

基于动态约束插值的PM10浓度空间分布模拟

李佳新

大连交通大学基础部理学院, 辽宁 大连

收稿日期: 2025年12月13日; 录用日期: 2026年1月6日; 发布日期: 2026年1月14日

摘要

PM10是一类极具代表性的大气污染物,其表面易吸附多环芳烃、重金属等有毒有害物质,进而对人体健康与生态环境构成多重威胁,相关研究因此受到广泛重视。对PM10污染的研究不仅能深入剖析其传输扩散规律,更为后续污染防控工作提供坚实的科学支撑。在PM10相关研究中,精准模拟其浓度空间分布是核心环节,对全面掌握PM10污染特征具有不可替代的作用。地面监测、数值模拟与空间插值是当前研究PM10浓度空间分布的主要技术手段,但数值模型对数据完整性要求较高,传统插值方法在观测稀疏区域难以达到理想精度。动态约束插值法(DCIM)已被证实可有效提升时空稀疏观测数据的利用率,为此,本研究提出一种基于伴随模型的动态约束插值方法,通过算法优化增强PM10浓度的空间重建能力。利用2015年3月2日至3月5日中国中南部85个地面监测站的观测数据,通过分步实验设计开展系统验证:首先验证DCIM在PM10数值模拟中的有效性,实验结果显示该方法显著提升了模拟精度、优化了模拟效果,充分证明其在该领域的适用性与可靠性;然后针对插值方法在稀疏数据场中的局限性,引入高阶守恒插值(PPM)算法,实验表明该算法在适配稀疏数据场场景方面具有显著优势。

关键词

PM10, 动态约束插值(DCIM), 数值模拟, 高阶守恒插值(PPM)

Spatial Distribution Simulation of PM10 Concentrations Based on Dynamic Constraint Interpolation

Jiaxin Li

College of Science, Basic Department, Dalian Jiaotong University, Dalian Liaoning

Received: December 13, 2025; accepted: January 6, 2026; published: January 14, 2026

Abstract

PM10 is a highly representative atmospheric pollutant whose surface readily adsorbs toxic substances

文章引用: 李佳新. 基于动态约束插值的 PM10 浓度空间分布模拟[J]. 应用数学进展, 2026, 15(1): 252-264.
DOI: 10.12677/aam.2026.151026

such as polycyclic aromatic hydrocarbons and heavy metals, posing multiple threats to human health and the ecological environment. Consequently, research in this area has garnered significant attention. Studies on PM₁₀ pollution not only provide in-depth insights into its transport and dispersion patterns but also offer robust scientific support for subsequent pollution prevention and control efforts. Accurate simulation of PM₁₀'s spatial concentration distribution is central to such research, playing an irreplaceable role in comprehensively understanding its pollution characteristics. Ground-based monitoring, numerical modeling, and spatial interpolation are the primary techniques currently used to study PM₁₀'s spatial distribution. However, numerical models demand high data completeness, while traditional interpolation methods struggle to achieve ideal accuracy in sparsely observed regions. Dynamic Constrained Interpolation Method (DCIM) has been proven effective in enhancing the utilization of spatiotemporally sparse observational data. Therefore, this study proposes a dynamic constrained interpolation method based on an adjoint model, which enhances the spatial reconstruction capability of PM₁₀ concentrations through algorithm optimization. Using observational data from 85 ground monitoring stations in central and southern China between March 2 and March 5, 2015, systematic validation was conducted through a stepwise experimental design: First, DCIM's effectiveness in PM₁₀ numerical simulation was validated. Results demonstrated significant improvements in simulation accuracy and optimized outcomes, fully confirming its applicability and reliability in this domain. Subsequently, addressing limitations of traditional interpolation methods in sparse data fields, the higher-order conservative interpolation (PPM) algorithm was introduced. Experiments revealed that this algorithm possesses distinct advantages in adapting to sparse data scenarios.

Keywords

PM₁₀, Dynamic Constrained Interpolation (DCIM), Numerical Simulation, High-Order Conservative Interpolation (PPM)

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

大气颗粒物 PM₁₀ 因其复杂的物理化学特性及对公众健康的显著影响, 长期以来一直是全球环境与气候研究的重点方向。PM₁₀ 可通过呼吸道深入肺泡, 诱发哮喘、慢性阻塞性肺病等呼吸系统疾病[1], 其表面吸附的多环芳烃(PAHs)和重金属等有毒物质, 还可通过氧化应激机制加剧心血管损伤与 DNA 突变风险[2]。流行病学调研数据显示, PM₁₀ 浓度升高时呼吸系统疾病相关死亡率也升高[3]。在生态系统层面, PM₁₀ 的长期沉降会改变土壤 pH 值与微生物群落结构, 同时遮蔽阳光抑制植物光合作用, 加速区域生态系统的脆弱化进程。

在区域环境与社会经济层面, 中国中南部作为国家核心经济增长区, 其主要工业城市与人口密集区域面临严重的 PM₁₀ 污染问题, 长期浓度超出世界卫生组织标准 2~3 倍甚至更高[4]。鉴于 PM₁₀ 对人体健康与生态环境的多重危害, 开展深入研究极具必要性。自 2013 年起, 中国环境保护部构建了全国空气质量监测网络, 该网络每小时发布 PM₁₀ 等空气污染物浓度数据, 并启动专项科研项目鼓励相关研究探索。通过分析不同区域 PM₁₀ 浓度分布特征, 可为空气污染状况的模拟预测及防控政策制定提供科学依据。

当前, 地面监测站能够提供高精度、高时间分辨率的观测数据, 已广泛应用于城市空气质量实时警报与健康暴露评估, 例如中国国家环境监测中心(CNEMC)已通过全国 1400 余个站点构建覆盖 339 个地级市的监测网络, 为《大气污染防治行动计划》的实施提供了关键数据支撑[5]。数值模型在 PM₁₀ 浓度

空间分布研究中发挥着关键作用,但这类模型对数据完整性的高度依赖限制了其应用场景[6]。传统空间插值技术如 Cressman 插值、Kriging 插值等虽在站点数据空间化中应用广泛,但在观测数据稀缺区域易出现过度平滑的问题,难以精准反映污染物浓度的真实空间分布特征。

动态约束插值(DCIM)通过观测数据与数值模型,为上述问题提供了创新解决方案,显著提升了稀疏数据的利用率。在海洋学领域,Zheng 等人利用 DCIM 方法重建渤海湾潮汐场,相比传统数学插值方法,均方根误差(RMSE)实现显著降低[7];在大气科学领域,Li 等人通过该方法将全国 PM_{2.5} 监测数据同化的空间分辨率提升至 5 千米,展现出其在大规模污染物分布重建中的潜力[8]。空间插值方法的准确性与精度一直是研究关注的重点,Cressman 插值虽具备简单高效的优势,但在处理远离观测站的格点时易产生较大误差,影响半径的不合理设定会进一步削弱插值效果。相比之下,高阶守恒插值(PPM)能够保证局部质量守恒,具备更高精度,适用于大气学、流体力学等领域。Li 等人(2017)应用 PPM 有效弥补了 Cressman 插值在大网格间距下的精度缺陷,其能够通过少量独立点构建更贴合实际的初始场,即便在大网格间距条件下仍能保持较高精度,为优化动态约束插值方法的模拟效果提供了新思路,对 PM₁₀ 浓度空间重建研究具有重要参考价值[9]。因此,本研究旨在结合地面监测数据与高阶守恒插值(PPM)优化算法,提升中国中南部 PM₁₀ 浓度空间重建的精度。

本文后续内容安排如下:第 2 节详细阐述所用材料与方法;第 3 节呈现实际实验与分析结果;第 4 节总结讨论并提出结论。

2. 材料与方法

2.1. 研究区域与数据

近年来,中国持续推进基础监测网络的科学规划与部署,截至 2022 年,已建成超过 1400 个地级监测站,覆盖 339 个地级市,形成了较为广泛的监测网络。本研究聚焦中国中南部地区,该区域人口密集、经济发展迅速,在全国城市体系中占据关键地位。研究选取该地区 85 个地面监测站,通过网络爬虫技术获取 2015 年主要城市的 PM₁₀ 观测数据,数据覆盖 2015 年 3 月 2 日 0 点至 3 月 5 日 0 点的 72 小时观测期。将每小时浓度值取平均得到日平均浓度,对于拥有多个监测站的城市,取所有可用测量值的平均值。85 个监测站的空间分布如图 1 所示。

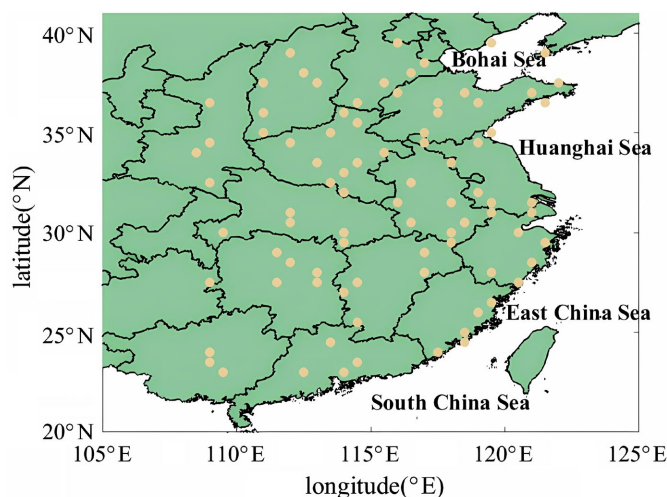


Figure 1. Spatial distribution of 85 basic monitoring stations in the central and southern regions

图 1. 中南部地区 85 个基础监测站的空间分布

2.2. 动态约束插值方法

伴随模型通过求解伴随方程计算目标函数梯度, 结合梯度下降法等优化算法逐步逼近最优解, 该方法广泛应用于海洋与大气研究中的参数优化及动态约束插值。例如, Wu 等人(2021)利用该方法优化 Ekman 模型参数[10], Xu 等人(2017)通过其高效模拟 2014 年 APEC 期间中国区域 PM_{2.5} 浓度分布[11], Li 等人(2021)成功模拟全国 PM_{2.5} 浓度分布[8]。这些研究证实, 该方法能有效提升时空稀疏数据的利用率与模拟精度, 尤其适用于大气污染物研究, 在全国 PM₁₀ 观测数据重建中具有显著优势。

2.2.1. 正向模型

PM₁₀ 浓度是空气质量评估的核心指标, 其空间分布特征与变化趋势的精准刻画可为空气质量管理提供重要依据。大气中的 PM₁₀ 受到传输、扩散、光化学转化(产生二次 PM₁₀)及湿干沉降等多重耦合过程影响, 其传输主要由风向与湍流扩散驱动。本研究采用二维对流扩散方程描述其动力学机制, 将一次与二次 PM₁₀ 合并为“源汇项”(S), 同时忽略垂直扩散与复杂化学过程, 建立简化质量守恒方程:

$$\frac{\partial C}{\partial t} + u \frac{\partial C}{\partial x} + v \frac{\partial C}{\partial y} = \frac{\partial}{\partial x} \left(A_H \frac{\partial C}{\partial x} \right) + \frac{\partial}{\partial y} \left(A_H \frac{\partial C}{\partial y} \right) + S$$

其中, C 表示 PM₁₀ 浓度, t 表示时间, x 和 y 为水平坐标, u 和 v 分别为 x 和 y 坐标处的水平风速, A_H 为水平粘度系数, S 表示源与汇。

$$C(t^0, x, y) = C^0(x, y)$$

$$\frac{\partial C}{\partial t} = 0 \quad \text{on } \Gamma^{\text{IN}}$$

$$\frac{\partial C}{\partial \vec{n}} = 0 \quad \text{on } \Gamma^{\text{OUT}}$$

在流入边界处施加常数边界条件, 而在流出边界处施加零梯度边界条件。该模型的数值方案与文献[12]中描述的方案一致。

2.2.2. 伴随模型

伴随模型是解决参数优化的强大工具。通过计算目标函数对参数的梯度, 并结合优化算法(如梯度下降法), 它们能逐步逼近最优解。在大气科学领域, 伴随模型常用于参数估计。该过程通过优化正向模型参数, 使模拟结果与观测数据的偏差最小化, 从而获得更精确的时空分布。

基于此, 我们构建了伴随模型并定义了成本函数, 以量化观测与模拟之间的差异:

$$J = \frac{1}{2} \int_{\Sigma} K (C - C_{\text{OBS}})^2 d\sigma$$

拉格朗日函数:

$$L = J + \left\{ p \times \left[\frac{\partial C}{\partial t} + u \frac{\partial C}{\partial x} + v \frac{\partial C}{\partial y} - \frac{\partial}{\partial x} \left(A_H \frac{\partial C}{\partial x} \right) - \frac{\partial}{\partial y} \left(A_H \frac{\partial C}{\partial y} \right) - S \right] \right\} d\sigma$$

其中, C 和 C_{OBS} 分别表示模拟与观测的 PM₁₀ 浓度; p 是 C 的伴随变量; K 是权重矩阵, 理论上为观测误差协方差矩阵的逆矩阵。假设误差不相关且权重相等, 则可简化为 1 (含观测值) 或 0 (无观测值)。为了最小化成本函数, 如下所示:

$$\frac{\partial L}{\partial p} = 0$$

$$\frac{\partial L}{\partial C} = 0$$

$$\frac{\partial L}{\partial A_H} = 0$$

$$\frac{\partial L}{\partial S} = 0$$

2.2.3. 独立点方案

在空间上选取部分网格点作为“独立点”。其他网格点的 PM10 浓度通过独立点浓度值进行线性插值计算。该方法既减少了需调整的网格点数量，又提升了反演结果的平滑度。设

$\{C_{ii,jj}, ii=1,2,\dots, in; jj=1,2,\dots, jn\}$ 为独立点的 PM10 浓度， $in \times jn$ 为独立点总数，则网格点 (i,j) 的 PM10 浓度值为：

$$C_{i,j} = \sum_{ii,jj} (\phi_{i,j,ii,jj} \times C_{ii,jj})$$

$\phi_{i,j,ii,jj}$ 是线性插值系数：

$$\phi_{i,j,ii,jj} = W_{i,j,ii,jj} / \sum_{ii,jj} W_{i,j,ii,jj}$$

$W_{i,j,ii,jj}$ 是 Cressman 插值法的权重系数：

$$W_{i,j,ii,jj} = (R^2 - r_{i,j,ii,jj}^2) / (R^2 + r_{i,j,ii,jj}^2)$$

$r_{i,j,ii,jj}$ 是一个独立点。该点 (ii,jj) 到网格点 (i,j) 的距离为 d ，影响半径为 R 。独立点的梯度公式可直接推导。在初始场反演中，我们采用独立点方案。空间上，每四个网格点选取一个独立点，影响半径设定为 4° 。独立点的初始 PM10 浓度设定为 $50 \mu\text{g}/\text{m}^3$ 。

2.2.4. 插值方法

1. Cressman 插值

模型运行过程中插值方法的选择对模拟结果影响显著，直接关系到空间数据处理能力和模拟结果的准确性。不同插值方法在原理、算法及适用性方面存在差异。例如：Cressman 插值法基于距离与浓度变化采用加权分布法估算未知点浓度；Kriging 插值法运用空间统计原理，考虑空间自相关性与变异性以生成更精确连续的浓度分布；正交多项式拟合通过构建多项式函数拟合数据点，适用于分布均匀且趋势明显的的数据，可生成平滑的浓度分布曲线。

这些插值方法可归纳为以下公式：

$$\hat{Z}(x_0) = \sum_{i=1}^n \lambda_i z(x_i)$$

Cressman 插值法是一种距离加权插值方法。因其模拟精度高、操作简便且适用于网格化数据，常被应用于气象研究。将 Cressman 插值法应用于观测数据处理，其定义为：

$$C(x,y) = \frac{\sum_{i=1}^n W_i z_i}{\sum_{i=1}^n W_i}$$

W_i 是第 i 个邻近观测点的权重，通常根据 x_0 与 x_i 的距离使用以下公式计算：

$$W_i = \frac{R - d_i}{R}$$

此处, R 表示影响半径, d_i 表示从 x_0 到 x_i 的距离。

2. 高阶守恒插值法(PPM)

PPM 方法应用于动态约束插值, 二维计算域划分为规则欧拉单元, 单元边界与中心定义如下:

$$\Omega_{i,j} = [x_{i-1/2}, x_{i+1/2}] \times [y_{j-1/2}, y_{j+1/2}]$$

$$x_i = \frac{1}{2}(x_{i-1/2} + x_{i+1/2})$$

$$y_i = \frac{1}{2}(y_{j-1/2} + y_{j+1/2})$$

设 (x_i, y_i) 表示一个独立点, $C_{i,j}$ 表示该独立点处的参数值。对于二维空间, 计算采用分步策略: 先沿 x 方向进行, 再沿 y 方向进行。本节中, 鉴于 x 方向计算原理与 y 方向类似, 我们将重点阐述 x 方向保守插值过程以突出关键点并避免冗余。通过这一系列计算步骤, 最终获得 $t = t_n$ 时间处 $\Omega_{i,j}$ 的值。

本应用中 PPM $[\mathcal{R}C^n]_i(x)$ 的具体应用如下:

$$[\mathcal{R}C^n]_i(x) = C_{i-1/2}^n + \frac{x - x_{i-1/2}}{h} \left(\Delta C_i^n + C_{6,i}^n \frac{x_{i+1/2} - x}{h} \right)$$

$$\Delta C_i^n = \langle C^n \rangle_{i+1/2} - \langle C^n \rangle_{i-1/2}$$

$$C_{6,i}^n = 6 \left(C_i^n - \frac{\langle C^n \rangle_{i+1/2} - \langle C^n \rangle_{i-1/2}}{2} \right)$$

插值分布 $[\mathcal{R}C^n]_i(x)$ 定义为, 其中 $\Omega_i = (x_{i-1/2}, x_{i+1/2})$, $\langle C^n \rangle_{i+1/2}$ 表示 $x_{i+1/2}$ 点处的数值。为获得 PPM 的二阶近似表达式, 给出 $\langle C^n \rangle_{i+1/2}$ 如下公式:

$$\langle C^n \rangle_{i+1/2} = \frac{C_i^n + C_{i+1}^n}{2}$$

然后, 我们可以使用以下方程获得点上所有值:

$$\begin{aligned} [\mathcal{R}C^n]_i(x) = & \left[\frac{1}{2} - \frac{x - x_{i-1/2}}{2h} - \frac{3(x - x_{i-1/2})(x_{i+1/2} - x)}{2h^2} \right] C_{i-1}^n \\ & + \left[\frac{1}{2} + \frac{3(x - x_{i-1/2})(x_{i+1/2} - x)}{h^2} \right] C_i^n \\ & + \left[\frac{x - x_{i-1/2}}{2h} - \frac{3(x - x_{i-1/2})(x_{i+1/2} - x)}{2h^2} \right] C_{i+1}^n \end{aligned}$$

2.3. 模型设置

计算域覆盖东经 105° 至 125° 、北纬 20° 至 40° , 水平分辨率是 $0.5^\circ \times 0.5^\circ$, 网格为 x 方向 43 个单元、 y 方向 41 个单元。时间步长是 3600 秒。模拟周期自 2015 年 3 月 2 日 00:00 起至 2015 年 3 月 5 日 00:00 止, 共计 72 小时。背景风数据覆盖 2015 年 3 月 2 日至 5 日, 空间分辨率为 $0.5^\circ \times 0.5^\circ$, 时间分辨率是 1 小时。迭代步数是 100 次。模型在流入边界采用常数边界条件, 在流出边界采用无梯度边界条件。依据 2013 年环境保护部颁布的中国新环境空气质量标准, 入流边界 PM10 浓度限值设定为 $35.0 \mu\text{g}/\text{m}^3$ 。水平

扩散系数取值 $100.0 \text{ m}^2/\text{s}$, 背景值设定为 $50 \text{ }\mu\text{g}/\text{m}^3$ 。为提高模拟精度, 每四个网格选取一个独立网格进行优化, 其余网格的 PM10 浓度通过 Cressman 法计算。初始场通过对初始时刻观测 PM10 数据进行 Cressman 插值获得, 以确保动态模型模拟结果的准确性。

3. 实际实验与分析

3.1. 实验设计

为开展动态约束插值性能的计算研究, 我们实施了两组实际实验。第一组实验验证动态约束插值法的有效性, 为进一步探索该方法与不同插值技术结合应用于 PM10 数值模拟研究时的效果, 我们设计了第二组实验, 旨在验证优化算法的动态约束插值法在 PM10 数值模拟中的优势。所有实验均采用交叉验证设计以确保结果可靠性, 因交叉验证是评估插值结果的有效方法[13] [14]。

本实验采用四折交叉验证方案评估动态约束插值法的有效性。具体而言, 随机抽取四分之一观测数据排除在同化处理之外, 该部分数据作为独立观测数据用于验证同化结果, 称为“检验数据”。其余四分之三观测数据作为模型输入数据, 称为“同化数据”。

采用平均绝对误差(MAE)和均方根误差(RMSE)作为误差指标。MAE 能直观反映模拟值与观测值的偏差幅度, 是评估模拟精度的常用指标, 但仅考虑误差绝对值而忽略正负方向, 可能在正负误差相互抵消时无法准确反映真实误差; RMSE 通过对误差进行平方处理, 全面考量正负误差, 且对较大偏差更敏感, 能更综合地评估模拟精度。定义如下:

$$\text{MAE} = \sum_{i=1}^N |M_i - O_i| / N$$

$$\text{RMSE} = \left[\frac{1}{N} \sum_{i=1}^N (M_i - O_i)^2 \right]^{\frac{1}{2}}$$

此处, N 表示检验数据点的总数 ($i=1, 2, \dots, N$), 而 M_i 和 O_i 分别代表 PM10 浓度的模拟值与观测值。

3.2. 第一组结果与讨论

本研究采用基于 Cressman 插值的动态约束插值方法, 模拟 2015 年 3 月 2 日 0:00 至 3 月 5 日 0:00 的 PM10 浓度数据, 通过比较模型应用前后成本函数和误差的变化(实验 PE11)评估方法性能; 设计实验 PE12 和 PE13, PE12 计算监测站日均浓度后采用 Cressman 方法直接插值, 得到日均 PM10 浓度空间分布; PE13 以地面监测站实际观测数据为模型输入, 通过伴随模型优化初始场和源汇项等关键参数运行模型得到 PM10 浓度空间分布。通过对比两种方法的浓度空间分布与 85 个地面监测站的日均浓度分布, 评估动态约束插值方法的模拟准确性。

2015 年 3 月 2 日至 3 月 5 日的四折交叉验证实验结果如表 1 所示, 四组实验的约化成本函数下降趋势如图 2 所示。表 1 显示, 同化后验证点观测值与模拟值的 MAE 从 $44.53 \text{ }\mu\text{g}/\text{m}^3$ 降至 $34.34 \text{ }\mu\text{g}/\text{m}^3$, 降幅达 22.89%; RMSE 显著降低 30.41%, 表明动态约束插值方法有效减小了模拟结果与实际观测值的误差, 使模拟值更接近真实 PM10 浓度。图 2 显示成本函数随迭代过程持续下降, 进一步证实该方法在提高模拟精度方面的有效性。

图 3 展示了 85 个地面监测站的 3 天平均观测结果, 图 4 展示了传统 Cressman 插值法和基于 Cressman 的动态约束插值法(DCIM)的 PM10 浓度分布图。对比可见, 两种方法的平均浓度分布图与平均观测图高度相似, 均能有效模拟 PM10 浓度的空间分布特征。其中, 传统 Cressman 插值法作为经典方法已得到广泛应用[15], 而 DCIM 方法产生的分布特征更平滑、空间连续性更优, 这得益于 2015 年 3 月中国中南部

处于春季过渡阶段，受东亚冬季风残余影响，区域盛行偏北风，同时气温回升导致边界层高度抬升，湍流扩散能力增强，这些气象条件直接影响 PM10 的传输与扩散过程。从物理机制来看，动态约束插值法 (DCIM)通过耦合二维对流扩散方程与伴随模型，将水平风速(u 、 v)、水平扩散系数(A_H)等参数纳入优化过程，能够动态响应区域气象场的空间差异，例如在风速较大的边界区域，DCIM 通过梯度优化捕捉到 PM10 的远距离传输特征，而在湍流扩散较弱的盆地或河谷区域，则通过源汇项调整还原浓度累积效应，从而更精准地捕捉 PM10 在复杂气象条件下的传输扩散规律。此外，DCIM 结果与 2015 年春季中国环境空气中 PM10 浓度的空间分布特征具有一致性，进一步验证了 DCIM 在空间分布模拟中的有效性。

需要注意的是，本研究的监测站点在空间分布上存在明显稀疏性，尤其是陕西、山西和内蒙古边境地区及内蒙古东部，85 个监测站在该区域的平均间距超过 50 公里，远大于中南部平原城市群的站点密度。这种稀疏性导致观测数据无法充分反映区域内 PM10 的微尺度空间变异，进而影响插值方法的优化效果：传统 Cressman 插值法模拟的 PM10 平均浓度约为 $106.017\text{ }\mu\text{g}/\text{m}^3$ ，虽低于观测浓度(约 $118.607\text{ }\mu\text{g}/\text{m}^3$)但数值更接近观测值；而 DCIM 方法模拟的平均浓度约为 $146.982\text{ }\mu\text{g}/\text{m}^3$ ，显著偏高，推测该区域 DCIM 方法存在过拟合现象，其核心原因是稀疏站点无法提供足够的约束信息，导致模型对局部观测值的敏感度异常升高，进而放大了局部浓度特征。

Table 1. MAE and RMSE of observed versus simulated values at assimilation and verification points (unit: $\mu\text{g}/\text{m}^3$)
表 1. 同化点与检验点观测值与模拟值的 MAE 和 RMSE (单位: $\mu\text{g}/\text{m}^3$)

	同化点		检验点	
	Before	After	Before	After
MAE	44.74	24.55	44.52	34.34
RMSE	619.99	39.47	80.81	56.23

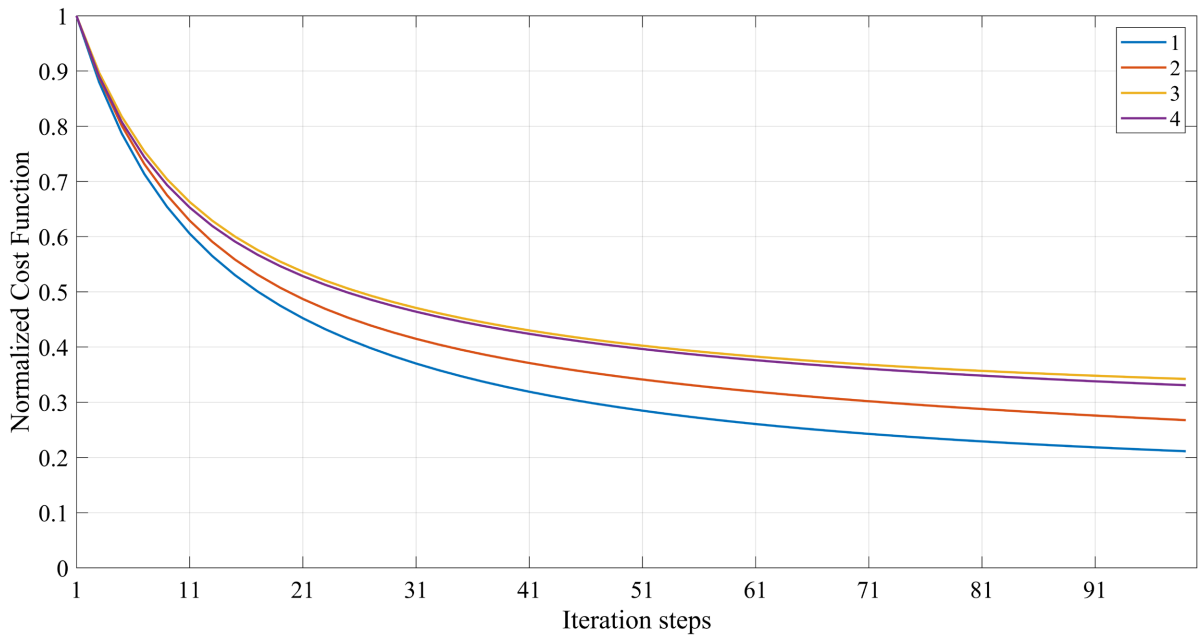


Figure 2. Changes in the cost function across iterations for the four experimental groups in the PE11 quadruple cross experiment
图 2. PE11 四折交叉实验中四个实验组迭代过程中代价函数的变化

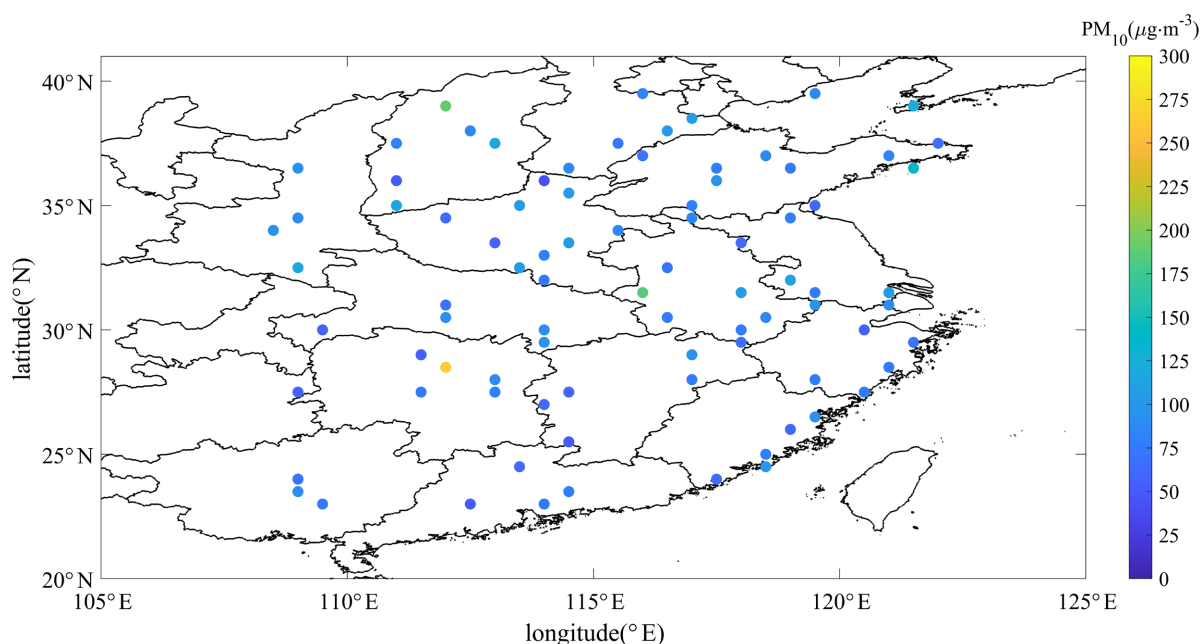


Figure 3. Daily average PM10 concentrations at 85 monitoring stations in central and southern China (00:00 on March 2, 2015 to 00:00 on March 5, 2015). The intensity of each point's color indicates its daily average PM10 concentration level

图 3. 中国中部和南部 85 个监测站的日均 PM10 浓度(2015 年 3 月 2 日 00:00 至 2015 年 3 月 5 日 00:00)。每个点的颜色深浅表示其日均 PM10 浓度的大小

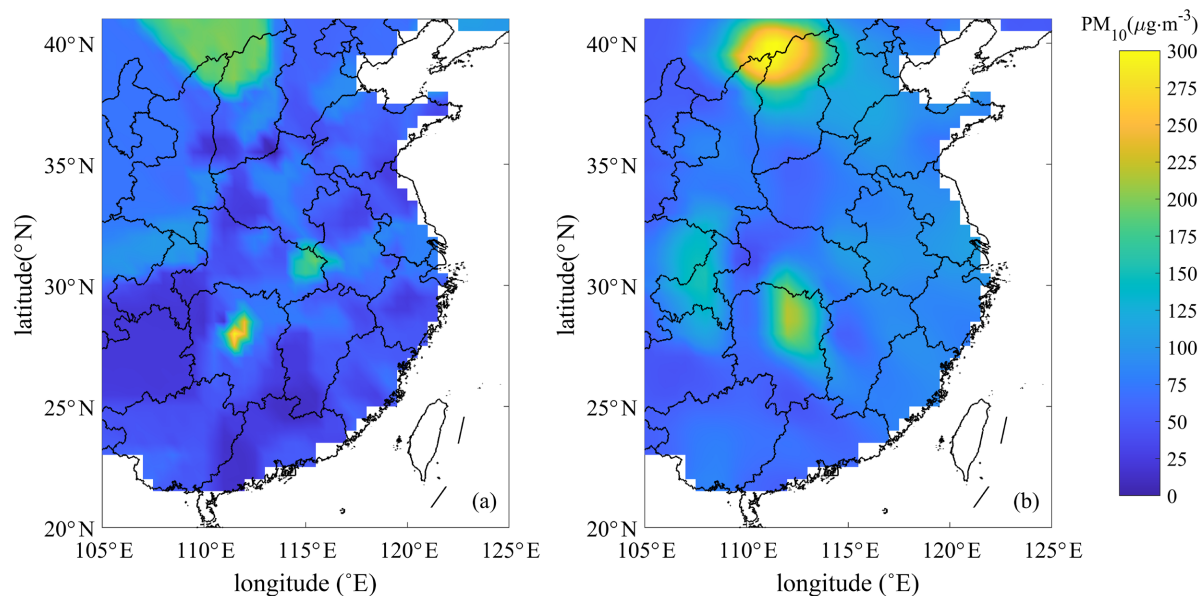


Figure 4. Average fields obtained from experiments PE12 and PE13 ((a): PE12, (b): PE13)

图 4. 实验 PE12 和 PE13 所得的平均场((a): PE12, (b): PE13)

3.3. 第二组结果与讨论

本研究采用基于 Cressman 插值和高阶守恒插值(PPM)的两种动态约束插值方法, 模拟分析 2015 年 3 月 2 日 0:00 至 3 月 5 日 0:00 的 PM10 浓度数据, 通过四折交叉验证实验评估结果并重建 PM10 浓度空间分布。

表 2 展示了两种方法在同化前后, 同化点和检验点观测值与模拟值之间的 MAE 和 RMSE 变化, 图 5 展示了实验 PE21 和 PE22 的误差折线图。图表显示, 与基于 Cressman 的插值方法相比, 基于 PPM 的动态约束插值方法在同化数据中产生的误差更小, 表明其对同化数据实现了更优的优化。表 2 显示, 基于 PPM 插值的动态约束插值方法在同化点的性能提升显著: MAE 降低了 59.6%, RMSE 降低了 95.36%; 而基于 Cressman 插值的方法同化点 MAE 降低了 45.1%, RMSE 降低了 93.63%。尽管 PPM 方法在检验点误差略高于 Cressman 方法, 但其同化点数据误差更小, 综合来看在兼顾同化点与检验点模拟精度方面更具优势。

Table 2. Comparison of MAE and RMSE between the two methods (unit: $\mu\text{g}/\text{m}^3$)
表 2. 两种方法的 MAE 与 RMSE 对比(单位: $\mu\text{g}/\text{m}^3$)

	同化点 MAE	检验点 MAE	同化点 RMSE	检验点 RMSE
	Before	After	Before	After
Cressman	44.74	24.55	44.53	34.34
RMSE	44.74	17.66	44.53	41.43

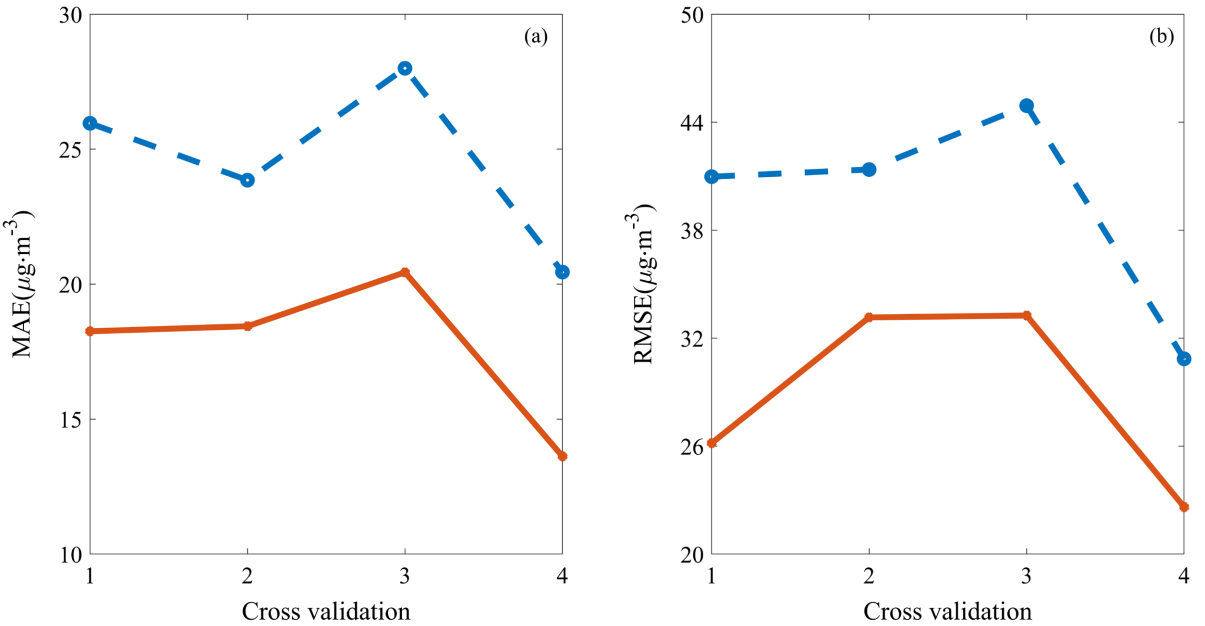


Figure 5. Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) of data assimilated using corresponding methods in the four-fold cross-validation experiment (blue dashed line: dynamic constrained interpolation based on Cressman interpolation; red solid line: dynamic constrained interpolation based on PPM interpolation)
图 5. 四折交叉实验中, 采用相应方法同化的数据的平均绝对误差(MAE)和均方根误差(RMSE) (蓝色虚线: 基于 Cressman 插值的动态约束插值; 红色实线: 基于 PPM 插值的动态约束插值)

图 6 展示了实验 PE21 和 PE22 所得平均浓度的空间分布: PE21 为基于 Cressman 插值的结果, PE22 为基于 PPM 插值的结果。对比可见, 两种方法的平均浓度分布与观测分布高度相似, 且均表现出平滑的分布模式, 整体一致性良好。值得注意的是, PE22 在陕西、山西和内蒙古边境地区及内蒙古东部的过拟合现象得到明显缓解, 其模拟的 PM10 平均浓度约为 $134.196 \mu\text{g}/\text{m}^3$, 相较于 PE21 的 $146.982 \mu\text{g}/\text{m}^3$, 更

接近观测值($118.607 \mu\text{g}/\text{m}^3$)。这一改善的核心原因是：PPM 的高阶守恒特性降低了模型对稀疏站点的敏感度，即使在站点间距较大的区域，也能通过多项式重构维持浓度分布的物理合理性，避免了 Cressman 插值和传统 DCIM 方法不足导致的极值偏离。从算法机理来看，PPM 方法之所以能缓解稀疏数据区域的过拟合问题，核心在于其高阶守恒特性与分段多项式重构逻辑：与 Cressman 插值基于距离加权的局部拟合不同，PPM 通过在每个网格单元内构建二次多项式分布，同时满足单元间的质量守恒和梯度连续性，能够在有限观测点的约束下，合理推断未观测区域的浓度变化趋势，避免对单一观测值的过度依赖。具体而言，PPM 通过计算单元边界的通量守恒(公式中 $\langle C^n \rangle_{i+1/2}$ 项)和内部梯度($C_{6,i}^n$ 项)修正，有效抑制了稀疏数据场景下的“局部极值放大”现象，这与 2015 年 3 月研究区域的气象背景高度契合：该时段区域扩散条件整体均匀，PM10 浓度空间梯度以平缓变化为主，PPM 的高阶平滑特性能够更好地匹配这种气象驱动下的浓度分布规律，而 Cressman 插值的距离加权逻辑则易受稀疏站点的局部异常值影响。

同时需要客观指出，本研究的站点分布稀疏性仍对 PPM 方法的性能产生了一定限制：在内蒙古东部部分无站点覆盖的区域，PE22 的模拟浓度仍存在约 15% 的偏差，这表明尽管 PPM 的算法机理更适配稀疏数据场景，但当站点间距超过其有效插值尺度时，仍难以完全消除数据稀缺带来的不确定性。

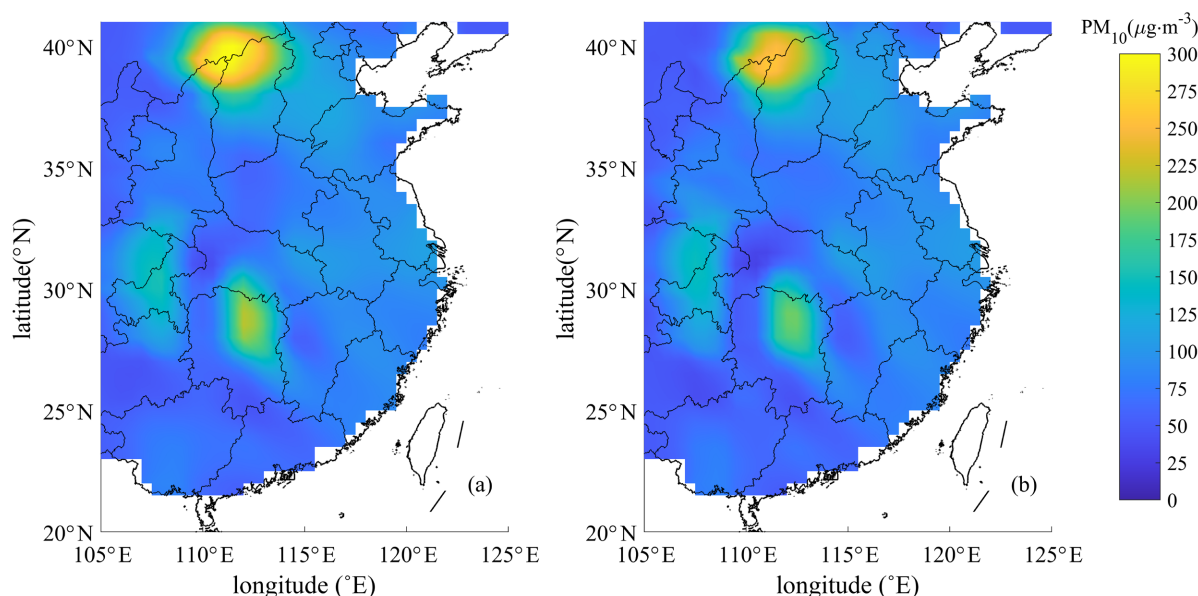


Figure 6. PM10 concentration distribution obtained from experiments PE21 and PE22 ((a): PE21; (b): PE22)

图 6. 实验 PE21 和 PE22 得到的 PM10 浓度分布((a): PE21; (b): PE22)

4. 结论与讨论

本研究提出的基于伴随模型的动态约束插值法(DCIM)，显著提高了中国中南部复杂地理区域 PM10 浓度的时空重建精度。该方法通过地面监测数据以及优化的高阶守恒插值(PPM)算法实现。借助伴随模型求解伴随方程以计算目标函数梯度，并结合迭代优化算法逼近最优解，该方法显著克服了传统插值法在观测稀疏区域的精度局限性。

从模拟精度和效果的角度出发，本研究将传统的 Cressman 插值法与基于 Cressman 的动态约束插值法进行了比较。结果表明，模拟值与观测值的对比显示，动态约束插值法有效降低了模拟误差，提高了精度。更重要的是，它生成了更平滑的空间分布，其核心优势在于通过物理模型约束减少了纯数学插值的主观性。这表明该方法能够有效地重建 PM10 的空间浓度分布，尽管在陕西、山西和内蒙古交界地区

以及内蒙古东部出现了过拟合现象。进一步分析表明,过拟合可能源于观测数据不足。当 Kriging 插值法应用于大网格间距的数据集时,其插值误差往往无法满足要求。观测数据不足导致反演问题的不适定性,这是一个需要解决的关键问题。

为解决观测数据不足导致的不适定性问题,本研究进一步引入了高阶守恒插值(PPM)算法。将该算法应用于独立点方案,有效缓解了反演问题的不适定性。将优化后的模型与未优化的模型进行比较,结果表明前者在陕西、山西和内蒙古的边界地区以及内蒙古东部部分缓解了过拟合现象,这得益于 PPM 的高阶守恒特性与分段多项式重构逻辑,能够在有限观测约束下维持浓度分布的物理合理性,避免对局部观测值的过度依赖。这一结果验证了 PPM 算法的独特优势,即在观测数据稀少的情况下能提供更优的插值质量和稳健的性能。因此,通过地面监测数据以及优化的高阶守恒插值(PPM)算法可作为模拟 PM10 分布的有效研究方法,并为制定中国中南部地区的空气污染控制政策提供参考。并且本研究开发的 DCIM 方法不仅为 PM10 模拟提供了一种新颖的技术范例,还为优化算法同化提供了一个可扩展的理论框架。

值得注意的是,当前研究区域的站点分布稀疏性对本研究结论的可靠性与普适性造成了不可忽视的影响:

从空间约束有效性来看,研究区域内 85 个监测站的空间分布不均衡,中南部平原城市群(如长江中下游、珠江三角洲)站点间距约为 20~30 公里,而内蒙古东部、晋陕蒙边境等区域的站点间距普遍超过 50 公里,部分区域甚至达到 80 公里以上,远超本研究模型的有效插值尺度(约 40 公里)。这种不均衡的稀疏分布导致观测数据无法完整捕捉 PM10 浓度的空间变异特征,尤其是在无站点覆盖的“数据空白区”,模型只能依赖数学假设与邻域站点的间接推断,使得插值结果的物理真实性难以保证。

对方法性能评估的干扰方面,站点稀疏性直接影响了交叉验证结果的客观性。在四折交叉验证中,检验数据与同化数据的空间重叠度较低,部分检验点周围缺乏足够的同化站点作为约束,导致 MAE、RMSE 等误差指标可能被低估或高估,例如晋陕蒙边境区域的检验点,因周边仅 1~2 个同化站点,其模拟误差无法真实反映方法的实际性能,进而可能导致对 DCIM 结合 PPM 方法优势的判断存在偏差。

对结论普适性的限制尤为明显:本研究得出“PPM 算法可有效缓解过拟合”的结论,主要基于春季短期数据与当前稀疏站点条件。但在站点更稀疏或稀疏模式不同的场景下(如西北干旱区、西南山区),PPM 算法的多项式重构逻辑可能因缺乏足够观测约束而失效,无法维持浓度分布的物理合理性;同时,站点稀疏导致本研究无法验证方法在不同污染梯度、不同地形条件下的适应性,使得“DCIM 结合 PPM 方法适用于中国中南部 PM10 浓度模拟”的结论仅能局限于当前站点覆盖范围内,难以推广至更广泛的区域。

本研究存在以下局限性,需在未来研究中进一步完善:

1. 模型简化带来的不确定性:本研究采用二维对流扩散方程,忽略了垂直扩散过程与复杂光化学转化机制,而实际 PM10 的传输扩散是三维耦合过程,且二次 PM10 的生成机制会显著影响浓度分布,模型简化可能导致高海拔或工业排放集中区域的模拟偏差;
2. 研究时段的局限性:实验数据仅涵盖 2015 年 3 月 2 日至 5 日的 72 小时,属于春季短期观测,未涵盖冬、夏、秋等不同季节的污染特征(如冬季供暖期 PM10 浓度峰值、夏季降水湿沉降影响),导致方法的季节普适性尚未得到充分验证;
3. 站点分布稀疏的影响:尽管 PPM 算法缓解了过拟合,但研究区域内部分区域(如内蒙古东部、晋陕蒙边境)的监测站点间距过大,仍导致插值结果存在一定偏差,稀疏站点无法提供足够的空间约束信息,是影响模拟精度的关键因素;
4. 数据来源的单一性:仅利用地面监测数据进行同化与验证,缺乏卫星遥感(如 MODIS AOD)、数值模型输出等多源数据的融合,可能导致观测误差被放大,且无法有效弥补地面站点的空间覆盖缺口。

为提升研究的鲁棒性与普适性,未来可从以下方面开展拓展研究:

1. 扩展研究时段至典型季节(如冬季 1 个月、夏季 1 个月),涵盖不同污染水平与气象条件(如沙尘、

暴雨、静稳天气), 系统验证 DCIM 结合 PPM 方法在不同场景下的适应性; 2. 优化模型结构, 引入三维对流扩散方程与二次 PM10 生成模块, 提升模型的物理完整性; 3. 融合多源数据(地面监测、卫星遥感、数值模型输出)构建多维数据集, 通过数据同化技术减少单一数据源的误差, 同时弥补地面站点的空间稀疏性。

参考文献

- [1] Forastiere, F., Bisceglia, L., Giua, R. and Hoffmann, B. (2024) Health Consequences of Prolonged Exposure to Multiple Industrial Air Pollutants in the EU.
- [2] Naimabadi, A., Shirmardi, M., Goudarzi, G., *et al.* (2022) *In vitro* cytotoxicity effects of polycyclic aromatic hydrocarbons (PAHs) associated with PM10 during the Middle Eastern Dust (MED) storms in Ahvaz. *Arabian Journal of Geosciences* **15**, Article No. 531. <https://doi.org/10.1007/s12517-022-09834-5>
- [3] Li, M. and Mallat, L. (2018) Health Impacts of Air Pollution. SCOR Paper.
- [4] Lin, Y., Zhang, L., Fan, Q., Meng, H., Gao, Y., Gao, H., *et al.* (2022) Decoupling Impacts of Weather Conditions on Interannual Variations in Concentrations of Criteria Air Pollutants in South China—Constraining Analysis Uncertainties by Using Multiple Analysis Tools. *Atmospheric Chemistry and Physics*, **22**, 16073-16090. <https://doi.org/10.5194/acp-22-16073-2022>
- [5] Lin, X. (2022) Major Developments in China's National Air Pollution Policies in the Early 12th Five-Year Plan. Institute for Global Environmental Strategies.
- [6] Diémoz, H., Barnaba, F., Magri, T., Pession, G., Dionisi, D., Pittavino, S., *et al.* (2019) Transport of Po Valley Aerosol Pollution to the Northwestern Alps—Part 1: Phenomenology. *Atmospheric Chemistry and Physics*, **19**, 3065-3095. <https://doi.org/10.5194/acp-19-3065-2019>
- [7] Zheng, J., Mao, X., Lv, X. and Jiang, W. (2020) The M2 Cotidal Chart in the Bohai, Yellow, and East China Seas from Dynamically Constrained Interpolation. *Journal of Atmospheric and Oceanic Technology*, **37**, 1219-1229. <https://doi.org/10.1175/jtech-d-19-0212.1>
- [8] Li, N., Xu, J. and Lv, X. (2021) Application of Dynamically Constrained Interpolation Methodology in Simulating National-Scale Spatial Distribution of PM2.5 Concentrations in China. *Atmosphere*, **12**, Article 272. <https://doi.org/10.3390/atmos12020272>
- [9] Li, N., Liu, Y., Lv, X., Zhang, J. and Fu, K. (2017) The High Order Conservative Method for the Parameters Estimation in a PM2.5 Transport Adjoint Model. *Advances in Meteorology*, **2017**, 1-13. <https://doi.org/10.1155/2017/4626585>
- [10] Wu, X., Xu, M., Gao, Y. and Lv, X. (2021) A Scheme for Estimating Time-Varying Wind Stress Drag Coefficient in the Ekman Model with Adjoint Assimilation. *Journal of Marine Science and Engineering*, **9**, Article 1220. <https://doi.org/10.3390/jmse9111220>
- [11] Xu, M.J., Fu, K. and Lv, X.Q. (2017) Application of Adjoint Data Assimilation Method to Atmospheric Aerosol Transport Problems. *Advances in Mathematical Physics*, **2017**, p. 5865403. <https://doi.org/10.1155/2017/5865403>
- [12] Lv, B.L., Hu, Y.T., Howard, H.C., Armistead, G.R., *et al.* (2017) Daily Estimation of Ground-Level PM2.5 Concentrations at 4 km Resolution over Beijing-Tianjin-Hebei by Fusing MODIS AOD and Ground Observations. *Science of the Total Environment*, **580**, 235-244. <https://doi.org/10.1016/j.scitotenv.2016.12.049>
- [13] Iyengar, G., Lam, H. and Wang, T.Y. (2024) Is Cross-Validation the Gold Standard to Evaluate Model Performance? arXiv preprint arXiv:2407.02754.
- [14] Zou, M., Jiang, W.G., Qin, Q.H., Liu, Y.C. and Li, M.L. (2022) Optimized XGBoost Model with Small Dataset for Predicting Relative Density of Ti-6Al-4V Parts Manufactured by Selective Laser Melting. *Materials*, **15**, Article 5298. <https://doi.org/10.3390/ma15155298>
- [15] Liu, Y., Yu, J., Shen, Y. and Lv, X. (2016) A Modified Interpolation Method for Surface Total Nitrogen in the Bohai Sea. *Journal of Atmospheric and Oceanic Technology*, **33**, 1509-1517. <https://doi.org/10.1175/jtech-d-15-0250.1>