}} \boldsymbol{W}_{\boldsymbol{q}} \sim \mathcal{N}(\mathbf{0}, \sigma^2 d_{\boldsymbol{q}} \boldsymbol{I}). \tag{19}$$

请注意，若干因素决定了是否可以从线性模型（ $\boldsymbol{\eta}_{\boldsymbol{q}}$ 中唯一确定分量（19），其中包括传感器网络布局 $\mathcal{M}$、时间样本的数量

样本 L，以及底层物理过程的参数。以下命题给出了唯一估计 $\boldsymbol{\eta}_{\boldsymbol{q}}$ 中分量的条件。

命题 3. 对于 $\boldsymbol{k}_{j_1}$，$\boldsymbol{k}_{j_2} \in \mathcal{K}_q$ 和 $j_1 \neq j_2$，若 ***(i)*** $L \geq d_{\boldsymbol{q}}$，且 ***(ii)*** 以下条件至少有一个不成立：$\vec{\boldsymbol{v}}^T(\boldsymbol{k}_{j_1} - \boldsymbol{k}_{j_2}) = 0$ 和 $\boldsymbol{k}_{j_1}^T \boldsymbol{D} \boldsymbol{k}_{j_1} - \boldsymbol{k}_{j_2}^T \boldsymbol{D} \boldsymbol{k}_{j_2} = 0$，则 $\boldsymbol{B}_{\boldsymbol{q}} \boldsymbol{G}_{\boldsymbol{q}}$ 为满列秩，且 $\boldsymbol{\eta}_{\boldsymbol{q}}$ 中的所有谱系数均可由（*19*）唯一估计。

尽管上述命题表明，可以通过针对所有 $\boldsymbol{\eta}_{\boldsymbol{q}}$ 的线性模型组来估计 $\boldsymbol{q} \in \mathcal{Q}$，但直接求解这些独立的线性模型通常并不合适，原因如下：当 $\boldsymbol{B}_{\boldsymbol{q}} \boldsymbol{G}_{\boldsymbol{q}}$ 和 $\vec{\boldsymbol{v}}^T(\boldsymbol{k}_{j_1} - \boldsymbol{k}_{j_2})$ 都接近于零时，矩阵 $\boldsymbol{k}_{j_1}^T \boldsymbol{D} \boldsymbol{k}_{j_1} - \boldsymbol{k}_{j_2}^T \boldsymbol{D} \boldsymbol{k}_{j_2}$ 很容易变得病态或在计算上奇异。换言之，$\boldsymbol{B}_{\boldsymbol{q}} \boldsymbol{G}_{\boldsymbol{q}}$ 中的某些列可能几乎相同。因此，我们将所有 19 的线性模型（$\boldsymbol{q}_1$、$\boldsymbol{q}_2$、……、$\boldsymbol{q}_{|\mathcal{Q}|}$）合并，得到

$$\underbrace{\begin{pmatrix} \boldsymbol{\beta}_{\boldsymbol{q}_1} \\ \vdots \\ \boldsymbol{\beta}_{\boldsymbol{q}_{|\mathcal{Q}|}} \end{pmatrix}}_{\vec{\mathscr{Y}}^{(\mathrm{P\text{-}II})}} = \underbrace{\begin{pmatrix} \boldsymbol{B}_{\boldsymbol{q}_1} \boldsymbol{G}_{\boldsymbol{q}_1} & & \\ & \ddots & \\ & & \boldsymbol{B}_{\boldsymbol{q}_{|\mathcal{Q}|}} \boldsymbol{G}_{\boldsymbol{q}_{|\mathcal{Q}|}} \end{pmatrix}}_{\mathscr{X}^{(\mathrm{P\text{-}II})}} \underbrace{\begin{pmatrix} \boldsymbol{\eta}_{\boldsymbol{q}_1} \\ \vdots \\ \boldsymbol{\eta}_{\boldsymbol{q}_{|\mathcal{Q}|}} \end{pmatrix}}_{\boldsymbol{\eta}} + \underbrace{\begin{pmatrix} \boldsymbol{B}_{\boldsymbol{q}_1} \boldsymbol{W}_{\boldsymbol{q}_1} \\ \vdots \\ \boldsymbol{B}_{\boldsymbol{q}_{|\mathcal{Q}|}} \boldsymbol{W}_{\boldsymbol{q}_{|\mathcal{Q}|}} \end{pmatrix}}_{\mathscr{V}^{(\mathrm{P\text{-}II})}} \tag{20}$$

其中，$\vec{\mathscr{Y}}^{(\mathrm{P\text{-}II})}$ 是一个（$|\mathcal{Q}| \times L$）$\times 1$ 列向量，$\mathscr{X}^{(\mathrm{P\text{-}II})}$ 是一个（$|\mathcal{Q}| \times L$）$\times |\mathcal{K}|$ 矩阵，$\boldsymbol{\eta}$ 是一个$|\mathcal{K}| \times 1$ 列向量，$\mathscr{V}^{(\mathrm{P\text{-}II})} \sim \mathcal{N}$（$\mathbf{0}$，$\boldsymbol{\Sigma}^{(\mathrm{P\text{-}II})}$）以及 $\boldsymbol{\Sigma}^{(\mathrm{P\text{-}II})} = \sigma^2 \mathrm{diag}(\{d_{\boldsymbol{q}_i} \boldsymbol{I}\}_{i=1}^{|\mathcal{Q}|})$。

与问题 P-I（10）类似，我们再次得到如下正则化逆问题：

问题 P-II：$\min \frac{1}{2} (\vec{\mathscr{Y}}^{(\mathrm{P\text{-}II})} - \mathscr{X}^{(\mathrm{P\text{-}II})} \boldsymbol{\eta})^T (\boldsymbol{\Sigma}^{(\mathrm{P\text{-}II})})^{-1} (\vec{\mathscr{Y}}^{(\mathrm{P\text{-}II})} - \mathscr{X}^{(\mathrm{P\text{-}II})} \boldsymbol{\eta}) + \mathcal{R}(\boldsymbol{\eta})$。
其中，$\mathcal{R}(\boldsymbol{\eta})$ 在（12）中定义。 (21)

3.2 平移均匀采样

移位均匀传感器阵列或平台由两个嵌套的矩形网格系统（图 2c）组成，$\mathcal{M}^{(1)}$ 以及 $\mathcal{M}^{(2)}$，它们分别由两个一维配置点集的张量积定义：

$$\mathcal{M}^{(1)} = (\frac{m_1}{M_1}; m_1 = 0, 1, ..., M_1 - 1) \otimes (\frac{m_2}{M_2}; m_2 = 0, 1, ..., M_2 - 1) \tag{22}$$

$$\mathcal{M}^{(2)} = (\frac{m_1}{M_1} + \delta_1; m_1 = 0, 1, ..., M_1 - 1) \otimes (\frac{m_2}{M_2} + \delta_2; m_2 = 0, 1, ..., M_2 - 1) \tag{23}$$

其中，$0 < \delta_1 < M_1^{-1}$和 $0 < \delta_2 < M_2^{-1}$，且 $\boldsymbol{\delta} = (\delta_1, \delta_2)^T$表示两个传感器平台之间的空间位移。令 $y^{(1)}(l, \boldsymbol{m})$表示来自$\mathcal{M}^{(1)}$的、在时间 l 和位置 $\boldsymbol{m}$ 处的观测，其中 $\boldsymbol{m} = (\frac{m_1}{M_1}, \frac{m_2}{M_2})^T$。对于任意 $\boldsymbol{q} = (q_1, q_2)^T \in \mathcal{Q}$，基于第一个网格系统 $\mathcal{M}^{(1)}$ 在时间 l 的观测，在 $\beta^{(1)}(l, \boldsymbol{q})$处的傅里叶系数为：

$$\begin{aligned}\beta^{(1)}(l, \boldsymbol{q}) &= \frac{1}{M_1 M_2} \sum_{\boldsymbol{m}} y^{(1)}(l, \boldsymbol{m}) e^{-\imath 2\pi \boldsymbol{m}^T \boldsymbol{q}} = \frac{1}{M_1 M_2} \sum_{\boldsymbol{m}} \left\{ \sum_{\boldsymbol{k} \in \mathcal{K}} \left[\alpha(l, \boldsymbol{k}) + \varepsilon(l, \boldsymbol{k}) \right] e^{\imath 2\pi \boldsymbol{m}^T \boldsymbol{k}} \right\} e^{-\imath 2\pi \boldsymbol{m}^T \boldsymbol{q}} \\ &= \sum_{i \in \mathcal{I}_{\boldsymbol{q}}} \sum_{i \in \mathcal{J}_{\boldsymbol{q}}} \left\{ \alpha(l, \boldsymbol{q} + (iM_1, jM_2)^T) + \varepsilon(l, \boldsymbol{q} + (iM_1, jM_2)^T) \right\}\end{aligned} \tag{24}$$

其中集合 $\mathcal{I}_{\boldsymbol{q}}$ 和 $\mathcal{J}_{\boldsymbol{q}}$ 定义见（15）。

由于第二个网格系统 $\mathcal{M}^{(2)}$ 是在空间中将第一个网格系统 $\mathcal{M}^{(2)}$ 位移 $\boldsymbol{\delta}$ 后得到的，因此可以直接由从$\mathcal{M}^{(2)}$采集的数据得到傅里叶系数 $\beta^{(2)}(l, \boldsymbol{q})$：

$$\beta^{(2)}(l, \boldsymbol{q}) = \sum_{i \in \mathcal{I}_{\boldsymbol{q}}} \sum_{i \in J_{\boldsymbol{q}}} \left\{ \alpha(l, \boldsymbol{q} + (iM_1, jM_2)^T) + \varepsilon(l, \boldsymbol{q} + (iM_1, jM_2)^T) \right\} e^{\imath 2\pi \boldsymbol{\delta}^T (\boldsymbol{q} + (iM_1, jM_2)^T)} \tag{25}$$

其中，最后一项 $e^{\imath 2\pi \boldsymbol{\delta}^T (\boldsymbol{q} + (iM_1,\ jM_2)^T)}$ 是由从 $\mathcal{M}^{(1)}$到 $\mathcal{M}^{(2)}$的空间位移引起的。

因此，对于 $\boldsymbol{q} \in \mathcal{Q}$，可以定义一个集合

$$\mathcal{K}_{\boldsymbol{q}} = \{(k_1, k_2); k_1 = q_1 + iM_1, k_2 = q_2 + jM_2, i \in \mathcal{I}_{\boldsymbol{q}}, j \in \mathcal{J}_{\boldsymbol{q}}\}, \tag{26}$$

对于 $\boldsymbol{k} \in \mathcal{K}_{\boldsymbol{q}}$，我们有

$$\begin{pmatrix} \beta^{(1)}(l, \boldsymbol{q}) \\ \beta^{(2)}(l, \boldsymbol{q}) \end{pmatrix} = \begin{pmatrix} \mathbf{1}^T \\ \boldsymbol{b}_{\boldsymbol{\delta}, \boldsymbol{q}}^T \end{pmatrix} \mathrm{diag}(\boldsymbol{\eta}_{\boldsymbol{q}}) \boldsymbol{g}_{\boldsymbol{q}}(l) + \begin{pmatrix} \mathbf{1}^T \\ \boldsymbol{b}_{\boldsymbol{\delta}, \boldsymbol{q}}^T \end{pmatrix} \boldsymbol{\varepsilon}_{\boldsymbol{q}}(l) \tag{27}$$

对于 $l = 1$、2、.....、L。这里，$\mathbf{1}$ 和 $\boldsymbol{b}_{\boldsymbol{\delta}、\ \boldsymbol{q}} = \{e^{\imath 2\pi \boldsymbol{\delta}^T (\boldsymbol{q} + (iM_1,\ jM_2)^T)}\}_{i \in \mathcal{I}_{\boldsymbol{q}}、\ j \in \mathcal{J}_{\boldsymbol{q}}}$是长度为 $d_{\boldsymbol{q}} = |\mathcal{I}_{\boldsymbol{q}} \otimes \mathcal{J}_{\boldsymbol{q}}|$ 的列向量，而 $\boldsymbol{g}_{\boldsymbol{q}}(l)$、$\boldsymbol{\varepsilon}_{\boldsymbol{q}}(l)$ 和 $\boldsymbol{\eta}_{\boldsymbol{q}}$ 分别是通过.....得到的列向量

从 $\boldsymbol{g}(l)$、 $\boldsymbol{\varepsilon}(l)$ 和 $\boldsymbol{\eta}$ 仅保留与 $\boldsymbol{k} \in \mathcal{K}_{\boldsymbol{q}}$对应的分量。与第 3.1 节 $\mathcal{K}_{\boldsymbol{q}}$ 的讨论类似，由 $\mathcal{K}$ 中与 $\boldsymbol{q}$ 相对应的所有波数组成；这些波数会发生混叠，除非时间观测的数量L 足够大，否则无法唯一确定。结合所有采样时刻的 $\beta^{(1)}(l, \boldsymbol{q})$ 和 $\beta^{(2)}(l, \boldsymbol{q})$，可得

$$\underbrace{\begin{pmatrix} \beta^{(1)}(1,\boldsymbol{q}) \\ \beta^{(2)}(1,\boldsymbol{q}) \\ \vdots \\ \beta^{(1)}(L,\boldsymbol{q}) \\ \beta^{(2)}(L,\boldsymbol{q}) \end{pmatrix}}_{\boldsymbol{\beta}_{\boldsymbol{q}}^{(\text{P-III})}} = \underbrace{\begin{pmatrix} \mathbf{1}^T & & \\ \boldsymbol{b}_{\boldsymbol{\delta},\boldsymbol{q}}^T & & \\ & \ddots & \\ & & \mathbf{1}^T \\ & & \boldsymbol{b}_{\boldsymbol{\delta},\boldsymbol{q}}^T \end{pmatrix}}_{\boldsymbol{B}_{\boldsymbol{\delta},\boldsymbol{q}}} \underbrace{\begin{pmatrix} \operatorname{diag}(\boldsymbol{g}_{\boldsymbol{q}}(1)) \\ \vdots \\ \operatorname{diag}(\boldsymbol{g}_{\boldsymbol{q}}(L)) \end{pmatrix}}_{\boldsymbol{G}_{\boldsymbol{q}}} \boldsymbol{\eta}_{\boldsymbol{q}} + \underbrace{\begin{pmatrix} \mathbf{1}^T & & \\ \boldsymbol{b}_{\boldsymbol{\delta},\boldsymbol{q}}^T & & \\ & \ddots & \\ & & \mathbf{1}^T \\ & & \boldsymbol{b}_{\boldsymbol{\delta},\boldsymbol{q}}^T \end{pmatrix}}_{\boldsymbol{B}_{\boldsymbol{\delta},\boldsymbol{q}}} \underbrace{\begin{pmatrix} \boldsymbol{\varepsilon}_{\boldsymbol{q}}(1) \\ \vdots \\ \boldsymbol{\varepsilon}_{\boldsymbol{q}}(L) \end{pmatrix}}_{\boldsymbol{W}_{\boldsymbol{q}}} \tag{28}$$

其中 $\boldsymbol{\beta}_{\boldsymbol{q}}^{(\text{P-III})}$是一个 $2L \times 1$ 列向量，$\boldsymbol{B}_{\boldsymbol{\delta},\ \boldsymbol{q}} = \operatorname{diag}\{(\mathbf{1},\ \boldsymbol{b}_{\boldsymbol{\delta},\ \boldsymbol{q}})^T\}$ 是一个 $2L \times (L \times d_{\boldsymbol{q}})$ 块对角矩阵，其中 $d_{\boldsymbol{q}} = |\mathcal{K}_q|$，而 $\boldsymbol{G}_{\boldsymbol{q}}$ 是一个 $(L \times d_{\boldsymbol{q}})$ 矩阵。对于任意 $\boldsymbol{q} \in \mathcal{Q}$，我们有

$$\boldsymbol{\beta}_{\boldsymbol{q}}^{(\text{P-III})} = \boldsymbol{B}_{\boldsymbol{\delta},\boldsymbol{q}} \boldsymbol{G}_{\boldsymbol{q}} \boldsymbol{\eta}_{\boldsymbol{q}} + \boldsymbol{B}_{\boldsymbol{\delta},\boldsymbol{q}} \boldsymbol{W}_{\boldsymbol{q}} \tag{29}$$

其中 $\boldsymbol{B}_{\boldsymbol{\delta},\ \boldsymbol{q}} \boldsymbol{G}_{\boldsymbol{q}}$ 是一个 $2L \times d_{\boldsymbol{q}}$ 矩阵。由于 $0 < \delta_1 < M_1^{-1}$和 $0 < \delta_2 < M_2^{-1}$，$\boldsymbol{b}_{\boldsymbol{\delta},\ \boldsymbol{q}}$ 中的元素相同，因此 $2L \geq d_{\boldsymbol{q}}$ 是 $\boldsymbol{B}_{\boldsymbol{\delta},\ \boldsymbol{q}} \boldsymbol{G}_{\boldsymbol{q}}$满列秩的充分条件，即

$$L \geq \frac{d_{\boldsymbol{q}}}{2} \geq \frac{1}{2} \left\lfloor \frac{N_1 - 1}{M_1} + 1 \right\rfloor \left\lfloor \frac{N_2 - 1}{M_2} + 1 \right\rfloor. \tag{30}$$

如第 3.1 节所述，直接求解（29）中的各个线性模型并不适用于病态问题。因此，将所有 $\boldsymbol{q}_1$、 $\boldsymbol{q}_2$、 ……、$\boldsymbol{q}_{|\mathcal{Q}|}$ 的线性模型（29）结合起来，可得

$$\underbrace{\begin{pmatrix} \boldsymbol{\beta}_{\boldsymbol{q}_1} \\ \vdots \\ \boldsymbol{\beta}_{\boldsymbol{q}_{|\mathcal{Q}|}} \end{pmatrix}}_{\vec{\mathcal{Y}}^{(\text{P-III})}} = \underbrace{\begin{pmatrix} \boldsymbol{B}_{\boldsymbol{\delta},\boldsymbol{q}_1} \boldsymbol{G}_{\boldsymbol{q}_1} & & \\ & \ddots & \\ & & \boldsymbol{B}_{\boldsymbol{\delta},\boldsymbol{q}_{|\mathcal{Q}|}} \boldsymbol{G}_{\boldsymbol{q}_{|\mathcal{Q}|}} \end{pmatrix}}_{\mathcal{X}^{(\text{P-III})}} \underbrace{\begin{pmatrix} \boldsymbol{\eta}_{\boldsymbol{q}_1} \\ \vdots \\ \boldsymbol{\eta}_{\boldsymbol{q}_{|\mathcal{Q}|}} \end{pmatrix}}_{\boldsymbol{\eta}} + \underbrace{\begin{pmatrix} \boldsymbol{B}_{\boldsymbol{\delta},\boldsymbol{q}_1} \boldsymbol{W}_{\boldsymbol{q}_1} \\ \vdots \\ \boldsymbol{B}_{\boldsymbol{\delta},\boldsymbol{q}_{|\mathcal{Q}|}} \boldsymbol{W}_{\boldsymbol{q}_{|\mathcal{Q}|}} \end{pmatrix}}_{\mathcal{V}^{(\text{P-III})}} \tag{31}$$

其中，$\vec{\mathscr{Y}}^{\text{(P-III)}}$ 是一个（$2|\mathcal{Q}| \times L$）$\times 1$ 列向量，$\mathscr{X}^{\text{(P-III)}}$ 是一个（$2|\mathcal{Q}| \times L$）$\times |\mathcal{K}|$ 矩阵，$\boldsymbol{\eta}$ 是一个 $|\mathcal{K}| \times 1$ 列向量，$\mathscr{V}^{\text{(P-III)}} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}^{\text{(P-III)}})$、$\boldsymbol{\Sigma}^{\text{(P-III)}} = \sigma^2 \text{diag}(\{\boldsymbol{S}_i\}_{i=1}^{|\mathcal{Q}|})$ 以及 $\boldsymbol{S}_i = \text{diag}(d_{\boldsymbol{q}_i}\boldsymbol{I}、||\boldsymbol{b}_{\boldsymbol{\delta}、\boldsymbol{q}_i}||_2^2\boldsymbol{I})$，其中 $i = 1, 2、\ldots\ldots、|\mathcal{Q}|$。

与问题 P-II（21）类似，我们得到一个正则化逆问题：问题 P-III：$\min \frac{1}{2}(\vec{\mathscr{Y}}^{\text{(P-III)}} - \mathscr{X}^{\text{(P-III)}}\boldsymbol{\eta})^T(\boldsymbol{\Sigma}^{\text{(P-III)}})^{-1}(\vec{\mathscr{Y}}^{\text{(P-III)}} - \mathscr{X}^{\text{(P-II)}}\boldsymbol{\eta}) + \mathcal{R}(\boldsymbol{\eta})$，其中 $\mathcal{R}(\boldsymbol{\eta})$ 定义见（12）。

$$- \tag{32}$$

备注。问题 P-I 通过最小化时空域中正向模型输出与观测之间的平方距离来估计特殊系数 $\boldsymbol{\eta}$，而问题 P-II 和 P-III 则通过最小化谱域中正向模型输出的谱系数与观测的谱系数之间的平方距离来估计谱系数 $\boldsymbol{\eta}$。因此，当我们通过逆傅里叶变换，将由 P-II 和 P-III 得到的最优解从谱域转换回时空域时，该解在最小二乘意义上不再是最优的（即，在时空域中，正向模型输出与观测之间的平方距离并未最小化）。这是因为最小二乘估计量在变换下并非不变。如果计算成本不是首要考虑因素，我们建议使用问题 P-I（10）在时空域中求解逆问题。

4 使用交替方向乘子法求解问题 P-I、P-II 和 P-III

本节介绍求解逆问题 P-I、P-II 和 P-III 所需的算法（本节中，为避免歧义，省略上标 $.^{\text{P-I}}$、$.^{\text{P-II}}$ 和 $.^{\text{P-III}}$）。请注意，在这三个逆问题中，$\boldsymbol{\eta}$ 的维数为 $|\mathcal{K}| = N_1 \times N_2$。因此，即使 N_1 和 N_2 的规模适中，$\boldsymbol{\eta}$ 的维数也可能很大。对于大规模优化问题，交替方向乘子法（ADMM）是一个合理的选择。

我们首先将一般形式的无约束问题转化为

$$\min_{\boldsymbol{\eta}} \quad \frac{1}{2}(\vec{\mathscr{y}} - \mathscr{X}\boldsymbol{\eta})^T\boldsymbol{\Sigma}^{-1}(\vec{\mathscr{y}} - \mathscr{X}\boldsymbol{\eta}) + \mathcal{R}(\boldsymbol{\eta}) \tag{33}$$

转化为约束问题:

$$\min_{\boldsymbol{\eta},\boldsymbol{\psi}} \quad f(\boldsymbol{\eta}) + \mathcal{R}(\boldsymbol{\psi}), \qquad \text{s.t. } \boldsymbol{\eta} = \boldsymbol{\psi} \tag{34}$$

其中 $f(\boldsymbol{\eta}) = \frac{1}{2}(\vec{\mathscr{y}} - \mathscr{X}\boldsymbol{\eta})^T\boldsymbol{\Sigma}^{-1}(\vec{\mathscr{y}} - \mathscr{X}\boldsymbol{\eta})$。对于 $\rho > 0$，增广拉格朗日函数的缩放形式写为:

$$f(\boldsymbol{\eta}) + \mathcal{R}(\boldsymbol{\psi}) + \frac{\rho}{2}\|\boldsymbol{\eta} - \boldsymbol{\psi} + \boldsymbol{u}\|_2^2 + \frac{\rho}{2}\|\boldsymbol{u}\|_2^2 . \tag{35}$$

然后，ADMM 通过重复以下迭代来求解约束问题（34）（Zou 和 Hastie，2005；Ramdas 和 Tibshirani，2016）：

$$\boldsymbol{\eta}^{(i)} = \text{argmin}_{\boldsymbol{\eta}} f(\boldsymbol{\eta}) + \frac{\rho}{2}\left\|\boldsymbol{\eta} - \boldsymbol{\psi}^{(i-1)} + \boldsymbol{u}^{(i-1)}\right\|_2^2 \tag{36a}$$

$$\boldsymbol{\psi}^{(i)} = \text{argmin}_{\boldsymbol{\psi}} \mathcal{R}(\boldsymbol{\psi}) + \frac{\rho}{2}\left\|\boldsymbol{\eta}^{(i)} - \boldsymbol{\psi} + \boldsymbol{u}^{(i-1)}\right\|_2^2 \tag{36b}$$

$$\boldsymbol{u}^{(i)} = \boldsymbol{u}^{(i-1)} + \boldsymbol{\eta}^{(i)} - \boldsymbol{\psi}^{(i)} \tag{36c}$$

对于 $i = 1, 2, \cdots$。这些迭代满足：残差收敛（即 $\boldsymbol{\eta}^{(i)} - \boldsymbol{\psi}^{(i)} \to 0$ 当 $i \to \infty$时），目标函数收敛（即 $f(\boldsymbol{\eta}^{(i)}) + \mathcal{R}(\boldsymbol{\psi}^{(i)}) \to f^* + \mathcal{R}^*$ 其中 f^* 和 $\mathcal{R}^*$ 是原始最优值），以及对偶收敛（即 $\boldsymbol{u}^{(i)} \to \boldsymbol{u}^*$ 其中 $\boldsymbol{u}^*$ 是对偶解）。算法1概述了为求解（33）而设计的 ADMM 算法。补充材料中提供了算法1各步骤推导过程的技术细节。

备注。尽管问题 P-I、P-II 和 P-III 均可通过 ADMM 算法求解，但需要注意的是，问题 P-I 是在时空域中表述的，而问题 P-II 和 P-III 则是在谱域中构建的。因此，$\mathscr{X}^{(\text{P-I})}$式（9）中的设计矩阵是稠密矩阵，而 $\mathscr{X}^{(\text{P-II})}$ 和 $\mathscr{X}^{(\text{P-III})}$ 式（20）和式（31）中的设计矩阵是稀疏的（块对角），因此在 ADMM 算法中计算 $\mathscr{X}^T\boldsymbol{\Sigma}^{-1}\mathscr{X}$、$\mathscr{X}^T\boldsymbol{\Sigma}^{-1}$ 和 $\boldsymbol{\Sigma}^{-1}\mathscr{X}$的速度更快。此外，问题 P-II 和 P-III 使得可以截断

高频分量，因为在式（ $\mathscr{X}$（20）和（31）中，的每个块都对应一个频率层级。这进一步有助于缩短计算时间，详细内容见补充材料。

在许多应用中，可以对逆模型的输出添加非负约束（例如，检测到的排放物或初始条件需要为非负）。添加非负约束后，逆建模问题（33）变为：

$$\min_{\boldsymbol{\eta}} \frac{1}{2}(\vec{\mathscr{Y}} - \mathscr{X}\boldsymbol{\eta})^T \boldsymbol{\Sigma}^{-1}(\vec{\mathscr{Y}} - \mathscr{X}\boldsymbol{\eta}) + \mathcal{R}(\boldsymbol{\eta}),\ \text{s.t.}\ \mathscr{X}\boldsymbol{\eta} \geq 0 \tag{37}$$

在补充材料中，我们表明，通过修改第4节所述的交替方向乘子法算法，可以高效求解约束问题（37），从而使所提出的模型适用于更广泛的问题。

算法1： 求解反问题的交替方向乘子法

数据： $\mathscr{Y}$, $\mathscr{X}$, $\mathscr{V}$, λ_1, λ_2

$\boldsymbol{\eta}^{(0)}$, $\boldsymbol{\psi}^{(0)}$, $\boldsymbol{u}^{(0)}$, $\rho > 0$, $\omega > 0$ //初始化$i \leftarrow 1$

（外循环）

当未满足收敛准则**时**

$$\boldsymbol{\eta}^{(i)} \leftarrow \left(\mathscr{X}^T\boldsymbol{\Sigma}^{-1}\mathscr{X} + \tfrac{1}{2}\rho\boldsymbol{I}\right)^{-1}\left\{\mathscr{X}^T\boldsymbol{\Sigma}^{-1}\vec{\mathscr{Y}} + \rho(\boldsymbol{\psi}^{(\boldsymbol{i-1})} + \boldsymbol{u}^{(i-1)})\right\}$$

（内循环）

$\tilde{\boldsymbol{\psi}}^{(0)}$ //内循环初始化

$j \leftarrow 1$

当未满足收敛准则时**执行**

$$\tilde{\boldsymbol{\psi}}^{(j)} \leftarrow \tfrac{\rho}{2\lambda_2}\left(\boldsymbol{J}^T\boldsymbol{J} + \tfrac{1}{2}(\rho+\omega)\boldsymbol{I}\right)^{-1}(\boldsymbol{\eta}^{(i)} + \boldsymbol{u}^{(i-1)} + \omega\boldsymbol{\theta}^{(j-1)} - \omega\boldsymbol{v}^{(j-1)})$$

$$\boldsymbol{\theta}^{(j)} \leftarrow S_{\lambda_1/\omega}(\tilde{\boldsymbol{\psi}}^{(j)} + \boldsymbol{v}^{(j-1)}),\quad \boldsymbol{v}^{(j)} \leftarrow \boldsymbol{v}^{(j-1)} + \tilde{\boldsymbol{\psi}}^{(j)} - \boldsymbol{\theta}^{(j)}$$

$j \leftarrow j + 1$

$$\boldsymbol{\psi}^{(i)} \leftarrow \tilde{\boldsymbol{\psi}}^{(j)},\quad \boldsymbol{u}^{(i)} \leftarrow \boldsymbol{u}^{(i-1)} + \boldsymbol{\eta}^{(i)} - \boldsymbol{\psi}^{(i)}$$

$i \leftarrow i + 1$

---

5 数值示例

本节通过两个数值示例，展示所提出的逆模型的应用，并为该方法提供一些有益的见解。

5.1 示例I

我们首先在一个 40×40矩形网格上，模拟由偏微分方程 (1) 描述的对流扩散过程。对流扩散算子 $\mathcal{A}$ 的参数 $\vec{\boldsymbol{v}} = (0.5 \times10^{-2}, 0.5 \times 10^{-2})^T$, $\boldsymbol{D} = \text{diag}\{0.25 \times 10^{-3}\}$和 $\zeta = 0$取为：。初始条件包含三个空间稀疏瞬时源，分别为 $\Phi(t, \boldsymbol{s}) = \delta(t - 0)\sum_{j=1}^{3}\phi_j(\boldsymbol{s})$。这里，$\phi_j(\boldsymbol{s}) = \phi_0 \exp\{-\frac{||\boldsymbol{s}-\boldsymbol{s}_0^{(j)}||_2}{0.09}\}$，其中 $\phi_0 = 300$, $\boldsymbol{s}_0^{(1)} = (0.4, 0.2)^T$, $\boldsymbol{s}_0^{(2)} = (0.2, 0.4)^T$和 $\boldsymbol{s}_0^{(3)} = (0.5, 0.5)^T$。

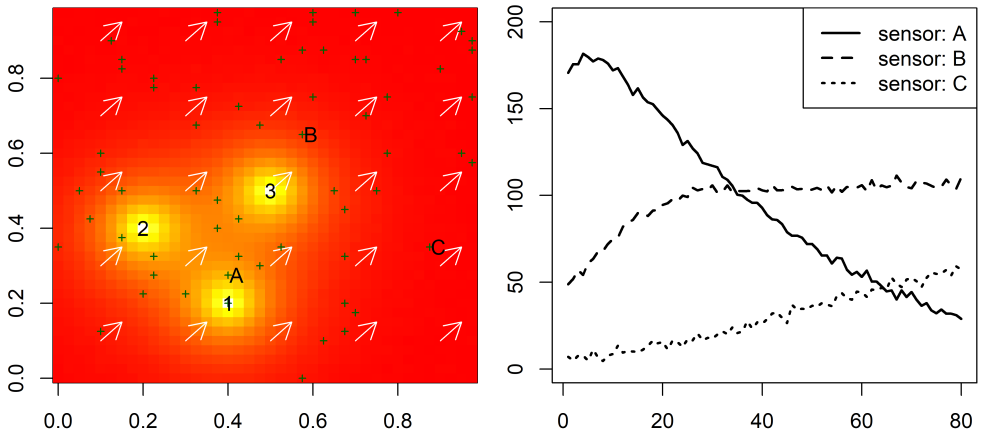

图 3：（左图）具有三个瞬时源的初始条件、速度场（以白色箭头表示）以及 64 个随机分布的传感器的位置；（右图）从选定的三个传感器“A”“B”和“C”获取的随时间变化的含噪传感器测量。

图 3 （左图）显示了初始条件、速度场（由箭头表示）以及 64 个随机分布的传感器位置（由小十字表示）。图 3 还突出显示了三个选定传感器“A”“B”和“C”的位置，随时间变化的测量值则显示在该图的右图中。测量误差服从 独立同分布的 均值为零、标准差为 2 的正态分布。传感器“A”发出的信号强度在过程（主要来自源1）迅速到达位置“A”时首先增强。随后，随着过程向外传播，信号逐渐减弱

并发生扩散。传感器“B”逐渐捕捉到信号（首先来自源 3，随后来自另外两个源），而传感器“C”由于远离三个源，缓慢地捕捉到相对较弱的信号。在整个空间域内缺乏时空过程“完整图景”的情况下，目标是估计初始条件。

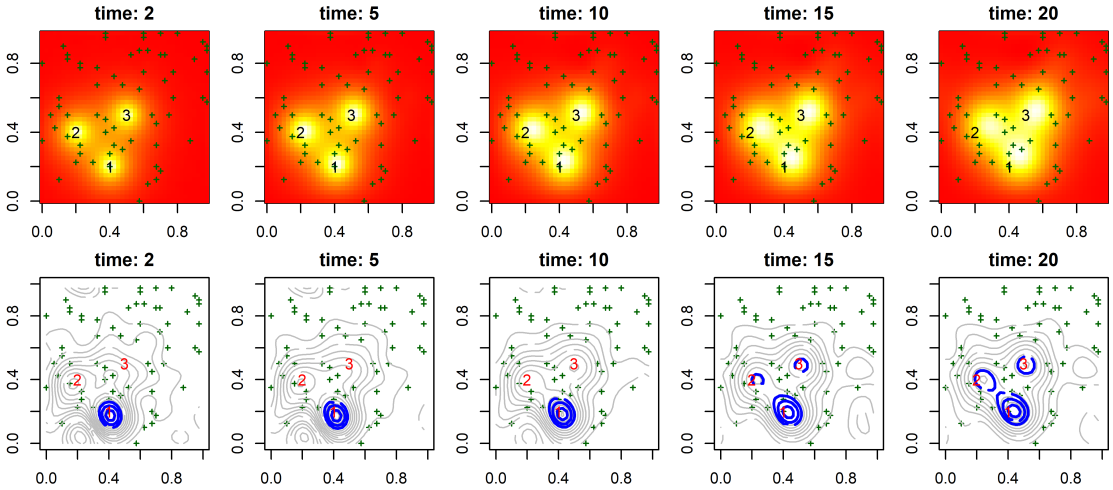

图4：（上排）对流扩散过程在时刻 2、5、10、15 和 20 的快照；（下排）基于来自 64 个传感器网络、截至时刻 2、5、10、15 和 20 的流数据检测到的空间源。

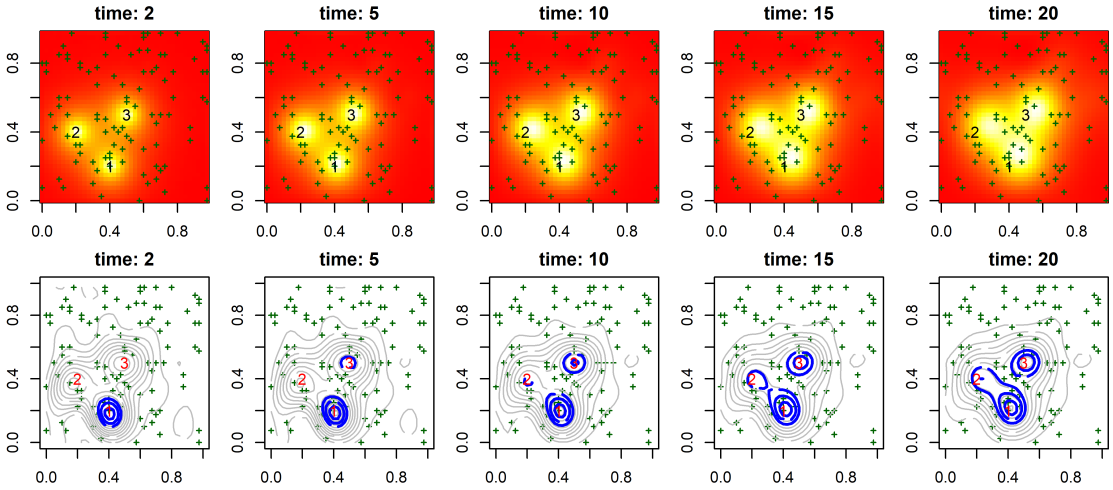

图 5：（上排）对流扩散过程在时刻 2、5、10、15 和 20 的快照；（下排）基于来自 100 个传感器网络、截至时刻 2、5、10、15 和 20 的流数据检测到的空间源。

图 4 的第一行展示了过程在时间 2、5、10、15 和 20 的快照。使用算法1求解不规则采样网格上的逆问题 P-I（10）后，

图4的第二行展示了利用截至时间 2、5、10、15 和 20 的传感器观测得到的估计的初始条件等高线图。粗蓝色水平集分别表示逆模型输出的第 75、85 和 95 百分位数。可以看到，源1在时间 2 时便迅速被检测到。这仅仅是因为恰好有一个传感器位于源1附近。可以看到，源2似乎也已被检测到（被一条等值线圈出）。然而，源2的估计强度弱于源1。时间 2 时完全无法检测到源 3，因为大多数传感器尚未接收到来自该源的任何信号。时间 15 时，源2和源 3 均被清晰检测到，因为下游传感器已接收到来自这两个源的信号。

预计扩大传感器网络的规模可以减少检测延迟。我们在现有传感器网络中随机增加了另外36个传感器（请注意，只有添加到源下游区域的传感器才可能有助于减少检测延迟）。图5 展示了更新后的结果：增加这36个传感器后，利用截至时间10的数据流即可检测到全部三个源。图4和5 充分展示了基于空间分布式传感器数据流的逆问题的动态特性。

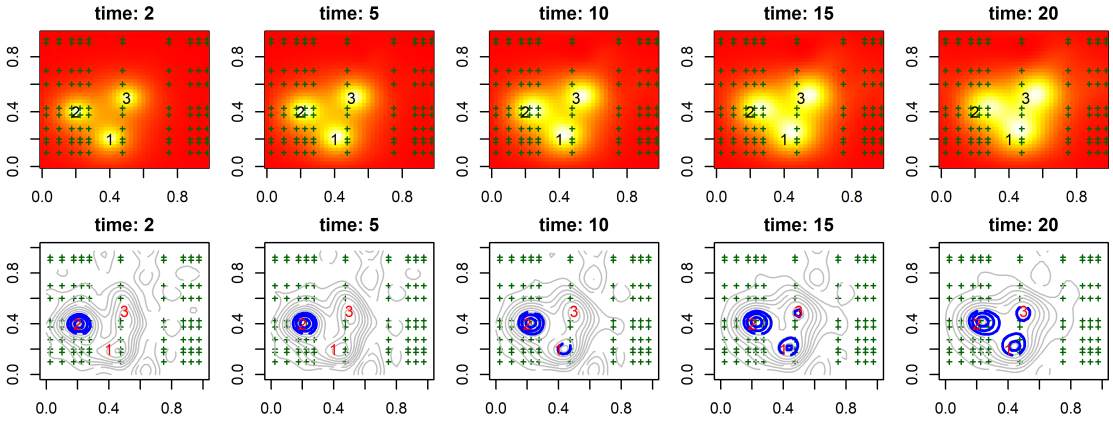

图6：（顶行）模拟对流扩散过程在时间2、5、10、15和20时的快照；（底行）基于来自非均匀采样网格、截至时间2、5、10、15和20的流数据检测到的空间源。

接下来，我们研究非均匀采样网格和移位采样网格下的逆问题P-II（21）和 P-III（32）。图6 展示了问题 P-II 的输出，其依据是

来自非均匀采样网格的传感器数据流，该网格由一个 10×10网格系统构成，它由 40×40 均匀网格系统生成，如第3.1节所述。与图4类似，图6 的第一行展示了过程在时间 2、5、10、15 和 20 时的快照，第二行则展示了利用截至时间 2、5、10、15 和 20 的流式观测得到的估计初始条件等高线图。粗蓝色水平集分别表示逆模型输出的第 75、85 和 95 百分位数。我们看到，由于源2靠近周边传感器，它几乎立即被检测到。源1和源3则较晚才被检测到，分别是在时间 10 和 15，此时下游区域的传感器才接收到这两个源发出的信号。

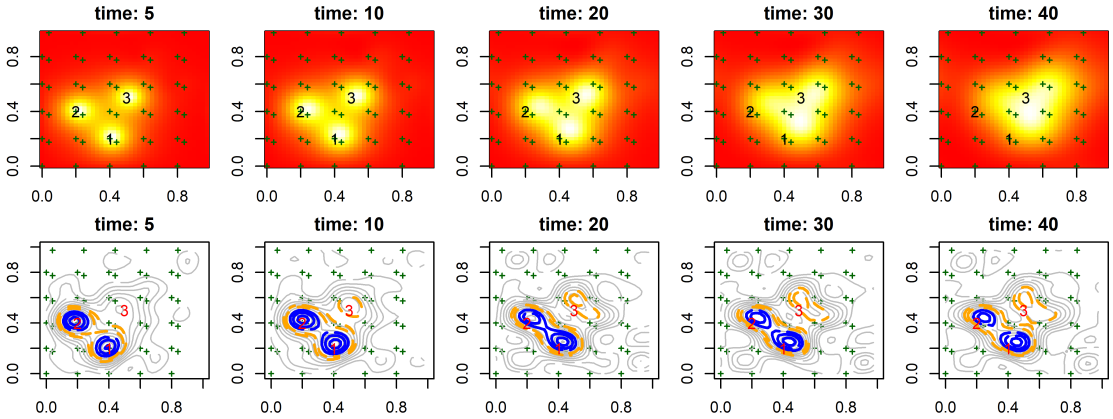

图7：（上排）模拟的对流扩散过程在时间 5、10、20、30 和 40 时的快照；（下排）基于来自两个平移均匀采样网格、截至时间 5、10、20、30 和 40 的流数据检测到的空间源。

图 7 展示了基于两个移位采样网格的传感器数据流得到的逆问题 P-III 动态输出。第一个采样网格是一个 5×5网格系统，而第二个网格则通过将第一个网格平移 $\boldsymbol{\delta} = (0.04,0.175)$ 获得。图 7第一行展示了时刻 5、10、20、30 和 40 的过程，第二行则展示了截至时刻 5、10、20、30 和 40 的模型输出。类似地，实线蓝色粗水平集分别表示模型生成输出的第 75、85 和 95 百分位数，而虚线粗水平集表示逆模型生成输出的第 50 和 60 百分位数。我们看到，源 1 和源 2 在时刻 5 就被迅速检测到，这是因为

它们靠近附近的传感器。源 3 则较晚才被检测到，因为源 3 下游区域的传感器此时才接收到信号。

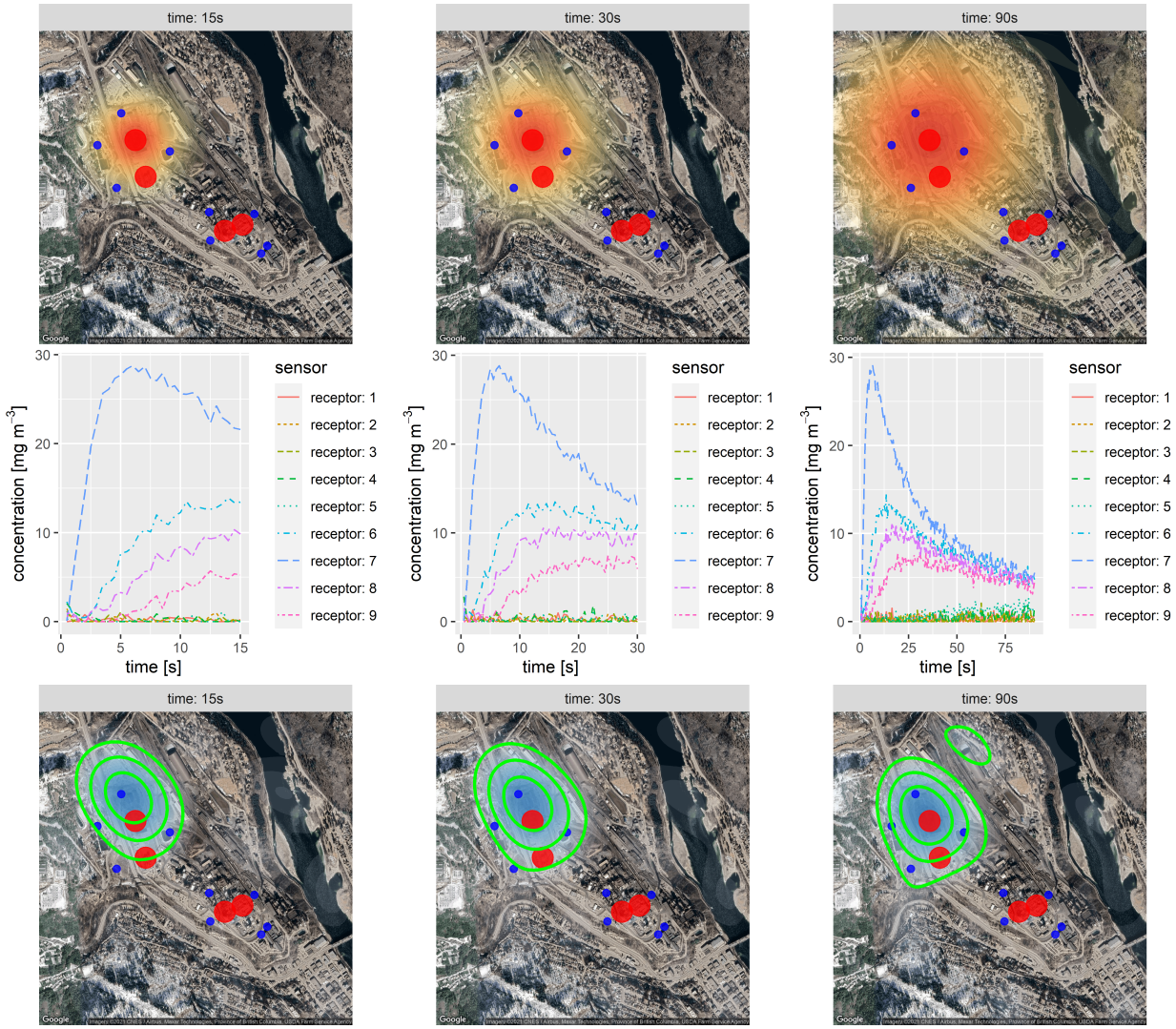

图8：纯扩散过程的反演建模：(a) 污染物释放后15、30和90秒的浓度；(b) 受体的地面测量值；(c) 逆模型输出的等高线图

示例I对所提出的逆模型提供了有益的见解，并利用空间分布式数据流成功揭示了逆问题的动态特性。当下游区域（如果存在）中的传感器检测到源发出的信号时，即可检测到该源。在补充材料中，我们比较了偏差

ˆ以及不同正则化方法下估计的初始条件 $\xi(0,\ \boldsymbol{s})$ 的均方误差（MSE），这些方法包括所提出的正则化、广义Lasso、弹性网络、 L_1和 L_2 正则化。

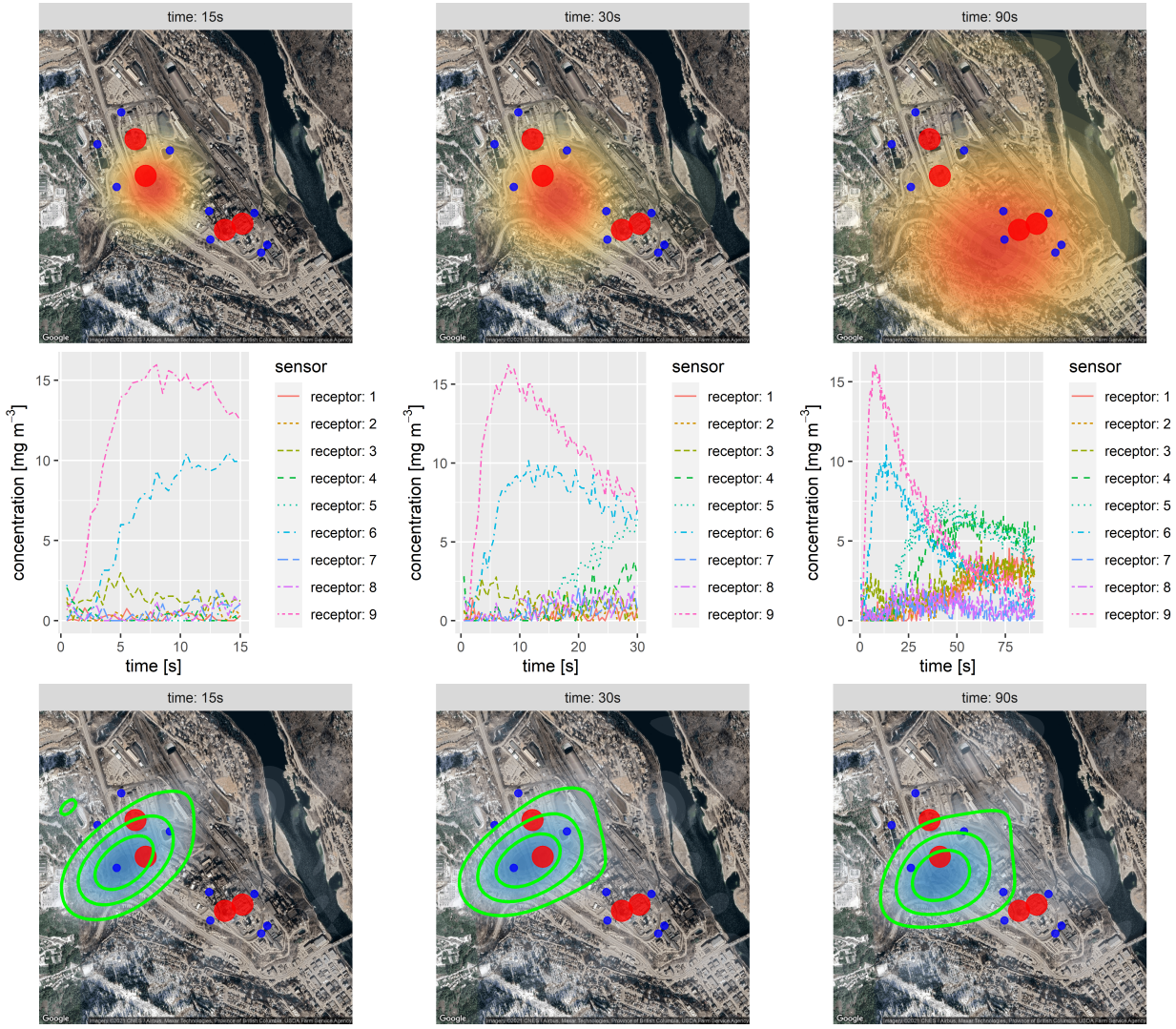

图9：对流扩散过程的反演建模：(a) 污染物释放后 15、30 和 90 秒的浓度；(b) 受体的地面测量值；(c) 逆模型输出的等高线图

5.2 示例 II

示例 II 重访了第 1.1 节中所述的启发性示例。该示例关注利用传感器估算意外 $ZnSO_4$ 泄漏的排放位置

监测数据；更多详情请参见第1.1节以及 Hosseini 和 Stockie (2016)。

大气中单一污染物的释放与输运可用平流-扩散方程很好地描述： $c(t,\boldsymbol{s}) + \vec{\boldsymbol{v}}^T \nabla c(t,\boldsymbol{s}) - \nabla \cdot [\boldsymbol{D} \nabla c(t,\boldsymbol{s})] = \Phi(t,\boldsymbol{s})$，其中 $c(t,\boldsymbol{s})$ 为污染物浓度，单位为 $[\text{mg m}^{-3}]$； $\vec{\boldsymbol{v}}$ 为风向量，单位为 $[\text{m s}^{-1}]$； $\boldsymbol{D}$ 为扩散率，单位为 $[\text{m}^2\,\text{s}^{-1}]$； $\Phi(t,\boldsymbol{s}) = \delta(t-0) \sum_{j=1}^{4} \phi_j(\boldsymbol{s})$为瞬时排放源，单位为$[\text{mg m}^{-3}\ \text{s}^{-1}]$。该方程是第 2 节中一般形式 (1) 的一个特例。

- 情景 1：纯扩散。首先考虑这样一种情景：$ZnSO_4$释放后，其传播由纯扩散过程驱动（即无风的情况）。图8 第一行显示了污染物释放后 15、30 和 90 秒时的浓度。如图所示，污染物从北面的第一个源释放，并按照纯扩散过程向各个方向传播。图8 第二行显示了随时间变化的含噪测量数据。图8 第三行给出了逆模型（10）利用截至 15、30 和 90 秒的流式观测生成的输出等高线图。如图8第一、二行所示，位于源北面的传感器最先检测到信号。因此，在 15 秒和 30 秒时，等高线图所示的峰值位于该传感器与实际源之间的某处。随着靠近该源的另外三个传感器提供更多数据（图8第二行），到 90 秒时，峰值逐渐接近实际源，并成功识别出排放源。这些观测证实了基于传感器数据流的所提反问题具有动态特性。

- 情景 2：平流与扩散。现在我们考虑一种更常见的情景：$ZnSO_4$释放后，其传播同时受到平流和扩散的驱动。污染物从空间域顶部数起的第二个源释放，并在风的作用下向东南方向传播。图9 第一行显示了污染物释放后 15、30 和 90 秒时的浓度。第二行显示了带噪声的

地面测量数据。第三行展示了逆模型（10）根据截至 15、30 和 90 秒的数据生成的输出等高线图。如图9的第一行和第二行所示，位于源西侧的传感器最先检测到信号。因此，在 15 秒和 30 秒时，等高线图所示的峰值位于该传感器与实际源之间的某处。随着位于源附近的其他传感器（见图9第二行）提供更多数据，到 90 秒时，峰值逐渐靠近实际源，准确定位了排放源。

• 敏感性分析。输入速度的准确性，即对流扩散算子 $\mathcal{A}$ 中的 $\vec{v}$，会显著影响逆模型的性能。设想在上述第二种情形中，如果指定的风向与真实风向相差甚远，我们就不能再期待模型给出准确结果。因此，我们进行敏感性分析，以考察模型对输入速度误设以及正则化调优参数 $\mathcal{R}$ 的稳健性。设 $\vec{v} = (v_x,\ v_y)^T = (10,\ -10)^T$ 为实际风向量 $[\mathrm{ms}^{-1}]$，而 $\vec{v}_{\text{输入}} = (v_x + \epsilon_{\text{风}x},\ v_y + \epsilon_{\text{风}y})^T$ 为输入逆模型且含有随机误差的风向量，$\epsilon_{\text{风}_x} \sim N(0,\ |v_x|\tau)$ 和 $\epsilon_{\text{风}_y} \sim N(0,\ |v_y|\tau)$；也就是说，误差方差与风向量的大小成正比，比例因子为 τ。

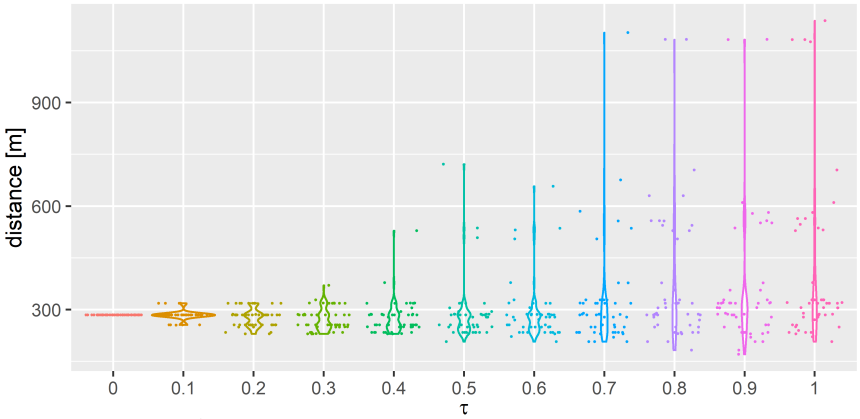

图10 ：针对输入速度矢量设定错误的敏感性分析 or.

考虑上述第二种情形，图10 展示了不同 τ 取值范围内检测源与实际源之间距离（单位：米）的小提琴图

从 0.1 到 1。对于每个 τ取值，实验重复进行 50 次，每次运行都会模拟一个新的输入风向量。可以看出，当 $\tau \leq 0.3$时，检测源与实际源之间的距离保持在 225 米至 380 米范围内。换言之，当输入风向量与实际风向量的偏差不大时，模型性能似乎较为稳健。在本问题中，由于实际风向量的水平和垂直分量均为 10 米/秒，因此当$^{-1}$时，输入风向量在水平方向和垂直方向上所含随机误差的标准差均为 1 米至 3 米。 $0.1 \leq \tau \leq 0.3$这种规格误差范围是合理的，在许多风速和风向均可观测的应用中都能达到（Ding 等人，2021年）。当 $\tau \geq 0.4$时，我们注意到检测误差的方差显著增大，这表明逆模型的性能如预期般迅速下降。

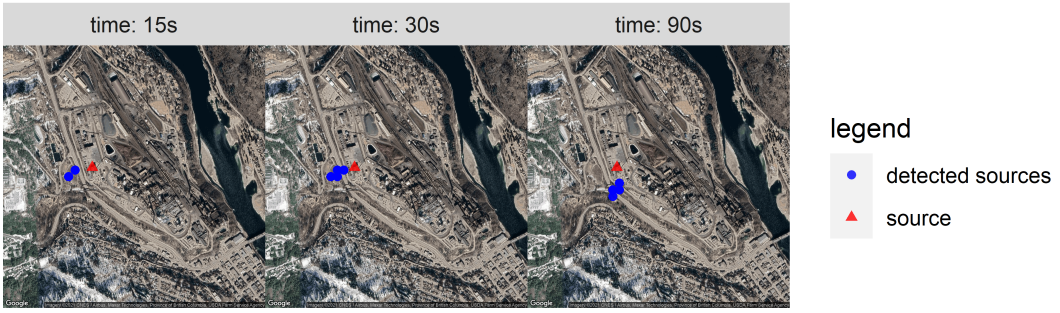

图11：在网格 $\{5, 10, 15, \cdots, 50\} \otimes \{5, 10, 15, \cdots, 50\}$中选取 λ_1和 λ_2的组合，对15、30和90秒时的检测源进行敏感性分析。

模型的表现还取决于正则化中调优参数的选择， λ_1和 λ_2。 $\mathcal{R}$需要注意的是，对于许多正向预测问题，例如 Lasso 回归、Ridge 回归、弹性网络等，可以通过交叉验证来选择调优参数。然而，由于真实源未知，也无法建立调优参数与检测精度之间的联系，交叉验证的思路不再适用于反问题。因此，研究所提出的逆模型对 λ_1 和 λ_2的稳健性（敏感性）更有意义。图11展示了在 15、30 和 90 秒时， λ_1 和

λ_2取自一个网格 $\{5, 10, 15, \cdots, 50\} \otimes \{5, 10, 15, \cdots, 50\}$。图中显示，检测源对调优参数的选择具有足够的稳健性。换言之，从相对宽泛的范围内选取不同组合的 λ_1和 λ_2，都能正确识别真实源，这在实践中无疑是理想的。

最后值得注意的是，如果初始条件严格用狄拉克函数建模（例如“点”源），时空过程的傅里叶级数将不收敛，所提出的方法可能会产生一种振荡解，称为 Gibbs 现象（Gibbs，1898年）。因此，如上文数值示例所示，若初始条件是光滑函数，所提出的方法效果良好。即使初始条件由点源构成，该模型也会在释放时刻稍后的时间点求解，此时点源因扩散而变得更为光滑。

6 结论

基于一种针对时空数据的基于偏微分方程的统计模型，本文提出了一种利用由三种空间采样方案生成的数据流对流扩散过程进行反演建模的方法。本文获得了能够唯一估计对流扩散过程初始条件的傅里叶系数的充要条件。本文还推导出了交替方向乘子法的详细迭代步骤，该方法能够以较高的计算效率求解反问题。此外，该算法还扩展为可处理关于模型输出的线性不等式约束。数值示例展示了所提出的逆模型对输入模型参数的稳健性，并揭示了基于传感器数据流的逆问题的动态特性。需要指出的是，本文考虑的是源自确定性偏微分方程的数据。一个重要的未来研究方向是将所提出的模型扩展至随机偏微分方程，旨在最小化正向分布与观测数据分布之间的距离。计算机代码可在 https://github.com/dnncode/inverse-model 获取。

致谢

我们感谢两位审稿人、副编辑和编辑提出的建设性意见，这些意见提升了论文的质量。

资助

本材料基于美国国家科学基金会通过编号为 2143695 的资助所支持的研究工作。

补充材料

补充材料提供了：(i) 命题1和2的证明，(ii) ADMM算法1的推导，(iii) 带非负约束的交替方向乘子法，(iv) 关于问题 P-I、P-II 和 P-III 计算时间的讨论，以及 (v) 不同正则化方案的数值比较。

参考文献

Apostol, B. F. (2019), “地震学中的一个反问题：从P波和S波推导震源参数，”《地震学杂志》, 23, 1017–1030. Beyrouthy, T., Fesquet, L., and Rolland, R. (2015), “数据采样与处理：均匀与非均匀方案，”*2015*年国际事件控制、通信与信号处理会议（*EBCCSP*），*DOI: 10.1109/EBCCSP.2015.7300665*. Chen, J., Kang, L., and Lin, G. (2021), “高斯过程辅助的物理定律主动学习，”*Technometrics*, 63, 329–342. Constantinescu, E. M., Petra, N., Bessac, J., and Petra, C. G. (2019), “受含随机项的基于微分方程模型约束的反问题的统计处理，”*arXiv: 1810.08557v2*. Deng, X., Lin, C. D., Liu, K. W., and Rowe, R. K. (2017), “用于包含定性和定量因素的计算机模型的加性高斯过程，”*Technometrics*, 59, 283–292. Ding, Y., Kumar, N., Prakash, A., Kio, A., Liu, X., Liu, L., and Li, Q. C. (2021), “风电场风力涡轮机时空性能比较案例研究，”《可再生能源》, 171, 735–746.

Eckhardt, S., Prata, A. J., Stebel, S. K., and Stohl, A. (2008), “利用卫星柱浓度测量和逆向输送建模估算火山喷发向大气中注入的二氧化硫垂直廓线，” 大气化学与物理, 8, 3881–3897. Gibbs, J. W. (1898), “傅里叶级数，” *https://doi.org/10.1038/059200b0*. Gramacy, R. B. (2020), 代理模型：应用科学的高斯过程建模、设计与优化, 博卡拉顿，佛罗里达州：Chapman & Hall/CRC. Gul, E., Joseph, R., Yan, H., and Melkote, S. N. (2018), “利用原位模拟器对加工仿真进行不确定性量化，” 质量技术期刊, 50, 253–261. Hosseini, B. and Stockie, J. M. (2016), “利用高斯烟羽模型估算空气中的无组织排放，” 大气环境, 141, 122–138. Hung, Y., Joseph, R., and Melkote, S. N. (2015), “具有函数响应的计算机实验分析，” *Technometrics*, 57, 35–44. Hwang, Y. D., Kim, H. J., Chang, W., Yeo, K. M., and Kim, Y. (2019), “通过逆物理模型进行贝叶斯污染源识别，” 计算统计与数据分析, 134, 76–92. Kadri, U. (2019), “海底弹性对撞击物体产生的声重力波传播的影响，” 科学报告, 9, 912. Klein, L. J., Muralidhar, R., Marianno, F. J., Chang, J. B., Lu, S. Y., and Hamann, H. F. (2016), “地理空间物联网：甲烷无组织泄漏监测框架，”载于国际地理信息科学会议短文论文集. Liu, X., Yeo, K. M., and Lu, S. Y. (2022), “物理对流扩散过程产生的时空数据统计建模，” 美国统计学会期刊（在线可查）, 117, 1482–1499. Mak, S., Sung, C. L., Wang, X., Yeh, S. T., Chang, Y. H., Joseph, R., Yang, V., and Wu, C. F. J. (2018), “用于设计评估和物理规律提取的大涡模拟高效代理模型，” 美国统计学会期刊, 113, 1443–1456. Martinez-Camara, M., Haro, B. B., Stohl, A., and Vetterli, M. (2014), “利用盲异常值检测对大气排放进行逆向输送建模的稳健方法，”地球科学模型开发, 7, 2303–2311. Miron, P., Beron-Vera, F. J., Olascoaga, M. J., and Koltai, P. (2019), “受马尔可夫链启发的MH370搜索，” *Chaos*, 29, 041105. Oates, C. J., Cockayne, J., Aykroyd, R. G., and Girolami, M. (2019), “用于工业水力旋流器设备时变状态估计的贝叶斯概率数值方法，”美国统计学会期刊, 114, 1518–1531. Pal, P. and Vaidyanathan, P. P. (2010), “嵌套阵列：一种提升自由度的阵列处理新方法，” *IEEE*信号处理汇刊, 58, 4167–4181.

Qian, E., Karamer, B., Peherstorfer, B., 和 Willcox, K.（2019），“Lift & Learn：面向大规模非线性动力系统的物理信息机器学习”，*Oden*研究所报告 *19-18*。Qin, G. 和 Amin, M. G.（2021），“利用两个均匀子阵的移动平台DOA估计结构化稀疏阵列设计”，《信号处理》，180，107872。Raissi, M., Perdikaris, P., 和 Karniadakis, G. E.（2019），“物理信息神经网络：用于求解涉及非线性偏微分方程的正问题和逆问题的深度学习框架”，《计算物理学杂志》，378，686–707。Ramdas, A. 和 Tibshirani, R. J.（2016），“趋势滤波的快速灵活ADMM算法”，《计算与图形统计学杂志》，25，839–858。Sauer, A., Gramacy, R. B., 和 Higdon, D.（2021），“深度高斯过程代理模型的主动学习”，*arXiv:2012.08015*。Sigrist, F., Kunsch, H. R., 和 Stahel, W. A.（2015），“基于随机偏微分方程的大规模时空数据集建模”，《皇家统计学会杂志：*B*辑》，77，3–33。Tibshirani, R., Rosset, S., Zhu, J., 和 Knight, K.（2005），“通过融合套索实现稀疏性与光滑性”，《皇家统计学会杂志：*B*辑（统计方法）》，67，91–108。Venkataramani, R. 和 Bresler, Y.（2001），“多频带信号的最优次奈奎斯特非均匀采样与重建”，*IEEE*信号处理汇刊，49，2301–2313。Yao, B. 和 Yang, H.（2021），“逆向心电图建模的时空正则化”，《*IISE*医疗保健系统工程汇刊》，11，11–23。Yeo, K. M., Hwang, Y. D., Liu, X., 和 Kalagnanam, J.（2019），“利用广义多项式混沌开发hp逆模型”，《计算机方法及应用力学与工程》，347，1–20。Zhang, B. Y., Cole, D. A., 和 Gramacy, R. B.（2021），“高斯过程代理模型的距离分布设计”，*arXiv:1812.02794*。Zou, H. 和 Hastie, T.（2005），“通过弹性网络进行正则化与变量选择”，《皇家统计学会杂志：*B*辑（统计方法）》，67，301–320。

用于估计时空对流-扩散过程初始条件的逆模型

刘晓

阿肯色大学工业工程系

Kyongmin Yeo

IBM T. J. 沃森研究中心

补充材料

1. 命题1的证明

证明。 为证明命题1，注意到

- 对于任意满足 $\vec{\boldsymbol{v}}^T(\boldsymbol{k}_{j_1} - \boldsymbol{k}_{j_2}) = 0$ 且$\boldsymbol{k}_{j_1}^T \boldsymbol{D} \boldsymbol{k}_{j_1} = \boldsymbol{k}_{j_2}^T \boldsymbol{D} \boldsymbol{k}_{j_2}$的 $\boldsymbol{k}_{j_1}$和 $\boldsymbol{k}_{j_2}$（$\boldsymbol{k}_{j_1}, \boldsymbol{k}_{j_2} \in \mathcal{K}$和 $j_1 \neq j_2$），由 $\gamma_j = -\boldsymbol{k}_j^T \boldsymbol{D} \boldsymbol{k}_j - \zeta - \imath \vec{\boldsymbol{v}}^T \boldsymbol{k}_j$ 可立即推出，两个向量（$g_{j_1}(1) \ldots, g_{j_1}(L)$）和（$g_{j_2}(1) \ldots, g_{j_2}(L)$）变得相同；

- 对于给定的 $\boldsymbol{s}$，复指数 $e^{\imath 2\pi \boldsymbol{s}^T \boldsymbol{k}_{j_1}}$和 $e^{\imath 2\pi \boldsymbol{s}^T \boldsymbol{k}_{j_2}}$在复数域 $\mathbb{C}$中线性相关。例如，存在 $u, v \in \mathbb{R}$ （其中至少一个不为零），使得 $(u + v\imath)\, e^{\imath 2\pi \boldsymbol{s}^T \boldsymbol{k}_{j_1}} = (u - v\imath)\, e^{\imath 2\pi \boldsymbol{s}^T \boldsymbol{k}_{j_2}}$；

- 对于一对采样位置 $\boldsymbol{s}$ 和 $\boldsymbol{s}'$，如果 $2\boldsymbol{k}_{j_1}(\boldsymbol{s} - \boldsymbol{s}')$ 和 $2\boldsymbol{k}_{j_2}(\boldsymbol{s} - \boldsymbol{s}')$ 返回的都是偶数，那么，$e^{\imath 2\pi \boldsymbol{s}^T \boldsymbol{k}} = e^{\imath 2\pi \boldsymbol{s}'^T \boldsymbol{k}}$。类似地，如果 $2\boldsymbol{k}_{j_1}(\boldsymbol{s} - \boldsymbol{s}')$ 和 $2\boldsymbol{k}_{j_2}(\boldsymbol{s} - \boldsymbol{s}')$ 返回的都是奇数，那么，$e^{\imath 2\pi \boldsymbol{s}^T \boldsymbol{k}} = -e^{\imath 2\pi \boldsymbol{s}'^T \boldsymbol{k}}$。

因此，矩阵 $\mathscr{X}^{(\text{P-I})}$ (9) 在命题1中的条件A和条件B之一成立时，具有线性无关的列。当 $ML \geq N$，$\mathscr{X}_{(\text{P-I})}$的秩为 N且满足命题1中的条件时， $\boldsymbol{\eta}$ 中的分量也可以被唯一确定。
■

2. 命题2的证明

证明：我们首先通过从 $\boldsymbol{G}$ 中消除冗余行（如果存在），构造一个新的矩阵 $\tilde{\boldsymbol{G}}$，如下所示：设 $\boldsymbol{g}_{i,\cdot}$ 为第 i 行 $\boldsymbol{G}$。对于 $i=2,\ ...N$，如果存在 $i'=1,\ ...,\ i-1$ 使得 $\boldsymbol{g}_{i,\cdot}=\boldsymbol{g}_{i',\cdot}$，则从 $\boldsymbol{G}$ 中消去行 $\boldsymbol{g}_{i,\cdot}$。从 $\boldsymbol{G}$ 中消去重复行后，当 $L \geq N$ 时，矩阵 $\tilde{\boldsymbol{G}}$ 为行满秩。

设 $\boldsymbol{H}=(\boldsymbol{h}_{\cdot,1},\boldsymbol{h}_{\cdot,2},...,\boldsymbol{h}_{\cdot,\tilde{N}})$ 为一个 $N\times N$ 矩阵，其中 $\boldsymbol{h}_{\cdot,j}$ 表示 j 的第 $\boldsymbol{H}$ 个列向量。对于任意列向量 $\boldsymbol{h}_{\cdot,j}$，其第 i 个元素 $h_{i,j}=1$ 若 $i\in\Psi_j$；否则，$h_{i,j}=0$。于是，(4)可重写为

$$\boldsymbol{Y}=\boldsymbol{F}(\boldsymbol{E}\boldsymbol{H})\tilde{\boldsymbol{G}}+\boldsymbol{V}. \tag{38}$$

由于 $\tilde{\boldsymbol{G}}$ 是 N 的行满秩（当 $L\geq N$ 时），我们将（38）重写如下：

$$\boldsymbol{Y}\tilde{\boldsymbol{G}}_R^{-1}=\boldsymbol{F}(\boldsymbol{E}\boldsymbol{H})+\boldsymbol{V}\tilde{\boldsymbol{G}}_R^{-1} \tag{39}$$

$\tilde{\boldsymbol{G}}_R^{-1}$ 是 $\tilde{\boldsymbol{G}}$ 的右逆。

设 $\boldsymbol{Y}\tilde{\boldsymbol{G}}_R^{-1}\equiv\tilde{\boldsymbol{Y}}=(\tilde{\boldsymbol{Y}}_1,\ \tilde{\boldsymbol{Y}}_2,\ .....,\ \tilde{\boldsymbol{Y}}_{\tilde{N}})$，$\tilde{\boldsymbol{F}}_i=\{f_{m,j}\}_{m=1,\ .....,\ M,\ j\in\Psi_i}$ 为一个 $M\times|\Psi_i|$ 矩阵，且 $\boldsymbol{V}\tilde{\boldsymbol{G}}_R^{-1}\equiv\boldsymbol{\Omega}=(\boldsymbol{\Omega}_1,\boldsymbol{\Omega}_2,\ .....,\ \boldsymbol{\Omega}_{\tilde{N}})$。那么，（39）定义了一个 $\tilde{N}$ 线性模型

$$\tilde{\boldsymbol{Y}}_i=\tilde{\boldsymbol{F}}_i\boldsymbol{\eta}_i+\boldsymbol{\Omega}_i,\qquad \forall i=1,2,...,\tilde{N} \tag{40}$$

其中，$\boldsymbol{\eta}_i$ 是一个列向量 $\{\eta(\boldsymbol{k}_j)\}_{j\in\Psi_i}$。因此，要唯一确定 $\boldsymbol{\eta}$ 中的所有分量 $\tilde{\boldsymbol{\eta}}$，充分条件是对于所有 $i=1,\ .....,\ N$ ■，$\text{rank}(\boldsymbol{F}_i)=|\Psi_i|$。

3. ADMM算法1的推导

交替方向乘子法通过重复以下迭代来求解约束问题（34）（Zou and Hastie，2005；Ramdas and Tibshirani，2016）：

$$\boldsymbol{\eta}^{(i)} = \text{argmin}_{\boldsymbol{\eta}} f(\boldsymbol{\eta}) + \frac{\rho}{2}\left\|\boldsymbol{\eta} - \boldsymbol{\psi}^{(i-1)} + \boldsymbol{u}^{(i-1)}\right\|_2^2 \tag{41a}$$

$$\boldsymbol{\psi}^{(i)} = \text{argmin}_{\boldsymbol{\psi}} \mathcal{R}(\boldsymbol{\psi}) + \frac{\rho}{2}\left\|\boldsymbol{\eta}^{(i)} - \boldsymbol{\psi} + \boldsymbol{u}^{(i-1)}\right\|_2^2 \tag{41b}$$

$$\boldsymbol{u}^{(i)} = \boldsymbol{u}^{(i-1)} + \boldsymbol{\eta}^{(i)} - \boldsymbol{\psi}^{(i)} \tag{41c}$$

对于 $i = 1, 2, \cdots$。注意，（41a）和（41b）中的两个最小化问题可以通过以下方式高效求解。

求解（41a）。对于（41a），可以证明:

$$\begin{aligned}
\boldsymbol{\eta}^{(i)} &= \text{argmin}_{\boldsymbol{\eta}} \frac{1}{2}\langle(\vec{\mathscr{Y}} - \mathscr{X}\boldsymbol{\eta}), \boldsymbol{\Sigma}^{-1}(\vec{\mathscr{Y}} - \mathscr{X}\boldsymbol{\eta})\rangle + \frac{\rho}{2}\langle\boldsymbol{\eta} - \boldsymbol{\psi}^{(i-1)} + \boldsymbol{u}^{(i-1)}, \boldsymbol{\eta} - \boldsymbol{\psi}^{(i-1)} + \boldsymbol{u}^{(i-1)}\rangle \\
&= \text{argmin}_{\boldsymbol{\eta}} - \frac{1}{2}\langle\boldsymbol{\eta}, \mathscr{X}^T\boldsymbol{\Sigma}^{-1}\vec{\mathscr{Y}}\rangle - \frac{1}{2}\langle\boldsymbol{\eta}, (\boldsymbol{\Sigma}^{-1}\mathscr{X})^T\vec{\mathscr{Y}}\rangle + \frac{1}{2}\langle\boldsymbol{\eta}, \mathscr{X}^T\boldsymbol{\Sigma}^{-1}\mathscr{X}\boldsymbol{\eta}\rangle \\
&\qquad + \frac{\rho}{2}\langle\boldsymbol{\eta}, \boldsymbol{\eta}\rangle - \frac{1}{2}\langle\boldsymbol{\eta}, \boldsymbol{\psi}^{(i-1)}\rangle + \frac{1}{2}\langle\boldsymbol{\eta}, \boldsymbol{u}^{(i-1)}\rangle \\
&= \frac{1}{2}\langle\boldsymbol{\eta}, (\mathscr{X}^T\boldsymbol{\Sigma}^{-1}\mathscr{X} + \frac{\rho}{2}\boldsymbol{I})\boldsymbol{\eta}\rangle + \langle\boldsymbol{\eta}, -\frac{1}{2}(\mathscr{X}^T\boldsymbol{\Sigma}^{-1} + (\boldsymbol{\Sigma}^{-1}\mathscr{X})^T)\vec{\mathscr{Y}} - \rho(\boldsymbol{\psi}^{(i-1)} - \boldsymbol{u}^{(i-1)})\rangle
\end{aligned} \tag{42}$$

其中，$\langle\cdot,\cdot\rangle$ 表示向量空间中的内积。

由于（42）右侧的梯度向量可表示为

$$(\mathscr{X}^T\boldsymbol{\Sigma}^{-1}\mathscr{X} + \frac{\rho}{2}\boldsymbol{I})\boldsymbol{\eta} - \frac{1}{2}(\mathscr{X}^T\boldsymbol{\Sigma}^{-1} + (\boldsymbol{\Sigma}^{-1}\mathscr{X})^T)\vec{\mathscr{Y}} - \rho(\boldsymbol{\psi}^{(i-1)} - \boldsymbol{u}^{(i-1)}), \tag{43}$$

我们得到　　将上述梯度向量设为零，即可得到（41a）的闭式解:

$$\boldsymbol{\eta}^{(i)} = (\mathscr{X}^T\boldsymbol{\Sigma}^{-1}\mathscr{X} + \frac{\rho}{2}\boldsymbol{I})^{-1}\left\{\frac{1}{2}(\mathscr{X}^T\boldsymbol{\Sigma}^{-1} + (\boldsymbol{\Sigma}^{-1}\mathscr{X})^T)\vec{\mathscr{Y}} - \rho(\boldsymbol{\psi}^{(i-1)} - \boldsymbol{u}^{(i-1)})\right\}. \tag{44}$$

求解（41b）。我们将（41b）重写为

$$\begin{aligned}
\boldsymbol{\psi}^{(i)} &= \text{argmin}_{\boldsymbol{\psi}} \mathcal{R}(\boldsymbol{\psi}) + \frac{\rho}{2}\left\|\boldsymbol{\eta}^{(i)} - \boldsymbol{\psi} + \boldsymbol{u}^{(i-1)}\right\|_2^2 \\
&= \text{argmin}_{\boldsymbol{\psi}} \lambda_1\|\boldsymbol{\psi}\|_1 + \lambda_2\|\boldsymbol{J}\boldsymbol{\psi}\|_2^2 + \frac{\rho}{2}\left\|\boldsymbol{\eta}^{(i)} - \boldsymbol{\psi} + \boldsymbol{u}^{(i-1)}\right\|_2^2 \\
&= \text{argmin}_{\boldsymbol{\psi}} \lambda_1\|\boldsymbol{\psi}\|_1 + \lambda_2\langle\boldsymbol{\psi}, (\boldsymbol{J}^T\boldsymbol{J} + \frac{\rho}{2}\boldsymbol{I})\boldsymbol{\psi}\rangle - \rho\langle\boldsymbol{\psi}, \boldsymbol{u}^{(i-1)} + \boldsymbol{\eta}^{(i)}\rangle
\end{aligned} \tag{45}$$

优化问题（45）同样可以使用交替方向乘子法进行数值求解。将（45）转化为约束问题可得：

$$\min_{\tilde{\boldsymbol{\psi}},\boldsymbol{\theta}} \quad \lambda_2\langle\tilde{\boldsymbol{\psi}},(\boldsymbol{J}^T\boldsymbol{J}+\frac{\rho}{2}\boldsymbol{I})\tilde{\boldsymbol{\psi}}\rangle-\rho\langle\tilde{\boldsymbol{\psi}},\boldsymbol{u}^{(i-1)}+\boldsymbol{\eta}^{(i)}\rangle+\lambda_1\left\|\boldsymbol{\theta}\right\|_1, \qquad \text{s.t. } \boldsymbol{\psi}=\boldsymbol{\theta}. \tag{46}$$

然后，对于给定的 $\omega>0$，通过重复以下步骤来求解约束问题（对于 $j=1,2,\ldots\ldots$）：

$$\begin{aligned}\tilde{\boldsymbol{\psi}}^{(j)} &= \operatorname{argmin}_{\tilde{\boldsymbol{\psi}}}\lambda_2\langle\tilde{\boldsymbol{\psi}},(\boldsymbol{J}^T\boldsymbol{J}+\frac{\rho}{2}\boldsymbol{I})\tilde{\boldsymbol{\psi}}\rangle-\rho\langle\tilde{\boldsymbol{\psi}},\tilde{\boldsymbol{\psi}}+\boldsymbol{\eta}^{(i)}\rangle+\frac{\omega}{2}\left\|\tilde{\boldsymbol{\psi}}-\boldsymbol{\theta}^{(j-1)}+\boldsymbol{v}^{(j-1)}\right\|_2^2 \\ &= \operatorname{argmin}_{\tilde{\boldsymbol{\psi}}}\frac{1}{2}\langle\tilde{\boldsymbol{\psi}},2\lambda_2(\boldsymbol{J}^T\boldsymbol{J}+\frac{\rho+\omega}{2}\boldsymbol{I})\tilde{\boldsymbol{\psi}}\rangle-\rho\langle\tilde{\boldsymbol{\psi}},\boldsymbol{u}^{(i-1)}+\boldsymbol{\eta}^{(i)}+\omega\boldsymbol{\theta}^{(j-1)}-\omega\boldsymbol{v}^{(j-1)}\rangle \\ &= \frac{\rho}{2\lambda_2}(\boldsymbol{J}^T\boldsymbol{J}+\frac{\rho+\omega}{2}\boldsymbol{I})^{-1}(\boldsymbol{u}^{(i-1)}+\boldsymbol{\eta}^{(i)}+\omega\boldsymbol{\theta}^{(j-1)}-\omega\boldsymbol{v}^{(j-1)}),\end{aligned} \tag{47}$$

$$\begin{aligned}\boldsymbol{\theta}^{(j)} &= \operatorname{argmin}_{\boldsymbol{\theta}} \quad \lambda_1\left\|\boldsymbol{\theta}\right\|_1+\frac{\omega}{2}\left\|\tilde{\boldsymbol{\psi}}^{(j)}-\boldsymbol{\theta}+\boldsymbol{v}^{(j-1)}\right\|_2^2 \\ &= S_{\lambda_1/\omega}(\tilde{\boldsymbol{\psi}}^{(j)}+\boldsymbol{v}^{(j-1)})\end{aligned} \tag{48}$$

其中 $S_{\lambda_1/\omega}(\cdot)$ 是一个软阈值算子

$$S_{\lambda_1/\omega}(x)=\begin{cases}x-\lambda_1/\omega, & \text{if } x>\lambda_1/\omega \\ x+\lambda_1/\omega, & \text{if } x<-\lambda_1/\omega \\ 0, & \text{otherwise}\end{cases} \tag{49}$$

且 $\boldsymbol{v}^{(j)}=\boldsymbol{v}^{(j-1)}+\tilde{\boldsymbol{\psi}}^{(j)}-\boldsymbol{\theta}^{(j)}$。

4. 带非负约束的交替方向乘子法

我们证明，通过修改第 4 节所述的交替方向乘子法（ADMM）算法，可以高效求解约束问题（37）。首先，通过引入惩罚函数和变量代换，将（37）中的不等式约束替换为等式约束：

$$\min_{\boldsymbol{\eta}} \quad \frac{1}{2}(\vec{\mathscr{Y}}-\mathscr{X}\boldsymbol{\eta})^T\boldsymbol{\Sigma}^{-1}(\vec{\mathscr{Y}}-\mathscr{X}\boldsymbol{\eta})+\mathcal{R}(\boldsymbol{\eta})+\mathcal{I}(\tilde{\boldsymbol{\eta}}) \qquad \text{s.t. } \mathscr{X}\boldsymbol{\eta}-\tilde{\boldsymbol{\eta}}=0 \tag{50}$$

其中，当 $\tilde{\boldsymbol{\eta}}\geq 0$ 时，$\mathcal{I}(\tilde{\boldsymbol{\eta}})=0$；否则，$\mathcal{I}(\tilde{\boldsymbol{\eta}})=\infty$。然后，对于 $\rho>0$，

（50）的增广拉格朗日函数的缩放形式写为：

$$f(\boldsymbol{\eta})+\mathcal{R}(\boldsymbol{\eta})+\mathcal{I}(\tilde{\boldsymbol{\eta}})+\frac{\rho}{2}\left\|\mathscr{X}\boldsymbol{\eta}-\tilde{\boldsymbol{\eta}}+\boldsymbol{u}\right\|_2^2+\frac{\rho}{2}\left\|\boldsymbol{u}\right\|_2^2. \tag{51}$$

其中 $f(\boldsymbol{\eta})=\frac{1}{2}(\vec{\mathscr{Y}}-\mathscr{X}\boldsymbol{\eta})^T\boldsymbol{\Sigma}^{-1}(\vec{\mathscr{Y}}-\mathscr{X}\boldsymbol{\eta})$。

类似于(41），求解（51）需要使用交替方向乘子法重复执行以下迭代 ：

$$\boldsymbol{\eta}^{(i)}=\operatorname{argmin}_{\boldsymbol{\eta}} f(\boldsymbol{\eta})+\mathcal{R}(\boldsymbol{\eta})+\frac{\rho}{2}\left\|\mathscr{X}\boldsymbol{\eta}-\tilde{\boldsymbol{\eta}}^{(i-1)}+\boldsymbol{u}^{(i-1)}\right\|_2^2 \tag{52a}$$

$$\tilde{\boldsymbol{\eta}}^{(i)}=\max(0,\mathscr{X}\boldsymbol{\eta}^{(i)}+\boldsymbol{u}^{(i-1)}) \tag{52b}$$

$$\boldsymbol{u}^{(i)}=\boldsymbol{u}^{(i-1)}+\mathscr{X}\boldsymbol{\eta}^{(i)}-\tilde{\boldsymbol{\eta}}^{(i)} \tag{52c}$$

对(56b)和(56c)的计算很简单，而计算上的难点在于最小化问题(68a)。幸运的是，再次使用交替方向乘子法，可以通过将(56a)转换为约束问题来求解(56a)：

$$\min_{\boldsymbol{\eta}} f(\boldsymbol{\eta})+\mathcal{R}(\boldsymbol{\psi})+\frac{\rho}{2}\left\|\mathscr{X}\boldsymbol{\psi}-\boldsymbol{c}\right\|_2^2,\quad \text{s.t. } \boldsymbol{\eta}=\boldsymbol{\psi} \tag{53}$$

其中 $\boldsymbol{c}=\tilde{\boldsymbol{\eta}}^{(i-1)}-\boldsymbol{u}^{(i-1)}$。然后，用于求解(53)的交替方向乘子法迭代如下：

$$\boldsymbol{\eta}^{(i)}=\operatorname{argmin}_{\boldsymbol{\eta}} f(\boldsymbol{\eta})+\frac{\rho}{2}\left\|\boldsymbol{\eta}-\boldsymbol{\psi}^{(i-1)}+\boldsymbol{v}^{(i-1)}\right\|_2^2 \tag{54a}$$

$$\boldsymbol{\psi}^{(i)}=\operatorname{argmin}_{\boldsymbol{\psi}} \mathcal{R}(\boldsymbol{\psi})+\frac{\rho}{2}\left\|\mathscr{X}\boldsymbol{\psi}-\boldsymbol{c}\right\|_2^2+\frac{\rho}{2}\left\|\boldsymbol{\eta}^{(i)}-\boldsymbol{\psi}+\boldsymbol{v}^{(i-1)}\right\|_2^2 \tag{54b}$$

$$\boldsymbol{v}^{(i)}=\boldsymbol{v}^{(i-1)}+\boldsymbol{\eta}^{(i)}-\boldsymbol{\psi}^{(i)} \tag{54c}$$

仔细考察(54)可得到以下关键观测结果：

- (59a) 与 (46a) 的形式完全相同，而 (59a) 的闭式解已由 (42) 给出。

- (59b) 可改写为

$$
\begin{aligned}
\boldsymbol{\psi}^{(i)} &= \operatorname{argmin}_{\boldsymbol{\psi}} \mathcal{R}(\boldsymbol{\psi}) + \frac{\rho}{2}\left\|\mathscr{X}\boldsymbol{\psi} - \boldsymbol{c}\right\|_2^2 + \frac{\rho}{2}\left\|\boldsymbol{\eta}^{(i)} - \boldsymbol{\psi} + \boldsymbol{v}^{(i-1)}\right\|_2^2 \\
&= \operatorname{argmin}_{\boldsymbol{\psi}} \lambda_1 \left\|\boldsymbol{\psi}\right\|_1 + \lambda_2 \left\|\boldsymbol{J}\boldsymbol{\psi}\right\|_2^2 + \frac{\rho}{2}\left\|\mathscr{X}\boldsymbol{\psi} - \boldsymbol{c}\right\|_2^2 + \frac{\rho}{2}\left\|\boldsymbol{\eta}^{(i)} - \boldsymbol{\psi} + \boldsymbol{v}^{(i-1)}\right\|_2^2 \\
&= \operatorname{argmin}_{\boldsymbol{\psi}} \lambda_1 \left\|\boldsymbol{\psi}\right\|_1 + \lambda_2 \langle \boldsymbol{\psi}, (\boldsymbol{J}^T\boldsymbol{J} + \frac{\rho}{2}\mathscr{X}^T\mathscr{X} + \frac{\rho}{2}\boldsymbol{I})\boldsymbol{\psi}\rangle - \rho\langle \boldsymbol{\psi}, \mathscr{X}\boldsymbol{c} + \boldsymbol{u}^{(i-1)} + \boldsymbol{\eta}^{(i)}\rangle
\end{aligned}
\tag{55}
$$

其形式与（45）完全相同，因此可以按照（46）至（49）中描述的相同步骤求解。

以上讨论表明，所提出的逆模型可以扩展以处理（针对模型输出的）线性不等式约束，从而使该模型能够适用于更广泛的问题。

5. 关于问题 P-I、P-II 和 P-III 的计算时间。

尽管问题 P-I、P-II 和 P-III 都可以通过 ADMM 算法求解，但需要注意的是，问题 P-I 是在时空域中表述的，而问题 P-II 和 P-III 则是在谱域中构建的。因此，$\mathscr{X}^{(\text{P-I})}$ 式（9）中的设计矩阵是稠密矩阵，而 $\mathscr{X}^{(\text{P-II})}$ 和 $\mathscr{X}^{(\text{P-III})}$ 式（20）和（31）中的设计矩阵是稀疏的（块对角），因此在 ADMM 算法中，计算 $\mathscr{X}^T\boldsymbol{\Sigma}^{-1}\mathscr{X}$、$\mathscr{X}^T\boldsymbol{\Sigma}^{-1}$和 $\boldsymbol{\Sigma}^{-1}\mathscr{X}$的速度更快。此外，与许多谱方法类似，问题 P-II 和 P-III 允许截断高频分量，因为式（$\mathscr{X}$ 20）和（31）中 $\mathscr{X}$的每个块都对应一个频率层级。这有助于进一步缩短计算时间，如图 12所示。

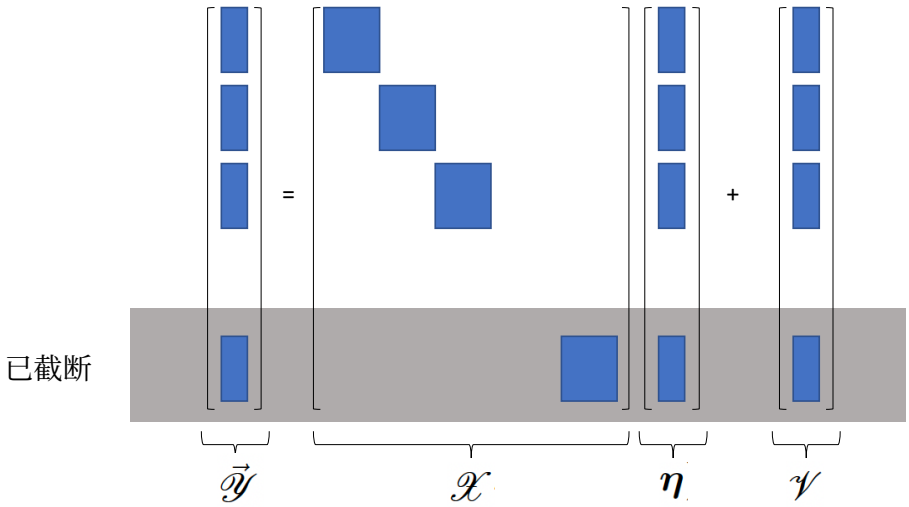

图12：在问题 P-II 和问题 P-III 中，$\mathscr{X}$ 是一个块对角矩阵，每个块对应一个频率；可以舍弃对应高频的块，以进一步减少计算时间。

6. 不同正则化选择的数值比较

ˆ 我们比较了采用不同正则化方式时，估计的初始条件 $\xi(0,\ s)$ 的偏差和均方误差，包括

方法1）。所提出的正则化

方法2）。广义Lasso

方法3）。弹性网络

方法4）。 L_1 仅正则化

方法5）。 L_2 仅正则化

比较基于论文中介绍的示例-I。我们重复进行“数据模拟—统计推断”过程 100 次，并计算估计的初始条件 $\xi(0,\ s)$ 的偏差和均方误差。

下一页的图（13）首先展示了采用不同正则化方法得到的估计初始条件。具体而言，该图第一行展示了过程在时刻 2、5、10、15 和 20 的快照。该图第 2 至第 6 行展示了利用截至时刻 2、5、10、15 和 20 的传感器观测得到的估计初始条件等高线图。粗蓝色水平集分别表示逆模型生成的输出的第 75、85 和 95 百分位数，而粗虚线水平集表示第 50 和第 60 百分位数。

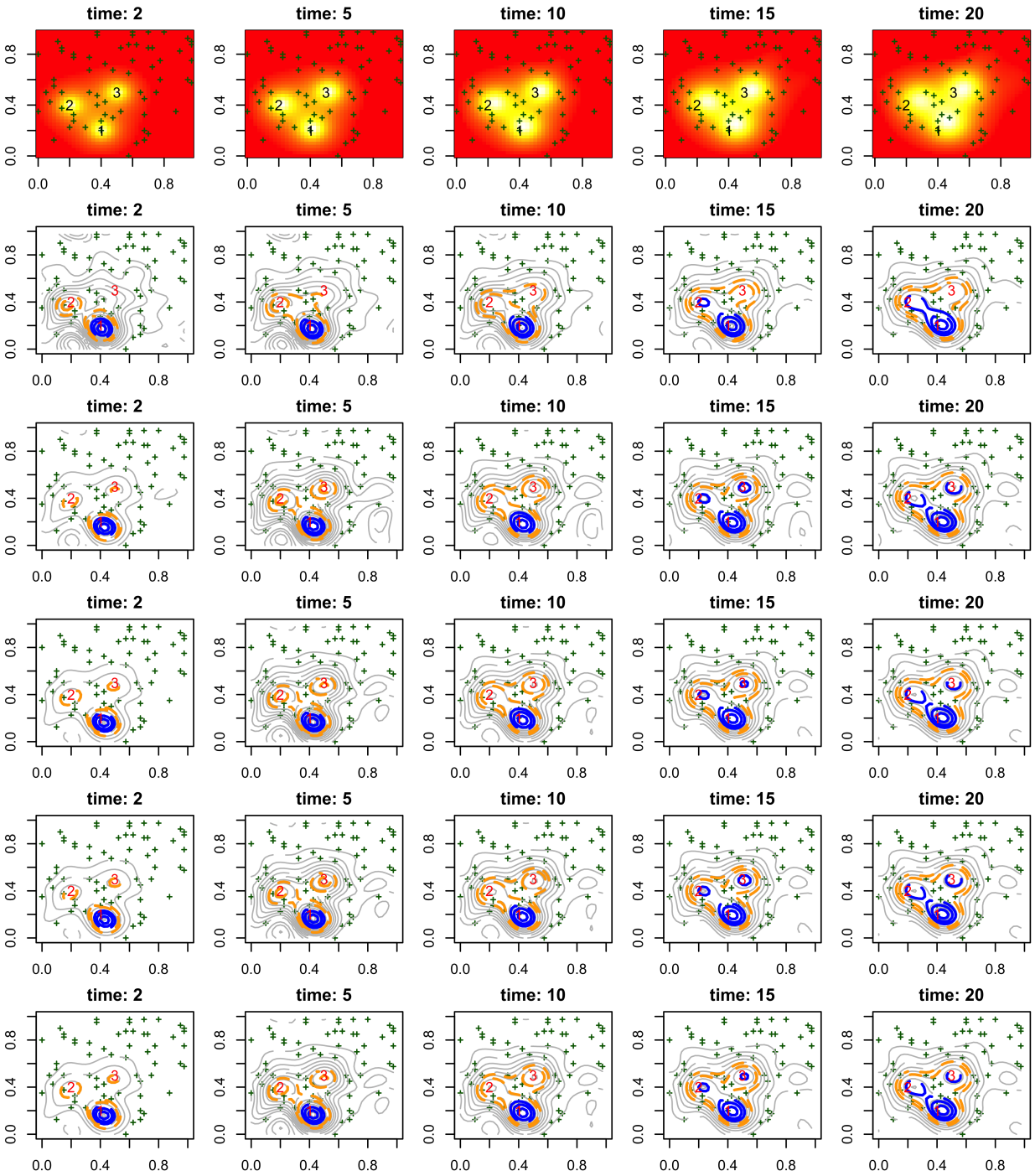

图13：（第1行）对流扩散过程在时间2、5、10、15和20时的快照；（第2至第6行）分别使用所提出的正则化（第2行）、广义Lasso（第3行）、弹性网络（第4行）、 L_1正则化（第5行）和 L_2 正则化（第6行）得到的时间2、5、10、15和20时的估计的初始条件。

接下来，图14 展示了所有正则化方案下估计的初始条件的偏差和均方误差。

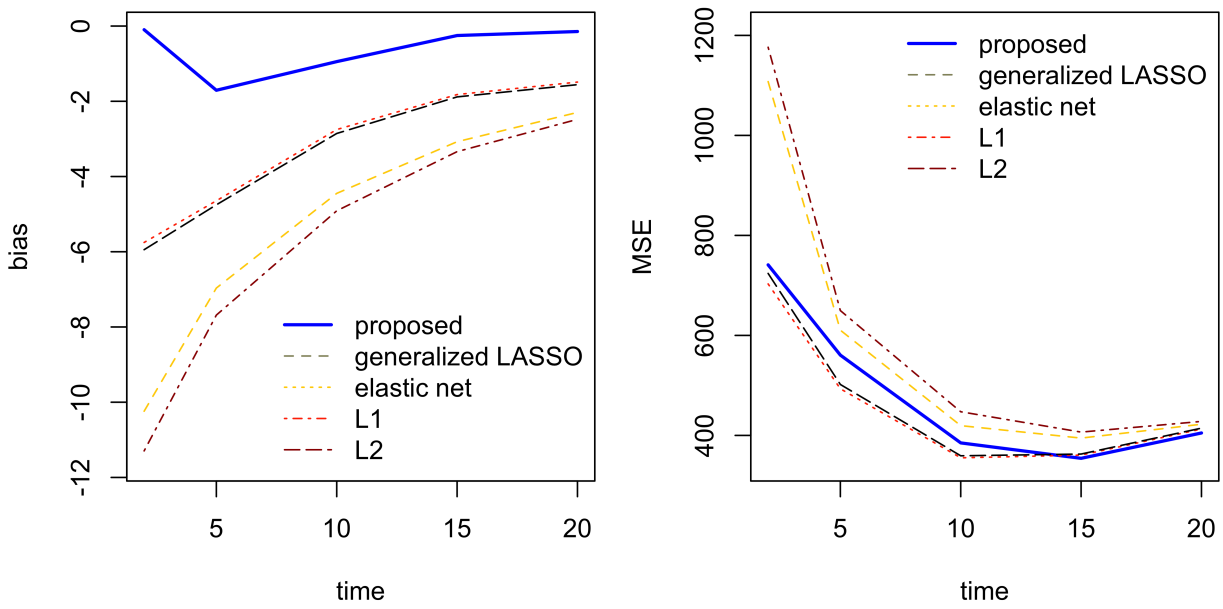

图14：基于不同正则化方案得到的估计初始条件的偏差和均方误差（MSE）。

可以看出：

◇ 所提出的正则化产生的偏差最低。 L_2 正则化和弹性网络产生的偏差最高，而 L_1 正则化和广义Lasso的表现相近。

◇ 所提出的正则化产生的均方误差略高于 L_1 正则化和广义Lasso，而 L_2 正则化和弹性网络产生的均方误差更高。

◇ 在我们的研究过程中，我们花费了大量时间尝试不同的正则化方式及其组合。我们发现的一个难题是， $\boldsymbol{\eta}$ 中的估计谱系数可能会随频率变化而大幅波动（例如，可能会看到对应较高频率的系数突然出现尖峰）。然而，由于本文关注的是平滑初始条件的估计，理想情况下，谱系数应随着频率的增加而逐渐衰减

（不一定是单调的）。单独添加 L_1 或 L_2 正则化并不能自动克服这一难题，尽管它能促进稀疏性。因此，我们对 Fussed Lass 进行了修改，并添加了第二项正则化，以帮助约束“相邻频率对应分量之间的平滑性”（防止相邻谱系数出现大幅波动）。因此，我们可能会看到估计的特殊系数逐渐衰减，从而产生平滑的初始条件。