

空气质量监测数据校准模型研究

摘 要

空气中 PM2.5、PM10、CO、NO₂、SO₂、O₃ 是污染生态环境和危害人类健康的“两尘四气”，对其浓度实时监测可以及时掌握空气质量，并及时对污染源头采取相应管理措施。

本文主要对国控点数据（准确）和自建点数据（有差异）的数据进行探索分析，寻找造成差异的主要因素，根据国控点数据对自建点数据进行数据校准，由于全程基于数据，先对数据质量进行分析，对缺失数据进行处理，对两个附件数据进行整理为进一步探索做数据准备。

对问题一，主要对国控点和自建点数据分别从空气质量因素的整体分布、分位数、因素之间的相关性、因素与气象因素之间的相关性等角度进行探索分析，发现部分异常数据，得出 PM2.5 与 PM10 的相关程度比较高，NO₂ 与 SO₂ 的有一定的负相关，O₃ 与其他因素相关较弱。自建点数据与国控点数据在 CO、NO₂、SO₂ 差异比较大。SO₂ 的异常数据比较集中。

对问题二，主要从零点漂移、量程漂移，交叉影响、气象因素等方面进行差异因素分析，确定以小时为单位间隔的均值表示零点漂移，以天为单位间隔的标准差表示量程漂移，用 MATLAB 计算出各自的漂移程度（表和表），用回归分析的方法寻找各种产生差异的因素，如对 CO 数据差异的交叉影响主要是 SO₂，其次是 O₃ 和 NO₂；对 NO₂ 数据差异的交叉影响的主要是 CO 和 SO₂ 等；造成 PM2.5 和 PM10 的差异的因素是风速，其次是湿度，然后是压强；造成 CO 的差异的气象因素较弱，主要是风速；造成自建点 NO₂ 的差异的主要是风速，其次是温度、压强和湿度，降雨量的影响较小等；

对问题三，为了以国控数据为参考，校准自建点数据，先将自建点数据零点归零，量程归一，然后再按照国控数据的零点和量程进行校准，建立数学模型，代入自建点各因素数据进行校正，得到绝对误差和相对误差都很小校准结果（表），校准效果如图所示，该模型能很好的校准自建点数据。

关键词

空气质量 均值 标准差 校准 数学模型 MATLAB 差异分析 回归分析



1 问题描述

1.1 基本情况

空气中 PM2.5、PM10、CO、NO2、SO2、O3 是污染生态环境和危害人类健康的“两尘四气”，对其浓度实时监测可以及时掌握空气质量，并及时对污染源头采取相应管理措施。现对空气检测有两种方式，一是国家监测控制站点，简称国控点，对“两尘四气”的数据有效监测，而且相对准确，但国控点费用高，布控较少，且数据发布时间滞后，不能立刻给出实时空气质量的监测和预报。二是自主研发的微型空气质量检测仪，改检测仪器费用小，检测空气数据的同时可以监测温度、湿度、风速、气压、降水等气象参数，对地区空气质量可以进行即时网格化监控。

由于微型空气质量检测仪所使用的电化学气体传感器在长时间使用后会产生一定的零点漂移和量程漂移，而非常规气态污染物（气）浓度变化对传感器存在交叉干扰，天气因素对传感器也会产生影响，在国控点附近布控的自建点，同一时间微型空气质量检测仪所采集的数据与该国控点的数据值存在一定的差异，因此，需要利用国控点每小时的数据对国控点近邻的自建点数据进行校准。

1.2 问题提出

两个附件分别提供了一个时间段内某个国控点每小时的数据和附近的一个自建点数据，时间相应于国控点间隔在 5 分钟以内，现需要建立数学模型研究：

问题一、探索性分析自建点数据与国控点数据数据；

问题二、对导致自建点数据与国控点数据造成差异的因素进行分析；

问题三、建立数学模型利用国控点数据对自建点数据进行校准。

2 条件假设

2.1 条件假设

- 1、假设国控点与自建点距离很近，所测数据能反映统一网格内空气质量；
- 2、假设国控点检测数据是准确的；
- 3、假设不管是国控监测仪器还是自建检测仪器都是正常运行，没有损坏；



- 4、附件中缺失的部门小时的数据不影响自建点数据校正；
- 5、两附件中记录有数据的都是在仪器正常工作的时候的数据；
- 6、假设对“两尘”的监测仪器主要采用散光法，对“四气”的监测主要采用电化学气体法；
- 7、对电化学气体法检测气体的影响，此处忽略电压变化所造成的影响。

3 模型准备

3.1 名词解释

为了很好的理解问题和解决本问题，首先对涉及名词解释如下：

零点漂移：在未进行维修、保养或调解的前提下，仪器按规定的时间运行后，仪器的读数与零输入之间的偏差。

量程漂移：在未进行维修、保养或调节的前提下，仪器按规定的时间运行后，仪器的读数与已知参数考值之间的偏差。

PM2.5：环境空气中空气动力学当量直径小于等于 $2.5\ \mu m$ 的颗粒物，也称可细颗粒物。

PM10：环境空气中空气动力学当量直径小于等于 $10\ \mu m$ 的颗粒物，也称可吸入颗粒物。

单位：CO 的含量单位： mg/m^3 ；PM2.5、PM10、NO2、SO2、O3 含量单位： $\mu g/m^3$ 。

3.2 数据准备

由数据可知，附件一给出了某监测点的 2018 年 11 月 14 日 10:00 至 2019 年 6 月 11 日 15:00，每隔一小时的国控点“两尘四气”数据，附件 2 给出了同一地点的 2018 年 11 月 14 日 10:02 至 2019 年 6 月 11 日 16:32，间隔不超过 5 分钟的自建点“两尘四气”和风速、压强、降水量、温度、湿度的数据，经仔细查看，在这个时段中，两个附件中均有部分小时的数据缺失，例如国控数据中 2018 年 11 月 15 日 10 时至 12 时的数据没有记录，自建点数据中 2018 年 12 月 1 日 21 时和 22 时段的数据缺失，特别是国控

点缺失的数据，对相应时间的自建点数据校准会产生影响，因此删除部门因数据缺失而无法校准的数据共 30803 个，根据假设 4，这部分缺失数据的处理将不影响数据校准，后面的数据分析和差异因素的实施将建立在新数据基础上。

3.3 问题分析

根据问题描述中的“两尘四气”是环境污染的主要有害物，中华人民共和国环境保护标准专门制定了对颗粒 PM2.5、PM10 和污染气体 CO、NO₂、SO₂、O₃ 的监测仪器、数据分析的相关标准[1]，不同的监测仪器读取的数据会有差异，要使用相对准确的监测数据对其他有差异的仪器读取的数据进行校准，校准数学模型是关键，在进行建模之前用统计学知识进行数据相关探索。

问题一、为了详尽的探索国控点数据和自建点数据，根据整理后的数据从均值、标准差、中位数、相关性、关联图等角度对各项因素进行探索性分析，国控点数据是相对准确的数据，将是自建点数据校准的参考。

问题二、不同的监测仪，因各种原因所得到的数据有差异，但环境是相同的，想要找出差异，均采用以时为单位，将两组数据各个因素分别从零点漂移因素、量程漂移因素、污染物（气）浓度变化交叉干扰因素、气象因素等分别进行差异对比。

目前型空气质量传感监测仪采用光散射法和电化学法监测大气中“两尘四气”其中 PM10、PM2.5 是采用光散射法，而 SO₂、NO₂、O₃、CO 是采用电化学法。

问题三、国控点数据是相对准确的数据，以国控点数据为准，根据问题二中探索的差异因素，将自建点数据校准。

4 建立模型与求解

4.1 问题一 探索性数据分析

4.1.1 国控点数据探索分析

(1) 整体分布

国控点数据是每小时一组，用 MATLAB 画出整理后的“两尘四气”六组数据的分布图（图 1）

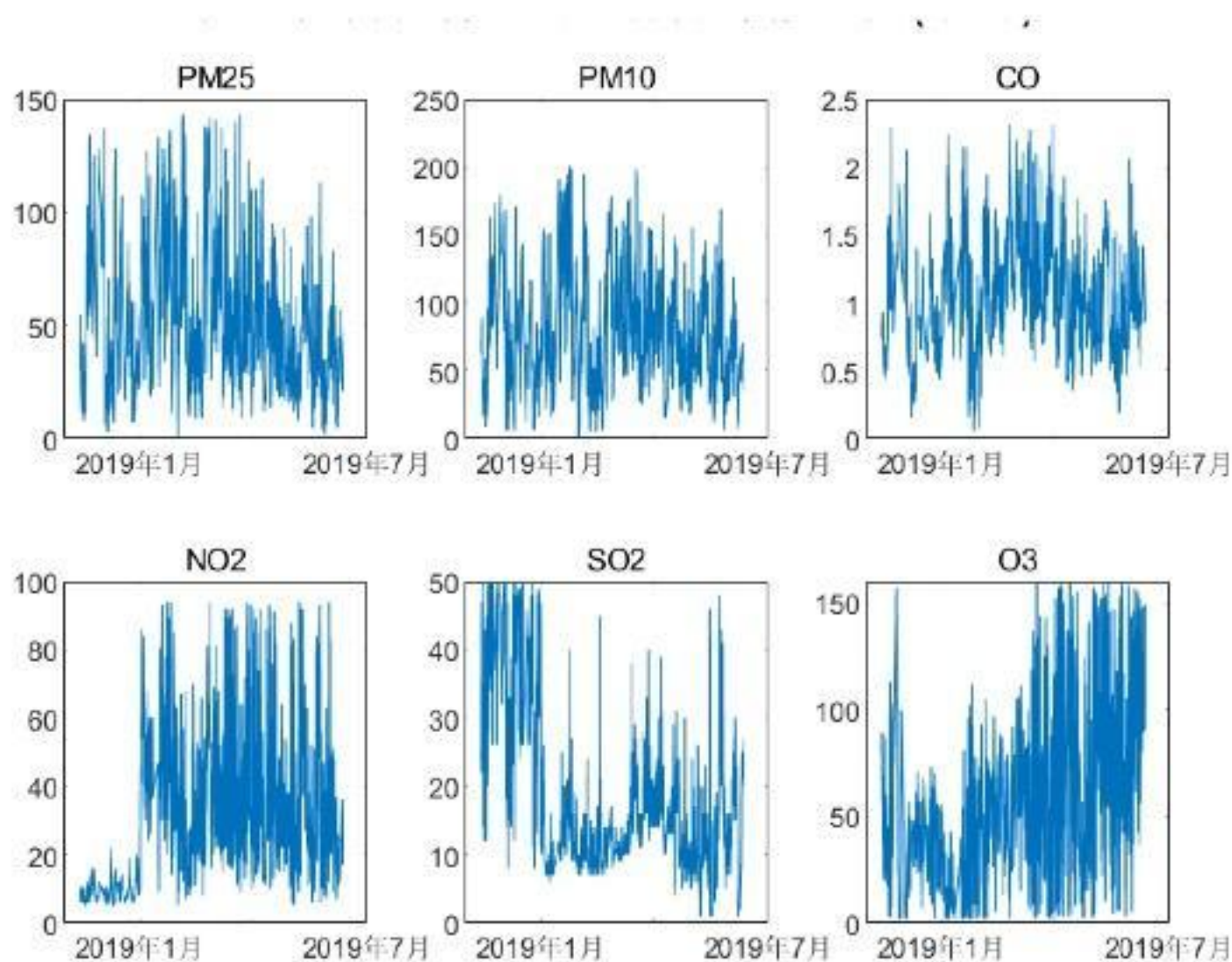


图 1 国控点时数据分布图

从图 1 中可以看出 PM2.5、NO₂、O₃ 的浓度在该时间段内的变化较大，主要集中在 0 到 100 微克每立方米；SO₂ 的在时间段内的浓度变化较为显著，且浓度值较小，主要集中在 0 到 30 微克每立方米；PM10 的浓度变化在时间段内的较小，主要集中在 50 到 100 微克每立方米；CO 在空气中的浓度相较于其他几种污染物而言偏大，主要集中在 0.5 到 1.5 毫克每立方米之间，即 500 到 1500 微克每立方米。

(2) 分位数

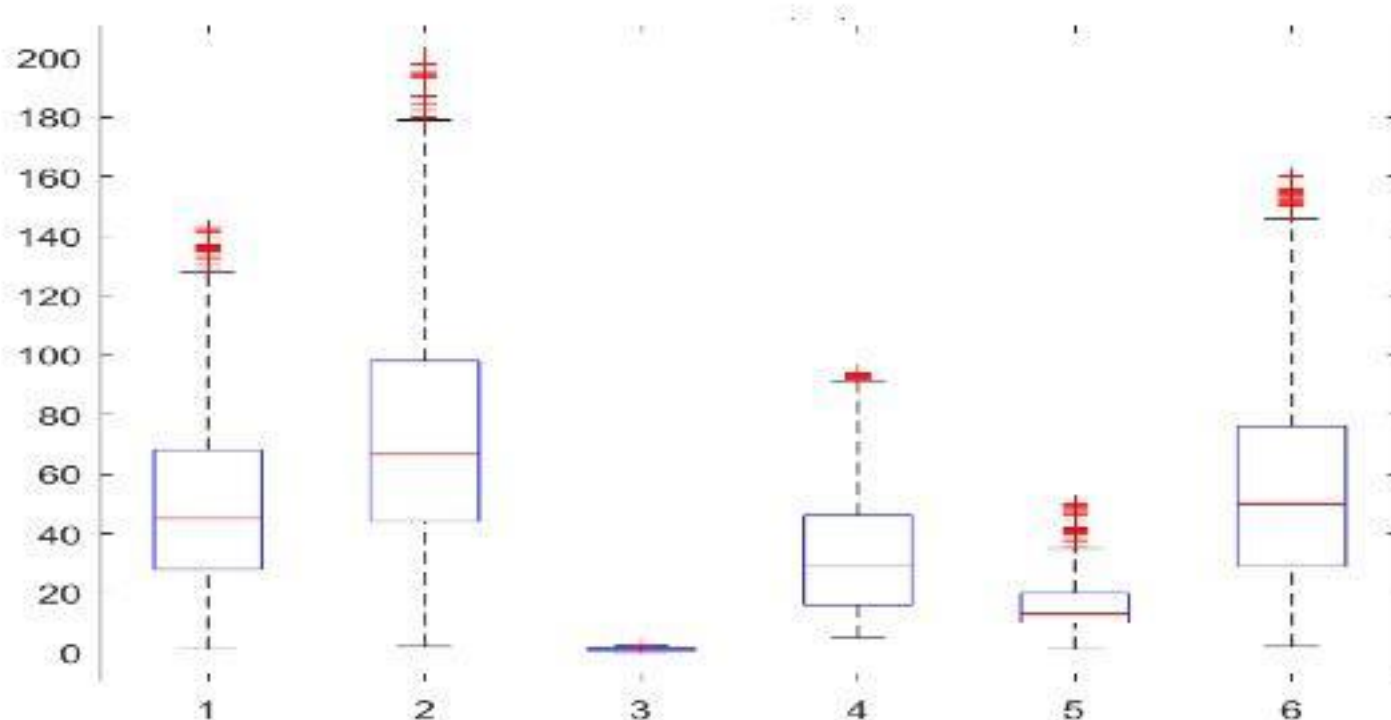


图 2 国控点数据箱位图

从箱位图（图 2）可以看出，国控点数据大部分在 10 分位到 90 分位之间，有少部分异常数据，其中 CO 的异常数据最少，其次是 NO₂ 的异常数据较少。

（3）相关性

相关性可以很好的反映因素之间的相关关系，是探索因素之间关系的一个重要手段，相关系数模型为

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

根据国控数据六个因素的数据，代入公式，用 MATLAB 计算相关系数如表（表 1）所示

表 1 国控点六因素的相关系数表（时）

| | PM2.5 | PM10 | CO | NO ₂ | SO ₂ | O ₃ |
|-----------------|---------|---------|---------|-----------------|-----------------|----------------|
| PM2.5 | 1.0000 | 0.8157 | 0.6624 | 0.2590 | 0.2713 | -0.2690 |
| PM10 | 0.8157 | 1.0000 | 0.5822 | 0.3064 | 0.3064 | -0.1765 |
| CO | 0.6624 | 0.5822 | 1.0000 | 0.2983 | 0.3119 | -0.2737 |
| NO ₂ | 0.2590 | 0.3064 | 0.2983 | 1.0000 | -0.3440 | -0.2544 |
| SO ₂ | 0.2713 | 0.3064 | 0.3119 | -0.3440 | 1.0000 | -0.2840 |
| O ₃ | -0.2690 | -0.1765 | -0.2737 | -0.2544 | -0.2840 | 1.0000 |

画出六因素关联图（图 3）：

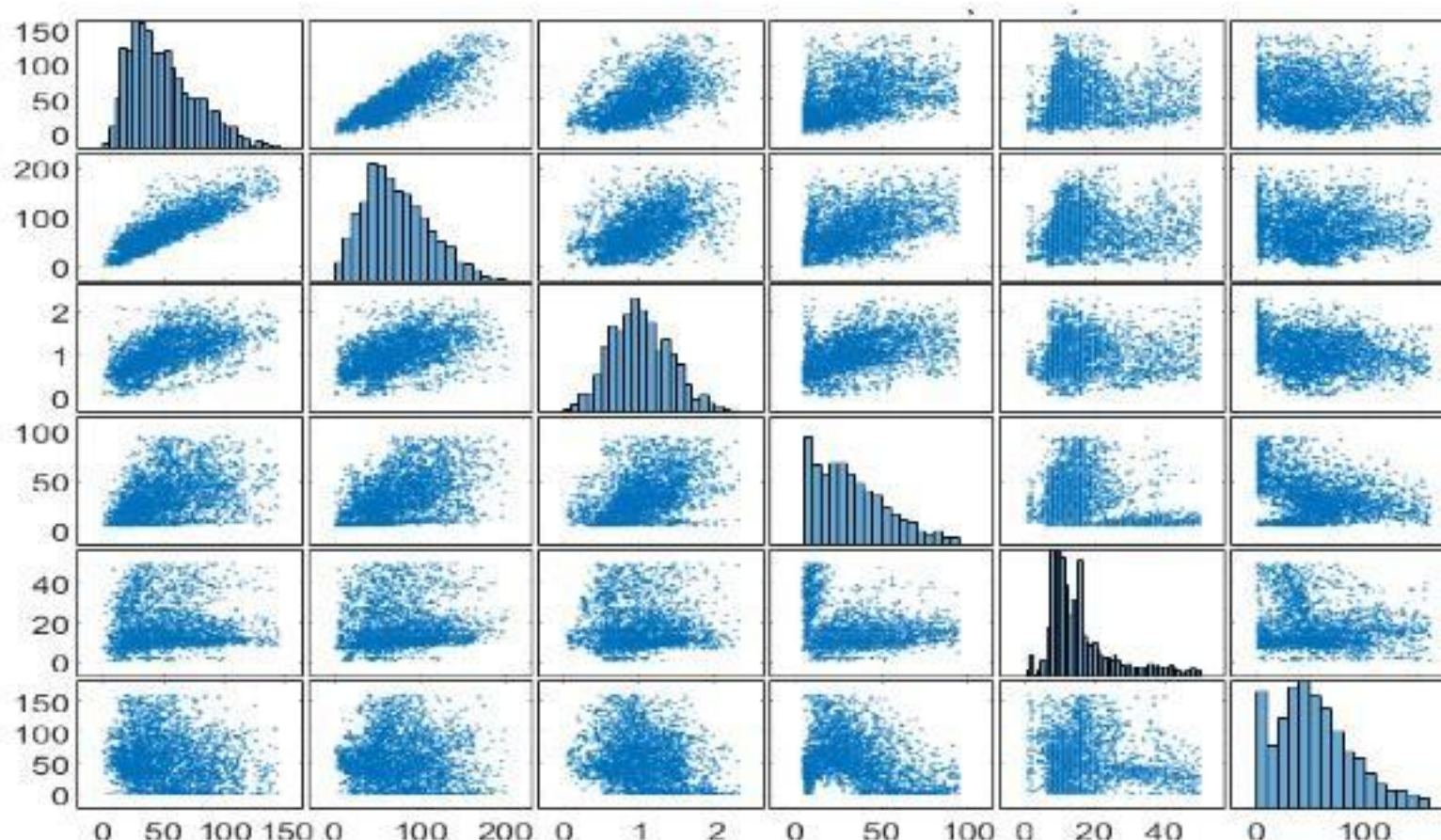


图 3 国控点数据（小时）六因素关联图

从表 1 和图 3 可得信息：PM2.5 与 PM10 相关性最大，PM2.5 与 CO 相关性其次，都是成正相关；NO2 与 SO2 线性相关程度相对较大，负相关；O3 与其他各因素直接相关性较小。

4.1.2 自建点数据探索分析

(1) 整体分布

自建点数据是 5 分钟以内一组，用 MATLAB 画出整理后的“两尘四气”六组数据的分布图（如图 4）和气象因素分布图（如图 5）

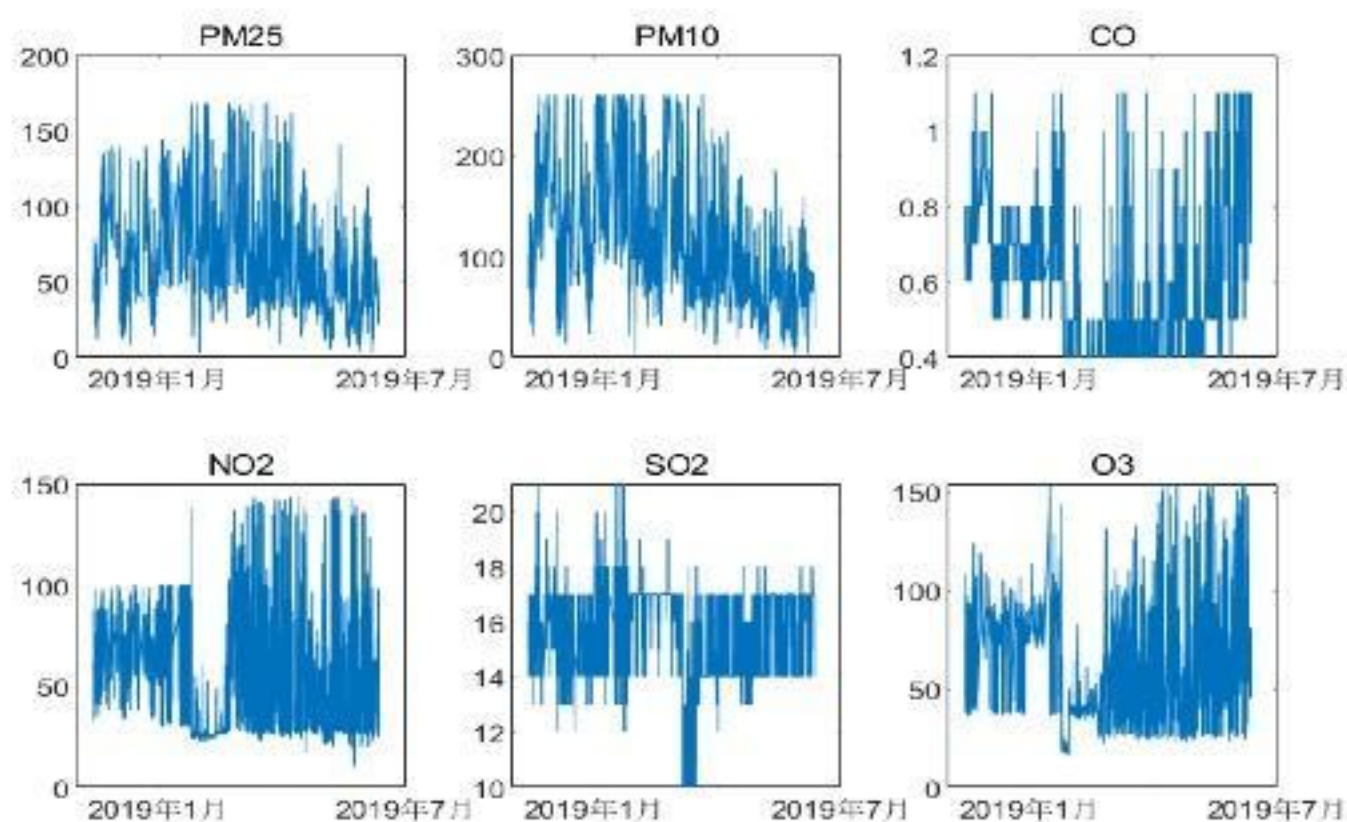


图 4 数据分布子图

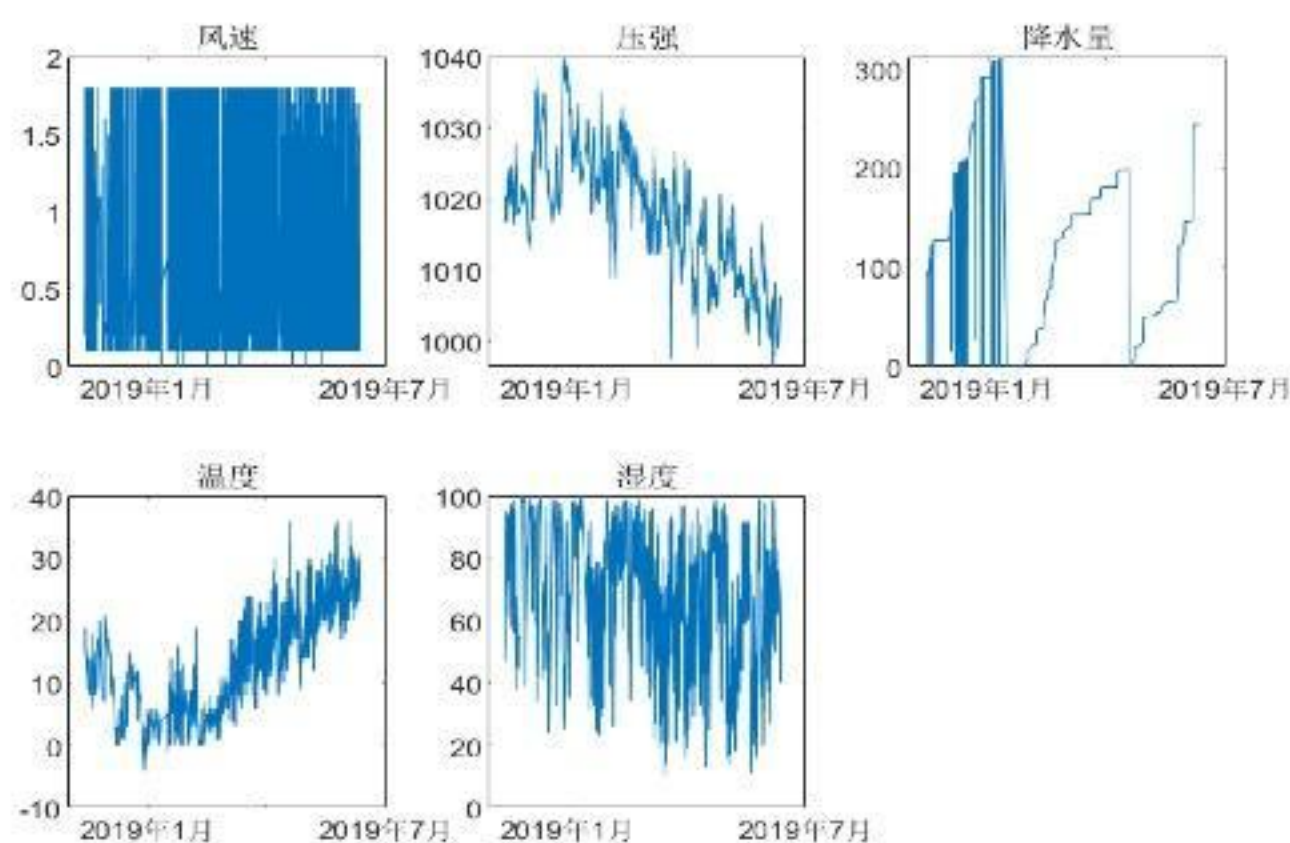


图 5 气象数据分布子图

从图中可以看出自建点的“两尘四气”的变化趋势与国控数据大致相当，温度和气压数据变化趋势明显，该段时间内风速长期维持，降水充足，温度适宜，湿度偏大

(2) 分位数

为很好了解数据的结构，画出箱体图（图 6）

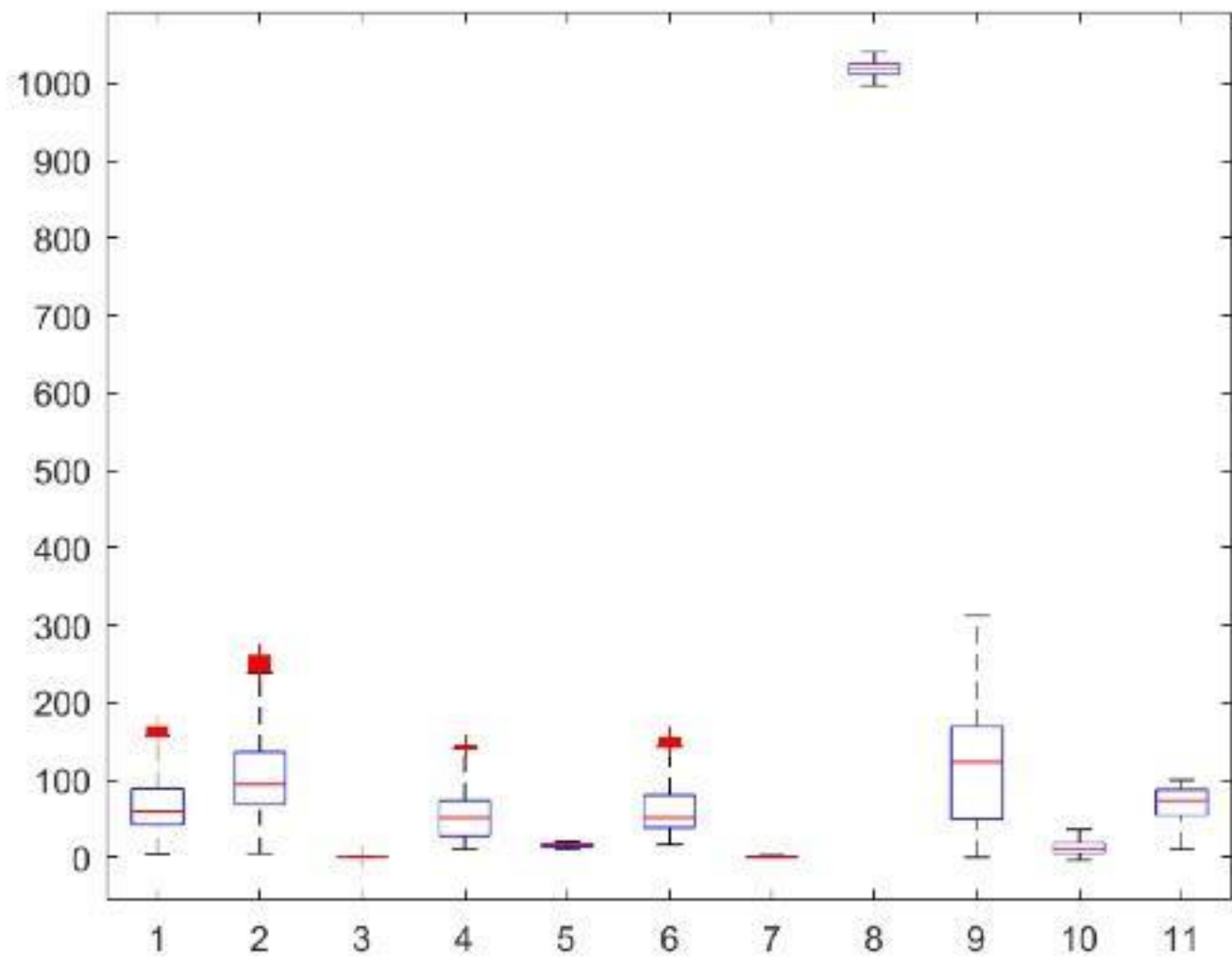


图 6 自建点数据箱体图

从箱体图（图 6）可以看出“两尘四气”中异常数据比较少，气象数据没有异常数据，数据质量比较高。

(3) 相关性

根据相关性公式，做出相关系数如表（表 2）

表 2 “两尘四气”与气象资料之间的相关系数

| | PM2.5 | PM10 | CO | NO ₂ | SO ₂ | O ₃ | 风速 | 压强 | 降雨量 | 温度 | 湿度 |
|-----------------|-------|------|------|-----------------|-----------------|----------------|-------|-------|-------|-------|------|
| PM2.5 | 1.00 | 0.95 | 0.13 | 0.26 | 0.42 | -0.11 | -0.14 | 0.10 | -0.02 | -0.27 | 0.33 |
| PM10 | 0.95 | 1.00 | 0.27 | 0.30 | 0.40 | 0.02 | -0.13 | 0.26 | 0.08 | -0.40 | 0.36 |
| CO | 0.13 | 0.27 | 1.00 | 0.41 | 0.19 | 0.43 | -0.11 | -0.10 | 0.19 | 0.20 | 0.07 |
| NO ₂ | 0.26 | 0.30 | 0.41 | 1.00 | 0.25 | 0.10 | -0.34 | -0.07 | 0.35 | 0.00 | 0.20 |

| | | | | | | | | | | | |
|-----------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| SO ₂ | 0.42 | 0.40 | 0.19 | 0.25 | 1.00 | -0.29 | -0.41 | 0.03 | -0.25 | -0.26 | 0.48 |
| O ₃ | -0.11 | 0.02 | 0.43 | 0.10 | -0.29 | 1.00 | 0.21 | 0.00 | 0.31 | 0.23 | -0.34 |
| 风速 | -0.14 | -0.13 | -0.11 | -0.34 | -0.41 | 0.21 | 1.00 | 0.19 | 0.12 | -0.03 | -0.32 |
| 压强 | 0.10 | 0.26 | -0.10 | -0.07 | 0.03 | 0.00 | 0.19 | 1.00 | 0.21 | -0.85 | 0.04 |
| 降雨量 | -0.02 | 0.08 | 0.19 | 0.35 | -0.25 | 0.31 | 0.12 | 0.21 | 1.00 | -0.12 | 0.05 |
| 温度 | -0.27 | -0.40 | 0.20 | 0.00 | -0.26 | 0.23 | -0.03 | -0.85 | -0.12 | 1.00 | -0.40 |
| 湿度 | 0.33 | 0.36 | 0.07 | 0.20 | 0.48 | -0.34 | -0.32 | 0.04 | 0.05 | -0.40 | 1.00 |

为直观反映因素之间的相关性，画出自建点的“两尘四气”与环境个因素的关联图（图 7）

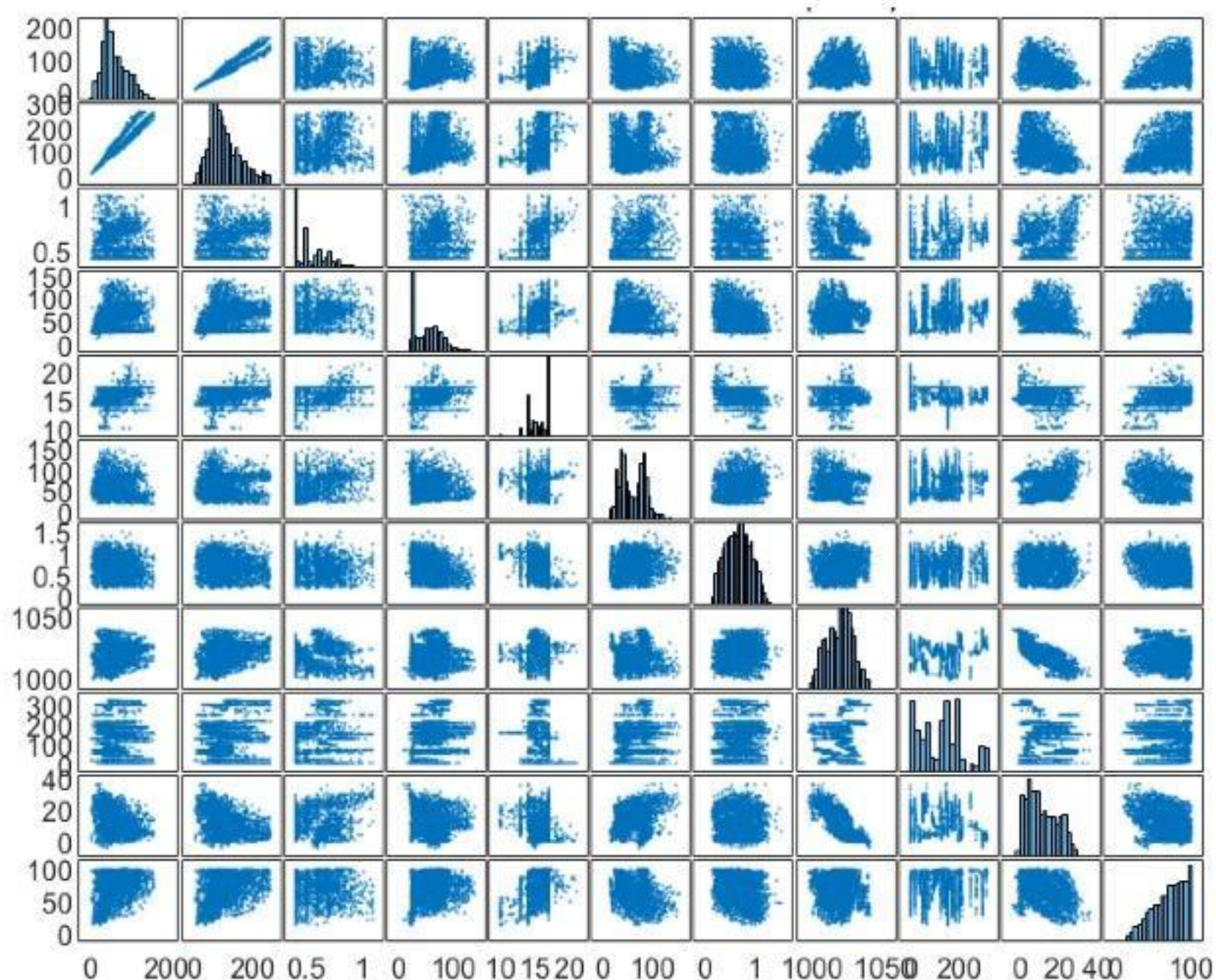


图 7 自建点关联图（时）

从相关系数和因素关联图可以看出，对于“两尘四气”的各因素的相关性，与国控

点基本一致，对于气象数据与“两尘四气”的各因素之间的相关系较为复杂，其中 SO₂ 与湿度和温度有一定相关性；NO₂ 余降雨量和风速有一定相关性，与压强和温度相关性小；CO 与其他气象因素相关性小；PM_{2.5}、PM₁₀ 与温度和湿度有一定相关性，与降雨量的相关性小。

4.1.3 国控点与自建点之间的数据探索分析

由所给数据可知，国控点数据是每小时一组，二自建点数据是五分钟以内一组，为了更好的对比国控点数据和自建点数据之间的差异，将整理后的数据都统筹到每小时一组。通过数据观察可知，一小时内各个污染因素的变化不大，所以将所有时的数据统计一起，求其平均值，计算模型如下：

$$X_k = \frac{\sum_{i \in k} x_i}{n_k}$$

例如，2018 年 11 月 14 日 10:02 至 10:58 分的数据的平均值作为 10:00 整的数据，用 MATLAB 编程可得自建点时数据，为探索他们之间的关系，会出图形如图（图 8）

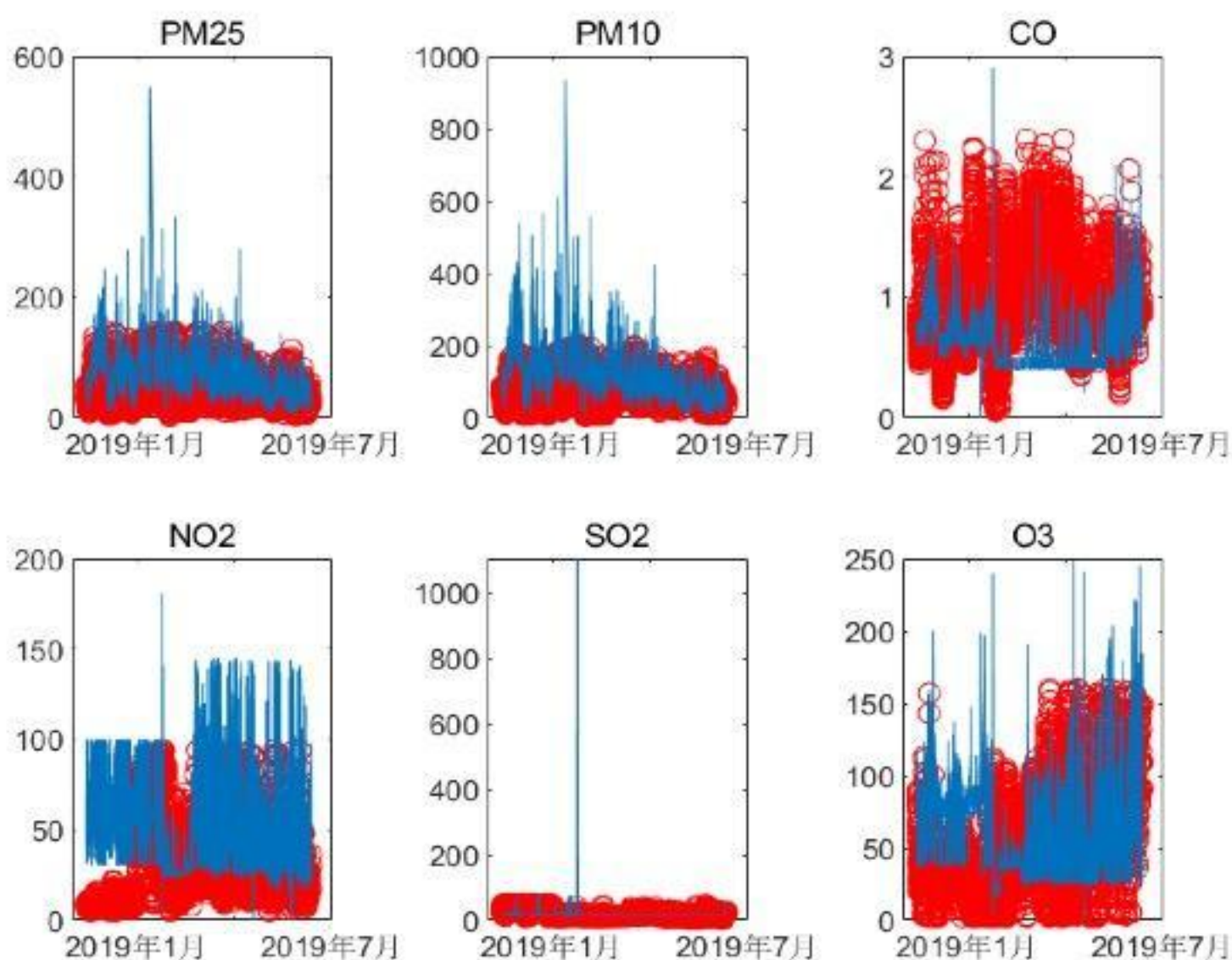


图 8 国控数据与自建数据分布图（蓝线条为自建点数据，红圈为国控点数据）

从图 8 中可知，两类数据基本反映了相同趋势的环境状态，但有差距，其中 PM_{2.5}、

PM10、O3 变化趋势差不多，CO、NO2、SO2 差异比较大。SO2 的异常数据比较集中。

4.2 问题二 造成自建点与国控点数据差异因素分析

4.2.1 造成差异的零点漂移分析

从问题一中的数据探索（图 8）可知，国控点数据与自建点数据的 PM2.5、PM10、O3 的差异相对较弱，CO、NO2、SO2 差异较大，因此用二者之差的绝对值作为其数据差异。

根据名词解释，零点漂移是指读数与零输入之间的差距，建立数学模型得

$$d_k = |Y_k - X_k| = \left| Y_k - \frac{\sum_{i \in k} x_i}{n_k} \right|$$

将自建点数据按平均的方法整合成以小时为间隔数据，代入数据，用 MATLAB 求出国控点数据与自建点数据在六个因素上的零点漂移如下表（表 3，选取了部分数据，完整数据见附件 1）：

表 3 六因素零点漂移差异（时）

| 时间 | PM2.5 | PM10 | CO | NO2 | SO2 | O3 |
|------------------|---------|---------|--------|---------|---------|---------|
| 2018/11/14 10:00 | 13.6500 | 22.1250 | 0.0460 | 47.4500 | 9.8750 | 27.0000 |
| 2018/11/14 11:00 | 8.7895 | 13.7895 | 0.0114 | 37.3684 | 6.6842 | 21.3158 |
| 2018/11/14 12:00 | 11.5200 | 25.1200 | 0.0480 | 34.6000 | 10.8400 | 18.3600 |
| 2018/11/14 13:00 | 6.1429 | 19.7143 | 0.0071 | 37.2857 | 11.0000 | 18.1905 |
| 2018/11/14 14:00 | 20.5455 | 26.9545 | 0.1050 | 52.8182 | 18.9091 | 29.6818 |
| 2018/11/14 15:00 | 16.1905 | 19.2381 | 0.0487 | 52.4762 | 21.7143 | 27.3333 |
| 2018/11/14 16:00 | 12.0000 | 17.0333 | 0.0480 | 61.0000 | 31.7000 | 20.4333 |
| 2018/11/14 21:00 | 19.0000 | 35.5789 | 0.0936 | 70.2632 | 22.0000 | 46.6316 |
| 2018/11/14 22:00 | 23.1731 | 45.8654 | 0.1840 | 63.2500 | 20.2692 | 40.7692 |
| 2018/11/14 23:00 | 17.5333 | 44.7667 | 0.1790 | 60.5667 | 20.3333 | 41.4333 |
| 2018/11/15 0:00 | 18.4583 | 41.3333 | 0.1355 | 57.4583 | 16.2917 | 26.3750 |

| | | | | | | |
|------------------|---------|---------|--------|---------|---------|---------|
| 2018/11/15 1:00 | 8.3478 | 33.8696 | 0.1216 | 49.7391 | 10.7826 | 19.3043 |
| 2018/11/15 2:00 | 2.2500 | 33.5833 | 0.0727 | 46.5833 | 9.7917 | 26.6250 |
| 2018/11/15 3:00 | 9.0000 | 33.0435 | 0.0347 | 45.5652 | 8.9565 | 23.6957 |
| 2018/11/15 4:00 | 15.2200 | 45.0800 | 0.0430 | 46.0600 | 7.5800 | 11.3200 |
| 2018/11/15 5:00 | 29.9231 | 68.1923 | 0.1656 | 64.1154 | 21.4231 | 51.1154 |
| 2018/11/15 6:00 | 29.9615 | 69.3846 | 0.0920 | 66.0385 | 18.2308 | 51.2692 |
| 2018/11/15 7:00 | 13.7407 | 65.3704 | 0.0060 | 67.3704 | 16.6667 | 49.3333 |
| 2018/11/15 8:00 | 13.9524 | 70.9524 | 0.0343 | 68.7619 | 17.4762 | 48.2857 |
| 2018/11/15 9:00 | 17.2273 | 60.5000 | 0.0421 | 71.1364 | 21.5909 | 50.6364 |
| 2018/11/15 13:00 | 13.3333 | 30.7500 | 0.0840 | 68.1250 | 20.6250 | 64.7500 |
| 2018/11/15 14:00 | 16.6250 | 46.5000 | 0.1062 | 72.1250 | 26.5833 | 67.2500 |
| 2018/11/15 16:00 | 15.6071 | 39.7500 | 0.0274 | 82.2143 | 34.6071 | 74.7500 |
| 2018/11/15 18:00 | 9.6400 | 15.2000 | 0.0950 | 65.3200 | 22.6000 | 61.2400 |
| 2018/11/15 19:00 | 13.7917 | 33.5000 | 0.1802 | 63.0417 | 19.2083 | 55.8333 |
| 2018/11/15 20:00 | 17.9130 | 53.7826 | 0.1110 | 69.2174 | 28.0435 | 65.7826 |
| ... | ... | ... | ... | ... | ... | ... |

4.2.2 造成差异的量程漂移分析

根据名词解释，量程漂移是指读数所度量的程度与实际度量程度的差异，建立数学模型如下：

$$l_r = |\sigma(x_r) - \sigma(y_r)|$$

由于，国控数据是按每小时给的，一小时只有一个数值，无法求出一小时内的量程，所以用日量程来表示，编程求解得到量程漂移如表（表 4，完整数据见附件 1）。

表 4 六因素量程漂移差异（天）

| 对应时间 | PM2.5 | PM10 | CO | NO2 | SO2 | O3 |
|------------------|---------|---------|---------|---------|---------|---------|
| 2018/11/14 10:00 | 3.7964 | 0.1612 | 0.0245 | 10.3456 | 5.7997 | 3.3973 |
| 2018/11/15 10:00 | 0.5830 | 3.3586 | 0.0096 | 12.6366 | 10.0527 | 8.7588 |
| 2018/11/16 10:00 | 3.3171 | 17.4448 | 0.0285 | 6.7959 | 4.2309 | 0.4781 |
| 2018/11/17 10:00 | 2.5784 | 1.2093 | 0.0083 | 11.1919 | 9.6140 | 7.1610 |
| 2018/11/18 10:00 | 1.1601 | 9.9634 | 0.2136 | 8.7443 | 6.8018 | 0.7485 |
| 2018/11/19 10:00 | 2.3312 | 8.4149 | 0.1212 | 5.9249 | 2.4112 | 6.7443 |
| 2018/11/20 10:00 | 3.6054 | 16.9900 | 0.0769 | 12.7276 | 5.8051 | 26.9299 |
| 2018/11/21 10:00 | 8.6482 | 2.2516 | 0.3057 | 3.1256 | 2.5142 | 6.3571 |
| 2018/11/22 10:00 | 2.7986 | 5.8490 | 0.1005 | 10.5374 | 4.2853 | 6.4997 |
| 2018/11/23 10:00 | 0.3303 | 10.8740 | 0.0349 | 11.4531 | 9.4092 | 0.0225 |
| 2018/11/25 10:00 | 4.2733 | 3.1232 | 0.0032 | 8.5923 | 0.6502 | 7.0297 |
| 2018/11/26 10:00 | 2.3428 | 7.7955 | 0.1346 | 9.9826 | 6.9743 | 6.5180 |
| 2018/11/27 10:00 | 2.5256 | 4.2555 | 0.0203 | 10.2788 | 0.4250 | 2.8546 |
| 2018/11/28 10:00 | 63.7003 | 57.2003 | 64.2003 | 58.2003 | 63.7003 | 62.7003 |
| 2018/11/29 10:00 | 56.7544 | 50.4957 | 63.3559 | 53.4369 | 62.8301 | 52.4082 |
| 2018/12/1 10:00 | 4.6889 | 9.1538 | 0.0471 | 9.1241 | 6.1243 | 1.0765 |
| 2018/12/2 10:00 | 8.7655 | 14.9677 | 0.0937 | 5.7230 | 4.0073 | 2.4247 |
| 2018/12/3 10:00 | 7.5656 | 21.9148 | 0.2348 | 11.3995 | 7.9331 | 3.3098 |
| 2018/12/4 10:00 | 6.8658 | 23.2946 | 0.0382 | 5.4516 | 4.6169 | 8.7025 |
| 2018/12/5 10:00 | 1.3317 | 9.0348 | 0.0753 | 14.3982 | 10.8410 | 5.9551 |
| 2018/12/6 10:00 | 1.0382 | 17.3375 | 0.1687 | 7.6602 | 4.9706 | 6.4981 |

| | | | | | | |
|------------------|---------|---------|--------|---------|--------|---------|
| 2018/12/7 10:00 | 0.2716 | 4.3743 | 0.0190 | 5.2123 | 4.5321 | 3.4665 |
| 2018/12/8 10:00 | 0.8375 | 4.8665 | 0.0848 | 3.1969 | 3.5043 | 0.1412 |
| 2018/12/9 10:00 | 0.8731 | 1.8851 | 0.0467 | 6.0823 | 8.8496 | 3.7491 |
| 2018/12/10 10:00 | 3.8848 | 17.4249 | 0.0314 | 11.1139 | 9.9421 | 0.3050 |
| 2018/12/11 10:00 | 12.3807 | 7.0771 | 0.1381 | 3.4769 | 4.7068 | 3.1091 |
| 2018/12/12 10:00 | 0.0167 | 4.7541 | 0.1484 | 5.1444 | 2.6941 | 12.0790 |
| 2018/12/13 10:00 | 5.3215 | 2.4335 | 0.0195 | 5.1847 | 4.8134 | 2.4011 |
| 2018/12/14 10:00 | 2.5564 | 5.2371 | 0.0132 | 7.1870 | 4.9068 | 0.3705 |
| 2018/12/15 10:00 | 9.8752 | 40.6976 | 0.1057 | 8.5914 | 5.8183 | 2.2465 |
| 2018/12/16 10:00 | 1.9390 | 4.5074 | 0.0214 | 4.4068 | 4.2700 | 7.6396 |
| ... | ... | ... | ... | ... | ... | ... |

4.2.3 造成差异的污染物（气）交叉干扰分析

根据假设 6，粉尘的监测仪器是采用光散射法，电化学传气体传感器是监测污染气体（“四气”）浓度的，因此，因污染气浓度变化对传感器会产生交叉干扰就只有气体，以国控点与自建点气体数据差异为因变量，其他气体及交叉项为自变量，建立多元回归模型

$$d_l = f(X_l) + \varepsilon_l \quad (\varepsilon_l \text{ 为第 } l \text{ 个随机影响因子})$$

由于有交叉影响，所以模型确定为

$$d_l = a_{l1}X_{l1} + a_{l2}X_{l2} + a_{l3}X_{l3} + a_{l4}X_{l1}X_{l2} + a_{l5}X_{l1}X_{l3} + a_{l6}X_{l2}X_{l3} + \varepsilon_l$$

$$, \quad l = 1, 2, 3, 4$$

其中 d_l 表示国控点与自建点的数据差异， X_l 表示气体中除了因变量以外的其他三个因子，用 MATLAB 编程，逐步回归分别得到四种气体的交叉影响因素的回归系数与置信区间及相关统计量分别如表 5—表 8

表 5 造成自建点 CO 浓度差异的交叉影响因素回归系数及统计量

| | 回归系数 | 置信区间 |
|-------|--------------|---|
| a_1 | 0.838640877 | TRUE |
| a_2 | -0.002637036 | [-0.00482860232935142,-0.00044546886846622] |
| a_3 | -0.032052149 | [-0.0419944858088612,-0.0221098120165993] |
| a_4 | -0.009595304 | [-0.0114641080368416,-0.00772649922774414] |
| a_5 | 0.000224751 | [9.91153190761922E-05,0.000350387573322156] |
| a_6 | 1.14E-05 | [3.78785390058088E-06,1.91014268788494E-05] |
| a_7 | 0.000755604 | [0.000627823318737813,0.000883384389445674] |

表 6 造成自建点 NO₂ 浓度差异的交叉影响因素回归系数及统计量

| | 回归系数 | 置信区间 |
|-------|--------------|--|
| a_1 | 16.15026762 | [-17.9457973684333,50.246332605258] |
| a_2 | 136.0616175 | [70.8450994663302,201.278135573797] |
| a_3 | -1.77845981 | [-3.85378328013802,0.2968636600797] |
| a_4 | -0.921094113 | [-1.288436889379,-0.55375133639111] |
| a_5 | -0.701362041 | [-4.48731361842689,3.08458953580456] |
| a_6 | -1.0645198 | [-1.28991687940105,-0.839122719993172] |
| a_7 | 0.097668639 | [0.0717528210197347,0.123584457839633] |

表 7 造成自建点 SO₂ 浓度差异的交叉影响因素回归系数及统计量

| | 回归系数 | 置信区间 |
|-------|-------------|--------------------------------------|
| a_1 | 156.9419301 | [128.924031206347,184.959828984221] |
| a_2 | 29.23625505 | [-27.0662326724419,85.5387427766777] |

| | | |
|-------|--------------|--|
| a_3 | -0.253693216 | [-0.611120506499765,0.103734074763409] |
| a_4 | -9.040501906 | [-10.7905617941779,-7.2904420168223] |
| a_6 | 3.367310093 | [-0.131515088393034,6.86613527502903] |
| a_7 | 0.015276172 | [-0.0069965580461346,0.0375489019998208] |

表 8 造成自建点 O3 浓度差异的交叉影响因素回归系数及统计量

| | 回归系数 | 置信区间 |
|-------|--------------|--|
| a_1 | 155.3082608 | [127.152075868251,183.464445679913] |
| a_2 | 32.99423239 | [-23.6729380680243,89.6614028538028] |
| a_3 | -0.310894888 | [-0.681511825021926,0.0597220490821466] |
| a_4 | -8.714017635 | [-10.5513080712488,-6.87672719936623] |
| a_5 | 0.121237047 | [-0.0865867791405965,0.329060872309548] |
| a_6 | 2.686654413 | [-1.00143250080032,6.37474132774853] |
| a_7 | 0.014843902 | [-0.00744009711945207,0.037127901449762] |

结论：通过对回归系数的绝对值的大小可判断，对自建点 CO 数据产生差异的交叉影响的很少，相对来说 SO2 影响要强一点，其次是 O3 和 NO2；对自建点 NO2 数据产生差异的交叉影响的主要是 CO 和 SO2；对自建点 O3 数据产生差异的交叉影响的主要是 CO 和 SO2；对自建点 SO2 数据产生差异的交叉影响的主要是 CO 和 O3。

4.2.4 造成差异的气象环境因素分析

气象因素也是影响检测数据之一，因此建立以国控与自建数据差异为因变量，气象因素为自变量的回归模型

$$d_l = f(X_l) + \varepsilon_l \quad (\varepsilon_l \text{ 为第 } l \text{ 个随机影响因子})$$

此处只需要得出哪些因素是产生差异的主要因素，因此建立线性回归模型，即

$$d_l = \beta_{l0} + \beta_{l1}X_{l1} + \beta_{l2}X_{l2} + \beta_{l3}X_{l3} + \beta_{l4}X_{l4} + \beta_{l5}X_{l5} + \varepsilon_l$$

$$(l = 1, 2, 3, 4, 5, 6)$$

用 MATLAB 编程，逐步回归分别得到六个因素与气象因素的回归系数与置信区间及相关统计量分别如表 9-表 14:

表 9 自建点 PM2.5 差异与气象因素的回归系数

| | 回归系数 | 置信区间 |
|------------------------------|----------|----------------------|
| β_0 | -254.222 | -349.1090, -159.3356 |
| β_1 | 3.73816 | 2.2975, 5.1789 |
| β_2 | 0.2394 | 0.1480, 0.3308 |
| β_3 | 0.0109 | 0.0069, 0.0148 |
| β_4 | -0.1757 | -0.2847, -0.0668 |
| β_5 | 0.3536 | 0.3313, 0.3759 |
| R2=0.459, F=551.438, P<0.001 | | |

表 10 自建点 PM10 差异与气象因素的回归系数

| | 回归系数 | 置信区间 |
|-----------|----------|---------------------|
| β_0 | -1726.74 | -1834.14, -1619.34 |
| β_1 | 5.1731 | 1.196823, 9.1494406 |
| β_2 | 1.6267 | 1.520307, 1.733066 |
| β_3 | 0.0508 | 0.04011, 0.061429 |
| β_5 | 1.3527 | 1.306957, 1.398514 |

表 11 自建点 CO 差异与气象因素的回归系数

| | 回归系数 | 置信区间 |
|--|------|------|
|--|------|------|

| | | |
|-----------|----------|-------------------|
| β_0 | -43.3052 | -46.9783,-39.6321 |
| β_1 | 0.1468 | 0.09105,0.20259 |
| β_2 | 0.044125 | 0.037714,0.044787 |
| β_3 | -0.0005 | -0.00062,-0.00031 |
| β_4 | 0.045 | 0.040831,0.049265 |
| β_5 | 0.0037 | 0.002838,0.004563 |

表 12 自建点 NO₂ 差异与气象因素的回归系数

| | 回归系数 | 置信区间 |
|-----------|----------|-------------------|
| β_0 | -1391.69 | -1590.14,-1193.24 |
| β_1 | 6.3493 | 3.336144,9.362355 |
| β_2 | 1.3103 | 1.119311,1.501463 |
| β_3 | 0.1043 | 0.096096,0.112545 |
| β_4 | 1.9264 | 1.69857,2.154226 |
| β_5 | 0.5695 | 0.52288,0.616052 |

表 13 自建点 SO₂ 差异与气象因素的回归系数

| | 回归系数 | 置信区间 |
|-----------|---------|-----------------------|
| β_0 | 725.433 | 634.0592223,816.8076 |
| β_2 | -0.6997 | -0.787557217,-0.61188 |
| β_3 | -0.0293 | -0.033127343,-0.02555 |

| | | |
|-----------|---------|-----------------------|
| β_4 | -0.7249 | -0.830300769,-0.61959 |
| β_5 | -0.0295 | -0.040567933,0.001381 |

表 14 自建点 O3 差异与气象因素的回归系数

| | 回归系数 | 置信区间 |
|-----------|----------|-----------------------|
| β_0 | -922.071 | -1202.831028,-641.311 |
| β_1 | -36.9613 | -41.22419272,-32.6985 |
| β_2 | 0.9017 | 0.631409188,1.172067 |
| β_3 | 0.1177 | 0.106089499,0.129361 |
| β_4 | -0.497 | -0.819292762,-0.17464 |
| β_5 | 0.35248 | 0.288897753,0.420716 |

结论：虽然有些回归优度不高，是由于气象数据离散程度较大，不确定因素较多，但其回归系数能反映出该因素对因变量的影响程度，有以上列表数据可知，造成自建点数据 PM2.5 和 PM10 的差异的因素是风速，其次是湿度，然后是压强；造成自建点数据 CO 的差异的气象因素较弱，主要是风速；造成自建点 NO2 的差异的主要是风速，其次是温度、压强和湿度，降雨量的影响较小；对自建点数据 SO2 差异的主要是温度和压强，不过是负相关影响；对自建点数据 O3 差异的主要影响因子是风速，其次是压强和温度，降雨量影响最小。

4.3 问题三 利用国控点数据，对自建点数据校准

4.3.1 建模与求解

根据问题一的数据探索和问题二的异影响因素分析，以国控数据为准确，建立校准模型。

首先，建立零点漂移和量程漂移校准模型。可能因为各种原因，如大自然现象，人为控制污染源等，空气中的“两尘四气”的含量长时段内不会明显呈现出某种规律，但在小时段内，如一日（24 小时）内，其变化不会很大，数据的均值和标准差相对比较稳定，因此以先将自建点数据 0—1 标准化，

$$x_i' = \frac{x_i - \bar{x}_k}{\sigma_k}, (i \in k)$$

再将标准化后的数据按照国控点数据的零点和量程进行校正，模型如下

$$\hat{x}_i = a_k + b_k x_i'$$

其中， a_k 为零点校正参数， b_k 量程校正参数，分别为

$$\begin{cases} a_k = y_k \\ b_k = \frac{y_k}{\bar{x}_k} \sigma_k \end{cases}$$

代入可得校正模型为

$$\hat{x}_i = y_k + \frac{y_k}{\bar{x}_k} \sigma_k \left(\frac{x_i - \bar{x}_k}{\sigma_k} \right), (i \in k)$$

代入数据，用 MATLAB 编程可算出校正结果如下表（表 7，只列出部分数据，完整数据见附件 2）：

表 7 基于国控点数据对自建点数据的校正结果表（部门数据，完整数据见附录 2）

| 时 间 | PM2.5 | PM10 | CO | NO2 | SO2 | O3 |
|-----------------------|----------|----------|----------|----------|----------|----------|
| '2018-11-14 10:02:00' | 35.36977 | 74.71678 | 0.851831 | 9.884854 | 24.79339 | 69.43396 |
| '2018-11-14 10:06:00' | 36.07717 | 78.14765 | 0.745352 | 9.087688 | 25.61983 | 78.49057 |
| '2018-11-14 10:09:00' | 34.66238 | 73.19195 | 0.745352 | 9.247121 | 24.79339 | 73.96226 |
| '2018-11-14 10:10:00' | 34.66238 | 71.66711 | 0.745352 | 9.406554 | 24.79339 | 75.4717 |
| '2018-11-14 10:14:00' | 33.95498 | 73.95436 | 0.745352 | 9.884854 | 24.79339 | 70.9434 |
| '2018-11-14 10:16:00' | 33.95498 | 73.95436 | 0.745352 | 9.247121 | 24.79339 | 75.4717 |
| '2018-11-14 10:18:00' | 33.95498 | 73.95436 | 0.745352 | 9.406554 | 24.79339 | 76.98113 |

| | | | | | | |
|-----------------------|----------|----------|----------|----------|----------|----------|
| '2018-11-14 10:19:00' | 36.78457 | 74.71678 | 0.745352 | 9.565988 | 26.44628 | 75.4717 |
| '2018-11-14 10:23:00' | 33.95498 | 71.66711 | 0.851831 | 9.247121 | 24.79339 | 78.49057 |
| '2018-11-14 10:24:00' | 33.95498 | 70.9047 | 0.745352 | 9.087688 | 24.79339 | 81.50943 |
| '2018-11-14 10:29:00' | 34.66238 | 73.95436 | 0.745352 | 8.928255 | 24.79339 | 80 |
| '2018-11-14 10:34:00' | 31.1254 | 67.85503 | 0.745352 | 9.087688 | 24.79339 | 75.4717 |
| '2018-11-14 10:39:00' | 33.24759 | 71.66711 | 0.745352 | 8.449956 | 24.79339 | 84.5283 |
| '2018-11-14 10:42:00' | 29.71061 | 64.80537 | 0.745352 | 8.768822 | 24.79339 | 86.03774 |
| '2018-11-14 10:43:00' | 31.1254 | 67.85503 | 0.745352 | 8.609389 | 24.79339 | 83.01887 |
| '2018-11-14 10:47:00' | 31.1254 | 70.14228 | 0.745352 | 8.768822 | 24.79339 | 83.01887 |
| '2018-11-14 10:48:00' | 29.00322 | 64.80537 | 0.745352 | 8.609389 | 24.79339 | 86.03774 |
| '2018-11-14 10:52:00' | 32.54019 | 69.37987 | 0.745352 | 7.971656 | 24.79339 | 87.54717 |
| '2018-11-14 10:53:00' | 29.71061 | 66.3302 | 0.745352 | 7.971656 | 24.79339 | 89.0566 |
| '2018-11-14 10:57:00' | 30.41801 | 66.3302 | 0.745352 | 8.768822 | 26.44628 | 89.0566 |
| '2018-11-14 11:02:00' | 32.16516 | 67.50858 | 0.689352 | 8.92849 | 21.54639 | 77.1131 |
| '2018-11-14 11:07:00' | 31.38065 | 69.17546 | 0.787831 | 10.09308 | 21.54639 | 78.44264 |
| '2018-11-14 11:08:00' | 31.38065 | 67.50858 | 0.689352 | 10.28717 | 21.54639 | 77.1131 |
| '2018-11-14 11:11:00' | 32.16516 | 69.17546 | 0.787831 | 8.92849 | 22.98282 | 79.77217 |
| '2018-11-14 11:12:00' | 31.38065 | 65.8417 | 0.787831 | 8.734393 | 21.54639 | 85.09032 |
| '2018-11-14 11:16:00' | 31.38065 | 65.8417 | 0.787831 | 9.704881 | 21.54639 | 82.43124 |
| '2018-11-14 11:17:00' | 30.59613 | 65.8417 | 0.689352 | 10.28717 | 21.54639 | 78.44264 |
| '2018-11-14 11:20:00' | 31.38065 | 65.8417 | 0.689352 | 9.704881 | 22.98282 | 83.76078 |
| '2018-11-14 11:23:00' | 31.38065 | 69.17546 | 0.689352 | 8.734393 | 21.54639 | 82.43124 |

| | | | | | | |
|-----------------------|----------|---------|----------|----------|----------|----------|
| '2018-11-14 11:25:00' | 32.94968 | 70.0089 | 0.689352 | 8.346198 | 22.98282 | 90.40846 |
| '2018-11-14 11:28:00' | 32.16516 | 70.0089 | 0.787831 | 8.346198 | 21.54639 | 89.07893 |
| ... | ... | ... | ... | ... | ... | ... |

用全部数据画出国控数据和自建点数据图，直观比较校准效果：

首先，在同一坐标内画出国控点数据图（蓝色线条）和自建点数据校正图（红色线条）（图），从图中不同进行比较发现，很大部分的重叠在一起，只有极少数的点没有重叠。为了进行直观对比，分开画出对比图（如图所示）

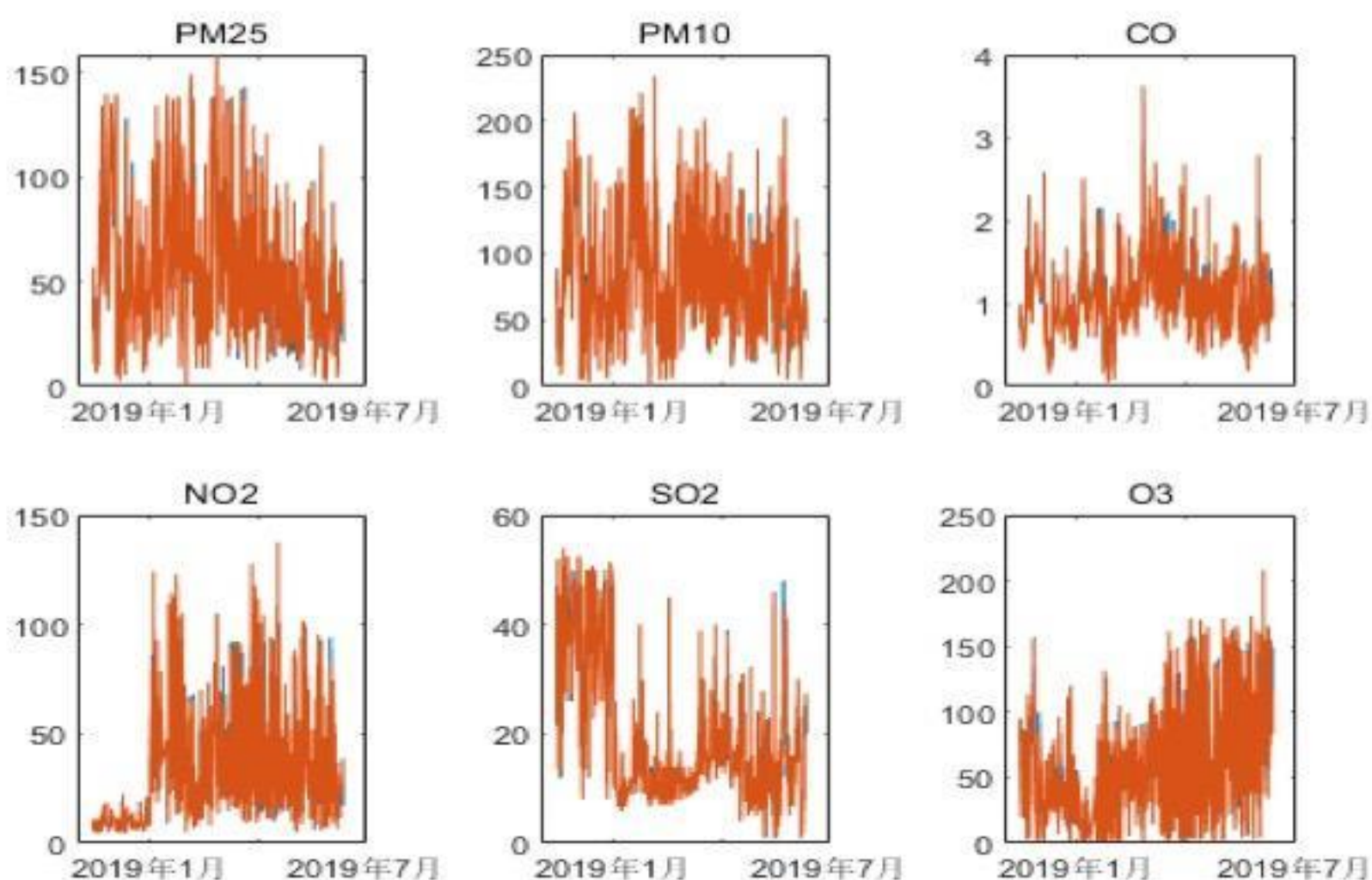


图9 同一坐标六因素国控点数据（蓝色线条）和自建点数据校正（红色线条）图

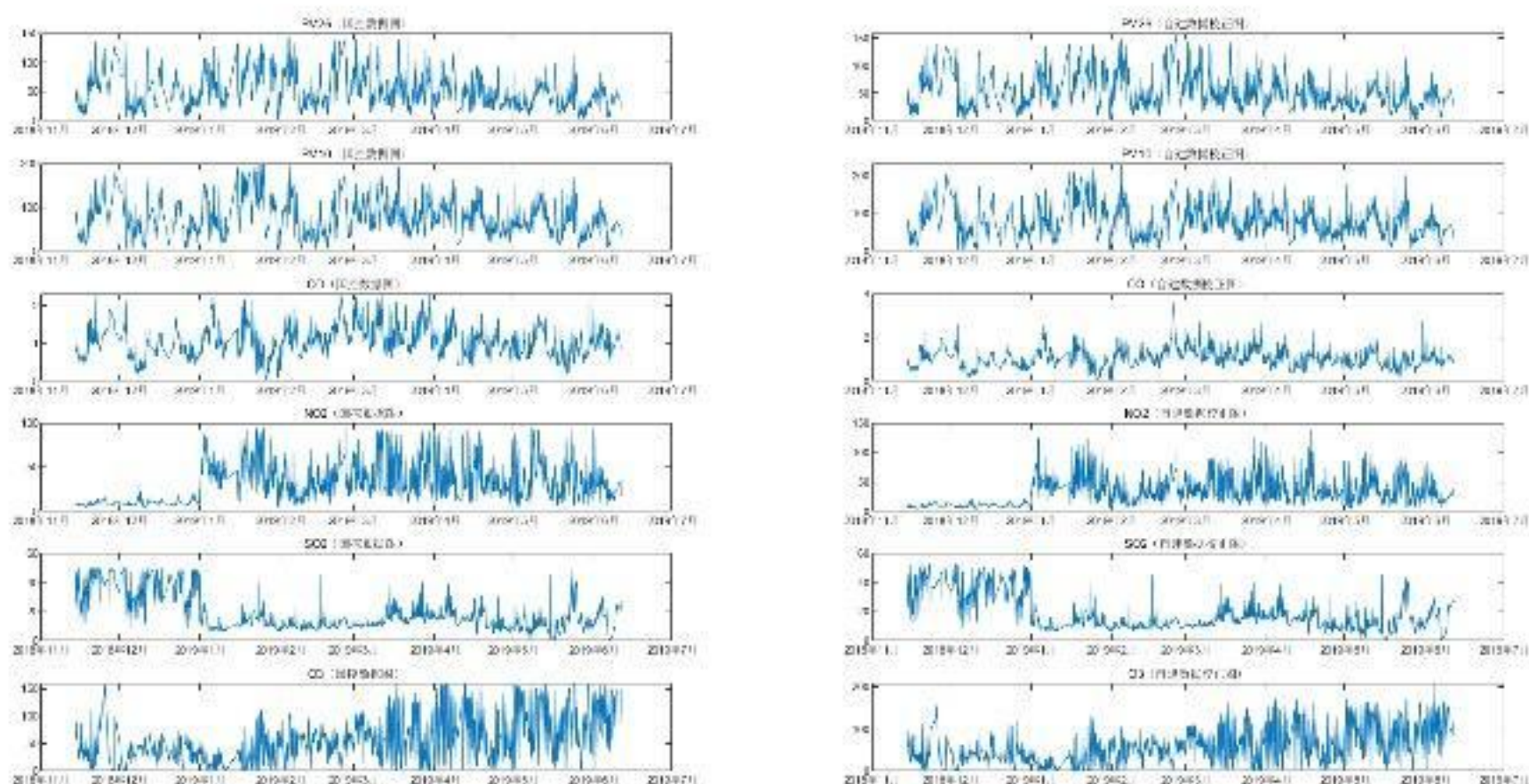


图 10 六因素国控数据（左边）与自建点校正数据（右边）对比图

4.3.2 误差分析

该模型将自建点所有的数据进行了校正，为了检验校准效果，需要用国控点数据与之进行对比，国控点是按小时给数据，为了进行检验，需要找到国控点时刻对应的自建点的时刻，用线性插值将自建点的整点时刻插值得到和国控点同一时刻的数值，数据见附录，然后进行误差分析，这里主要采用两种误差分析。

绝对误差：

$$w_{\text{绝}} = |\hat{x}_k - y_k|$$

相对误差：

$$w_{\text{相}} = \frac{|\hat{x}_k - y_k|}{y_k} \times 100\%$$

用 MATLAB 计算可得结果如下：

表 8 误差分析结果表

| | PM2.5 | PM10 | CO | NO2 | SO2 | O3 |
|------|--------|--------|--------|--------|--------|--------|
| 绝对误差 | 3.6848 | 5.7515 | 0.0548 | 4.0597 | 0.6950 | 4.9447 |

| | | | | | | |
|-----------|--------|--------|--------|---------|--------|---------|
| 相对误差(%) | 8.9343 | 8.6312 | 5.1324 | 10.8094 | 4.2573 | 10.1987 |
|-----------|--------|--------|--------|---------|--------|---------|

从误差可以看出校准效果非常好

5 模型评价

5.2 模型优缺点

模型优点：本文建立了通俗易懂的模型，特别是对问题三中对自建点数据的校准，误差很小

模型缺点：校准模型的校准效果好，但仅仅只用到了零点漂移和量程漂移，把其他交叉影响和气象影响都含进了读取的数据中心，校正模型中如果能包含交叉影响和气象数据的影响则更完美，如把模型改成

$$\hat{x}_i = y_k + \frac{y_k}{\bar{x}_k} \sigma_k \left(\frac{x_i - \bar{x}_k}{\sigma_k} \right) + \varepsilon(Z), (i \in k)$$

其中 $\varepsilon(Z)$ 是交叉影响和气象影响因素，由于时间有限，暂时不能在此做详细研究。

6 参考文献

- [1] 姜启源，大学数学实验，[M]，清华大学出版社。
- [2] MathWorks 中国.MATLAB Docu /.
- [6] 王正林，精通 Matlab7，[M]，电子工业出版社
- [7] 网址 <https://baike.baidu.com/item/%E8%B4%AD%E4%B9%B0%E5%8A%9B/794428?fr=Aladdin>
- [8] 张良均，数据挖掘：实用案例分析 [M].北京：机械工业出版社，2013.
- [9] 电化学气体传感器 <https://baike.baidu.com/item/电化学气体传感器/12605741>

附录 1

数据清除 25

导入数据 25

数据预处理 26

计算平均数、方差、中位数 27

小时 27

日 28

1.数据探索 28

2.分析差异 31

3.建立模型 32

数据清除

```
clear;clc
```

导入数据

```
[~,~,Country_data,Country_Time] = xlsread('CUMCM2019Problems_D_Country_data.xlsx',...  
  
'CUMCM2019Problems_D_Country_dat','A2:G4201','',@convertSpreadsheetExcelDates);  
[~,~,Self_data,Self_Time] = xlsread('CUMCM2019Problems_D_Self_data.xlsx',...  
  
'CUMCM2019Problems_D_Self_data','A2:L234718','',@convertSpreadsheetExcelDates);  
Country_data = double(string(Country_data(:,1:end-1)));  
Country_Time = datetime([Country_Time(:,end)].','ConvertFrom','Excel');  
Self_data = double(string(Self_data(:,1:end-1)));
```

```
Self_Time = datetime([Self_Time{:,end}].', 'ConvertFrom', 'Excel');
```

数据预处理

```
bt = string({"PM25";"PM10";"CO";"NO2";"SO2";"O3";"风速";"压强";"降水量";"温度";"湿度"});

figure

for i = 1:6

    subplot(2,3,i)

    plot(Country_Time,Country_data(:,i),'ro',Self_Time,Self_data(:,i))

    title(bt(i))

end

% 自建点相同时间取平均数

a = tabulate(Self_Time);

xtsj_dian = datetime(a([a{:,2}] >= 2,1));

for i = 1:size(xtsj_dian,1)

    wz = find(Self_Time == xtsj_dian(i),inf,'last');

    Self_Time(wz(2:end)) = [];

    Self_data(wz(1),:) = mean(Self_data(wz,:),1);

    Self_data(wz(2:end),:) = [];

end

figure

boxplot(Country_data)

title("国控点")

figure

boxplot(Self_data)

title("自建点")

[yc_data,yc_time,spot,data1,time1] = MC2(Country_data,Country_Time);

Country_data = data1;Country_Time = time1;
```



```

[yc_data1,yc_time1,spot1,data11,time11] = MC2(Self_data,Self_Time);

Self_data = data11;Self_Time = time11;

Hour = (datetime('2018-11-14 10:00:00'):duration('01:00:00'):datetime('2019-6-11 17:00:00'));

xc = [];ma = [];

for i = 1:size(Country_Time,1)

    list = find(Self_Time >= Hour(Hour == Country_Time(i)) & Self_Time < Hour(find((Hour == Country_Time(i)),inf,'last')+1),inf,'last');

    if ~isnan(list)

        ma = [ma list];

    else

        xc = [xc i];

    end

end

queshi = setdiff(1:size(Self_Time,1),ma);

Country_data_tichu = Country_data(xc,:);Country_Time_tichu = Country_Time(xc,:);

Country_data(xc,:) = [];Country_Time(xc,:) = [];

Self_data_tichu = Self_data(queshi,:);Self_Time_tichu = Self_data(queshi,:);

Self_data(queshi,:) = [];Self_Time(queshi,:) = [];

% 清除临时变量

clearvars ma list i xlh xc queshi data1 data11 time1 time11;

```

计算平均数、方差、中位数

小时

```

[Self_Mean_Hour,Self_Variance_Hour,Self_Median_Hour,time]=MC(Self_data,Self_Time,Country_Time,

```

```
duration('01:00:00')); % 自建点
```

日

```
Day = (datetime('2018-11-14 10:00:00'):duration('24:00:00'):datetime('2019-6-11 17:00:00'))';  
[Country_Mean_day, Country_Variance_day, Country_Median_day, time] =  
MC(Country_data, Country_Time, Day, duration('24:00:00')); % 国控点  
[Self_Mean_day, Self_Variance_day, Self_Median_day, Day] =  
MC(Self_data, Self_Time, Day, duration('24:00:00')); % 自建点
```

1. 数据探索

```
String_1 = string({"国控点检测物含量的整体分布散点图(小时)"; "国控点检测物含量的整体分布散点图(天)"; ...  
    "自建点检测物含量的整体分布散点图(小时)"; "自建点检测物含量的整体分布散点图(天)"; ...  
    "国控点检测物含量的整体分布散点图(天)中位数"; ...  
    "自建点检测物含量的整体分布散点图(小时)中位数"; "自建点检测物含量的整体分布散点图(天)中位数"  
    "; ...  
    "国控点检测物含量的整体分布散点图(天)标准差"; ...  
    "自建点检测物含量的整体分布散点图(小时)标准差"; "自建点检测物含量的整体分布散点图(天)标准差"  
    "; ...  
    "自建点检测物含量的整体分布散点图(5分钟内)"; "自建点气象的整体分布散点图(5分钟内)"; ...  
    "自建点气象的整体分布散点图(小时)均值"; "自建点气象的整体分布散点图(天)均值"; ...  
    "自建点气象的整体分布散点图(小时)中位数"; "自建点气象的整体分布散点图(天)中位数"; ...  
    "自建点气象的整体分布散点图(小时)标准差"; "自建点气象的整体分布散点图(天)标准差" ...  
    });  
  
String_2 = string({"国控点检测物含量关联图(小时)"; "国控点检测物含量关联图(天)"; ...  
    "自建点检测物含量关联图(小时)"; "国控点检测物含量关联图(天)"; ...  
    "国控点检测物含量关联图(天)中位数"; ...  
    "自建点检测物含量关联图(小时)中位数"; "国控点检测物含量关联图(天)中位数"; ...
```



```

        "国控点检测物含量关联图(天)标准差";...

        "自建点检测物含量关联图(小时)标准差";"国控点检测物含量关联图(天)标准差";...

        "自建点检测物含量关联图(5 分钟内)";"自建点气象关联图(5 分钟内)";...

        "自建点气象关联图(小时)均值";"自建点气象关联图(天)均值";...

        "自建点气象关联图(小时)中位数";"自建点气象关联图(天)中位数";...

        "自建点气象关联图(小时)标准差";"自建点气象关联图(天)标准差"...

    });

Cell_1 = {Country_Time;Day;Country_Time;Day;...

    Day;Country_Time;Day;...

    Day;Country_Time;Day;...

    Self_Time;Self_Time;...

    Country_Time;Day;...

    Country_Time;Day;...

    Country_Time;Day...

    };

Cell_2 = {Country_data;Country_Mean_day;Self_Mean_Hour;Self_Mean_day;...

    Country_Median_day;Self_Median_Hour;Self_Median_day;...

    Country_Variance_day;Self_Variance_Hour;Self_Variance_day;...

    Self_data(:,1:6);Self_data(:,7:end);...

    Self_Mean_Hour(:,7:end);Self_Mean_day(:,7:end);...

    Self_Variance_Hour(:,7:end);Self_Variance_day(:,7:end);...

    Self_Median_Hour(:,7:end);Self_Median_day(:,7:end)...

    };

c = cell(1,18);

for j = 1:18

    if j >= 12

        figure

        for i = 1:5

```

```

        subplot(2,3,i)

        plot(Cell_1{j},Cell_2{j}(:,i)) %作图

        title(bt(i+6))

    end

    subtitle(String_1(j))

else

    figure

    for i = 1:6

        subplot(2,3,i)

        plot(Cell_1{j},Cell_2{j}(:,i)) %作图

        title(bt(i))

    end

    subtitle(String_1(j))

end

figure

plotmatrix(Cell_2{j}) %关联图

title(String_2(j))

c{j}=corrcoef(Cell_2{j}); %求得对应的相关系数

end

% 自建点数据与国控点数据分析

figure

plotmatrix(Country_data,Self_Mean_Hour(:,1:6))%国与自的相关强度分析/时

title('自建点检测物含量与国控点检测物含量关联图(小时)')

figure

plotmatrix(Country_Mean_day(:,1:6),Self_Mean_day(:,1:6))%国与自的相关强度分析/天

title('自建点检测物含量与国控点检测物含量关联图(天)')

% 清除临时变量

clearvars i j bt String_1 String_2 Cell_1 Cell_2 Hour;

```


2.分析差异

```
a1 = Self_Mean_Hour(:,1:6) - Country_data;

a2=abs(a1)./Country_data;

a3=Self_Variance_day(:,1:6)-Country_Variance_day;

a3=abs(a3)./Country_Variance_day;

pj_wc = mean(abs(a1));bt = string({'PM25';'PM10';'CO';'NO2';'SO2';'O3'});

figure

for i =1:6

    subplot(2,3,i)

    plot(Country_Time,a1(:,i))

    title(bt(i))

end

suptitle("国控与自建的误差图")


a4 = Self_Mean_Hour(:,3:6);

X = [Self_Mean_Hour a4(:,1).*a4(:,2) a4(:,1).*a4(:,3)...

    a4(:,1).*a4(:,4) a4(:,2).*a4(:,3) a4(:,2).*a4(:,4) a4(:,3).*a4(:,4)];

rew1 = cell(6,17);rew2 = ones(6,17);rew3 = ones(6,17);rew4 = ones(6,17);rew = cell(6,17);

for i =1:6

    for j =1:17

        [b, bint,r,rint,stats] = regress(a1(:,i),[ones(size(Country_data,1),1) X(:,j)]);

        rew1{i,j} = b;

        rew2(i,j) = stats(1);rew3(i,j)= stats(2);rew4(i,j)= stats(3);

        r2 = corrcoef(a1(:,i),X(:,j));

        rew{i,j} = r2(2);
```

```

    end

end

sj_data1 = cell(4,1);sj_data3 = ones(4,1);sj_data4 = ones(4,1);sj_data5 = ones(4,1);sj_data =
cell(4,1);
for i = 1:4
    mac = Self_Mean_Hour(:,3:6);
    mac(:,i) = [];
    [b, bint,r,rint,stats] = regress(Self_Mean_Hour(:,i+2),[ones(size(Self_Mean_Hour,1),1)
mac ...
                                mac(:,1).*mac(:,3) mac(:,2).*mac(:,3)]];
    stepwise( Self_Mean_Hour(:,7:end),a1(:,i));
    sj_data1{i,1} = b;
    sj_data3(i) = stats(1);
    sj_data4(i) = stats(2);
    sj_data5(i) = stats(3);
    sj_data{1,i} = bint;
end

% 清除临时变量

clearvars i j a4 X b bint r rint stats r2

```

3.建立模型

```

[y,x1,x2] = MC11(Self_data(:,1:6),Self_Mean_Hour(:,1:6),Country_data,Country_Time,Self_Time);
y111 = interp1(Self_Time,y,Country_Time(2:end),'nearest');
wcz = Country_data(2:end,:) - y111;
jdwc = mean(abs(wcz));
xdwc = mean(abs(wcz) ./ Country_data(2:end,:)) .* 100

```



```

rwell = cell(2,6);rwe22 = cell(2,6);

for i = 1:6

    stepwise( Self_Mean_Hour(:,7:end),a1(:,i));

end

for i = 1:6

[b,      bint,r,rint,stats]      =      regress(a1(:,i),[ones(size(Country_data,1),1)
Self_Mean_Hour(:,7:end))]);

rwell{1,i} = b;rwell{2,i} = stats;

[b,      bint,r,rint,stats]      =      regress(Self_Mean_Hour(:,i),[ones(size(Country_data,1),1)
Self_Mean_Hour(:,7:end))]);

rwe22{1,i} = b;rwe22{2,i} = stats;

end

ace = ones(size(Self_data(:,1:6)));ace1 = ones(size(Self_data(:,1:6)));

for i = 1:6

    ym = rwe22{1,i}(1) + rwe22{1,i}(2).*Self_data(:,7) + rwe22{1,i}(3).*Self_data(:,8) +
rwe22{1,i}(4).*Self_data(:,9) + ...

        rwe22{1,i}(5).*Self_data(:,10) + rwe22{1,i}(6).*Self_data(:,10);

    ace1(:,i) = ym;

    ace(:,i) = Self_data(:,i) - ym;

end

mean(abs(ace))

[y__1,x__1,x__2] = MC11(ace1,Self_Mean_Hour(:,1:6),Country_data,Country_Time,Self_Time);

chazhi = interp1(Self_Time,y__1,Country_Time(2:end),'nearest');

mean(abs(Country_data(2:end,:) - chazhi) ./ Country_data(2:end,:)) .*100;%

```

```

bt1 = string({"PM25 ( 国控数据图 ) "; "PM10 ( 国控数据图 ) "; "CO ( 国控数据图 ) "; "NO2 ( 国控数据图 ) "; "SO2 ( 国控数据图 ) "; "O3 ( 国控数据图 ) "});

bt2 = string({"PM25 ( 自建数据校正图 ) "; "PM10 ( 自建数据校正图 ) "; "CO ( 自建数据校正图 ) "; "NO2 ( 自建数据校正图 ) "; "SO2 ( 自建数据校正图 ) "; "O3 ( 自建数据校正图 ) "});

figure

for i = 1:6

    figure

    subplot(2,1,1)

    plot(Country_Time(2:end,:),Country_data(2:end,i))

    title(bt1(i))

    subplot(2,1,2)

    plot(Country_Time(2:end,:),y111(:,i))

    title(bt2(i))

end

```

Published with MATLAB® R2018a

子程序1

```

function [Mean,Variance,Median,time]= MC(data,data1,time,a)

    Mean = ones(size(time,1)-1,size(data,2));

    Variance = ones(size(time,1)-1,size(data,2));

    Median = ones(size(time,1)-1,size(data,2));

    for i = 1:size(time,1)

        x1h = find(data1 >= time(i) & data1 < time(i) + a,inf,'last');

        Mean(i,:) = mean(data(x1h,:),1); %均值

        Variance(i,:) = std(data(x1h,:),1); % 标准差

        Median(i,:) = median(data(x1h,:),1); % 中位数

    end

    a = find(isnan(Mean(:,1)),inf,'last');

    Mean(a,:) = [];

```



```

        Variance(a,:) = [];

        Median(a,:) = [];

        time(a,:) = [];

    end

```

子程序2

```

function [yc_data,yc_time,spot,data1,time1] = MC2(data,time)

    spot = [];

    for i = 1:size(data,2)

        a = quantile(data(:,i),3);

        b = a(1) - 1.5 .* (a(3) - a(1));

        c = a(3) + 1.5 .* (a(3) - a(1));

        spot = union(spot,find(data(:,i) < b | data(:,i) > c,inf,'last')));

    end

    yc_data = data(spot);

    yc_time = time(spot);

    data1 = data(setdiff(1:size(data,1),spot),:);

    time1 = time(setdiff(1:size(time),spot),:);

end

```

子程序3

```

function [y,x1,x2] = MC11(Self_data,Self_Mean_Hour,Country_data,time,data1)

    y = ones(size(Self_data,1),6);x1 = ones(size(Self_data,1),6);x2 = ones(size(Country_data));

    for i = 1:size(time,1)

        x1h = find(data1 >= time(i) & data1 < time(i) + duration('01:00:00'),inf,'last');

        std = std(Self_data(x1h,:));

        if std(std == 0) == 0

            lis = Self_data(x1h,:);

            std(std == 0) = lis(1,find(std == 0,inf,'last'));

        end

    end

```

```
end

x = (Self_data(xlh,:) - Self_Mean_Hour(i,:)) ./ Std;
bk = (Country_data(i,:) ./ Self_Mean_Hour(i,:)) .* Std;

y(xlh,:) = Country_data(i,:) + bk .* x;
x1(xlh,:) = x;
x2(i,:) = bk;

end

end
```