

第五届全国研究生工业与经济金融大数据建模与计算大赛赛题（B）

$PM_{2.5}$ 浓度影响因素及时空规律探索

随着快速的工业化和大规模的能源消耗，全球正面临着严重的空气污染问题。空气污染直接关系到公共健康、经济的可持续性以及政府政策的制定。在众多大气污染物当中， $PM_{2.5}$ （指的是直径 2.5 微米或者更小的细颗粒物）是其中一个主要的污染物。一方面，对于公共健康来说，接触 $PM_{2.5}$ 空气污染物会显著增加自然因素导致的死亡率和多种疾病的死亡率。 $PM_{2.5}$ 可进入人体呼吸道，深入到肺部，甚至进入血液循环系统，对呼吸系统和心血管系统造成潜在威胁。因此，空气质量对于人们的健康至关重要。另一方面， $PM_{2.5}$ 污染还可能导致直接或者间接的经济损失。高水平的 $PM_{2.5}$ 污染通常会导致医疗支出的增加。减少 $PM_{2.5}$ 污染可以降低医疗开支、提高生产力和经济可持续性。综上所述，对 $PM_{2.5}$ 进行深入的研究和分析具有极其重要的意义。这不仅有助于减轻污染对人们健康和生计的威胁，还有助于推动经济的可持续发展。

问题一、收集了 2017 年 1 月 1 日至 2021 年 12 月 31 日某地区 13 个城市的 $PM_{2.5}$ 污染物数据（城市的经纬度已经给出），数据的采集频率为 3 小时一次。对 $PM_{2.5}$ 浓度进行时间和空间相关性的探索性数据分析。

问题二、大量研究表明 $PM_{2.5}$ 浓度受到气象因素的影响。为此收集了这些城市的气象因素数据，具体包括：

- **温度**：该变量为陆地、海洋或内陆水域表面以上 2 米处的空气温度，单位为开尔文，减去 273.15 即可转换为摄氏度。
- **降水量**：该变量为降落到地球表面的积聚的液体和冰冻的水，单位为米。
- **边界层高度**：该变量是地球表面附近的空气深度。当边界层高度较低时，会产生更高浓度的污染物。单位为米。
- **相对湿度**：该变量是水蒸气压力，作为空气饱和时的百分比。
- **地表气压**：该变量是陆地、海洋和内陆水域表面大气的压力。单位为帕斯卡。
- **风速**：提供了两个变量，分别是东向风速（U）和北向风速（V），是空气在地球表面上方 100 米处向东/北移动的水平速度，单位为米/秒。

有研究表明影响 $PM_{2.5}$ 浓度最主要的气象因素是风和湿度。例如，大量实证研究表明，在冬季的时候北风有利于驱散 $PM_{2.5}$ ，改善空气质量。建立模型探究哪些气象因素会显著影响 $PM_{2.5}$ 浓度，并且尝试解读其影响机制。

问题三、对 $PM_{2.5}$ 浓度的准确预测能够有效地加强污染物的监管和控制。利用所提供的数据对 $PM_{2.5}$ 浓度进行向前 1 期（3 小时）和向前 8 期（1 天）的预测，并且汇报均方误差（Root Mean Square Error, RMSE）评估模型的预测效果。提示：可以考虑空间自回归模型（Spatial VAR）、图神经网络模型（GNN）、大气污染物扩散模型或者自选其他可行的模型。

附件 数据文件

- 1. data.xlsx： 每条观测包括采集时间（精确到小时）、城市、 $PM_{2.5}$ 浓度以及气象变量。
- 2. location.csv： 13 个城市的经纬度。

变量名称	单位
温度	K
边界层高度	m
地表气压	Pa
降水量	m
相对湿度	%
U 水平风速	m/s
V 方向风速	m/s
PM2.5 浓度	$\mu g/m^3$