空气中PM2.5问题的研究

# PM2.5的相关因素分析

1.1原文方法和思路

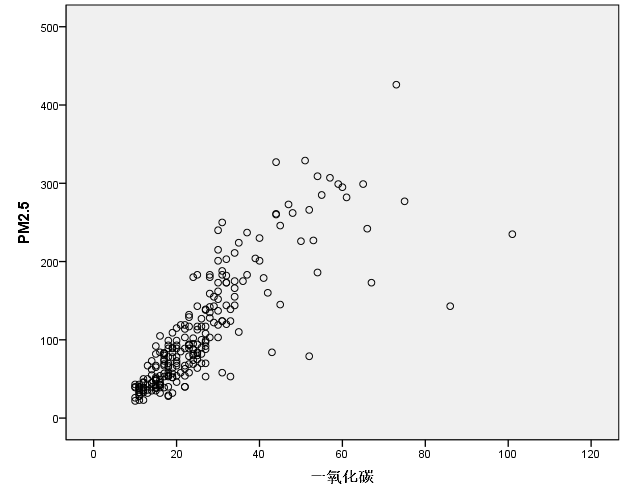
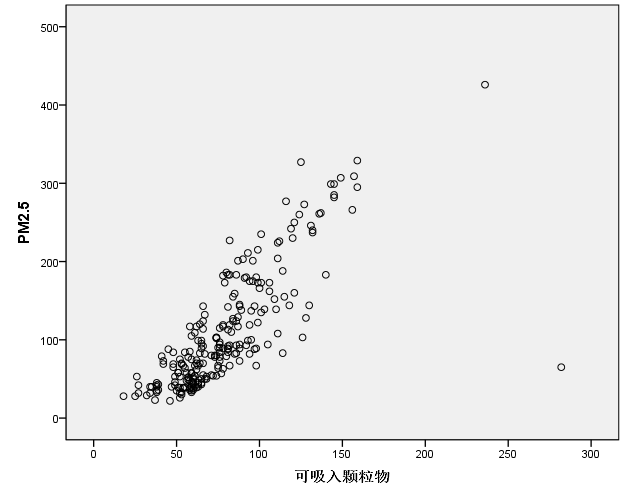
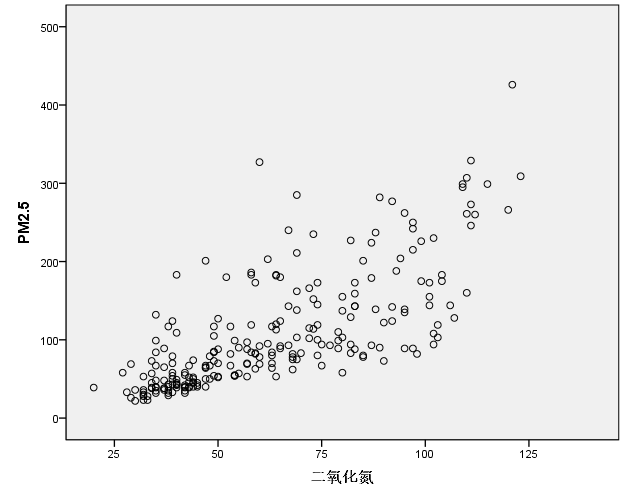
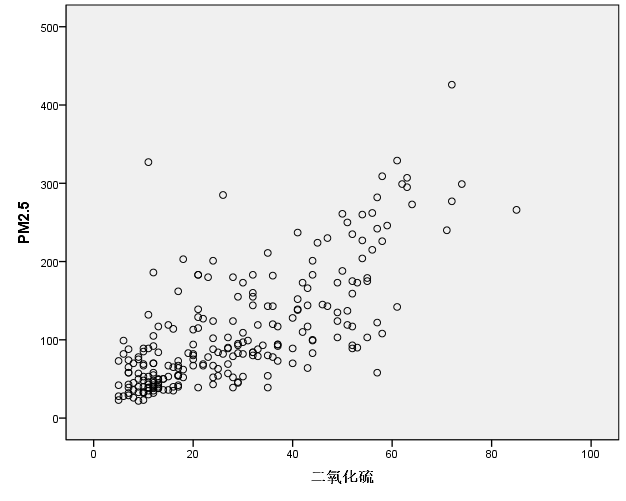
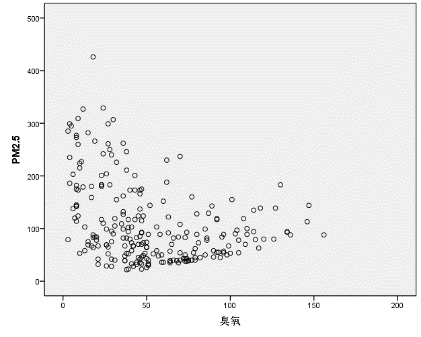
题目要求依据附件1或附件2中的数据或自行采集数据，利用或建立适当的数学模型，对AQI中6个基本监测指标的相关与独立性进行定量分析，尤其是对其中PM2.5（含量）与其它5项分指标及其对应污染物（含量）之间的相关性及其关系进行分析。

南京航空航天大学的同学对此问题的解决思路如下：

1. PM2.5分别与其它5项指标作散点图，分析PM2.5与其它5项指标的相关性；
2. 使用SPSS进行相关性分析，得到6个变量的相关系数；
3. PM2.5与其它5项指标作拟合，最大为0.8，不理想，得出无法用单一变量刻画PM2.5形成的结论；
4. PM2.5分别与其它5项指标作多元线性回归，，结果依旧不理想，所以根据残差去掉异常点，经过多次循环最终得到的多元回归方程。

1.2解题过程重现

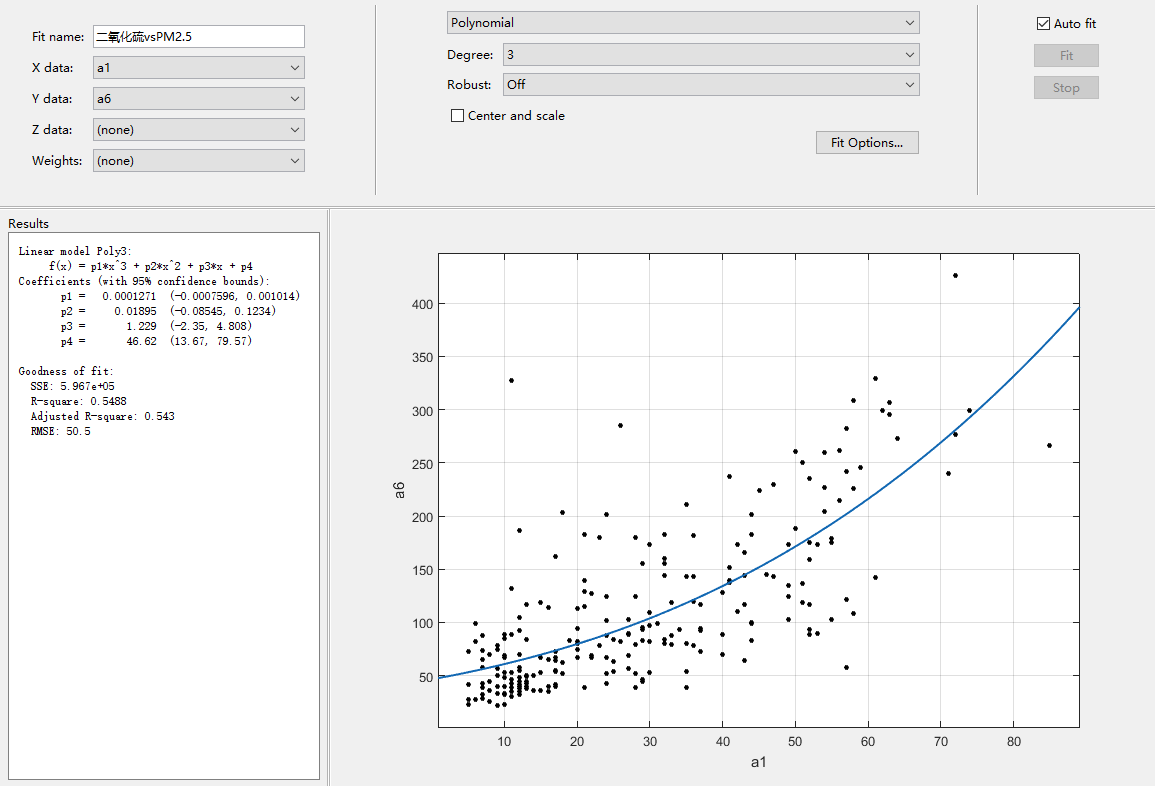
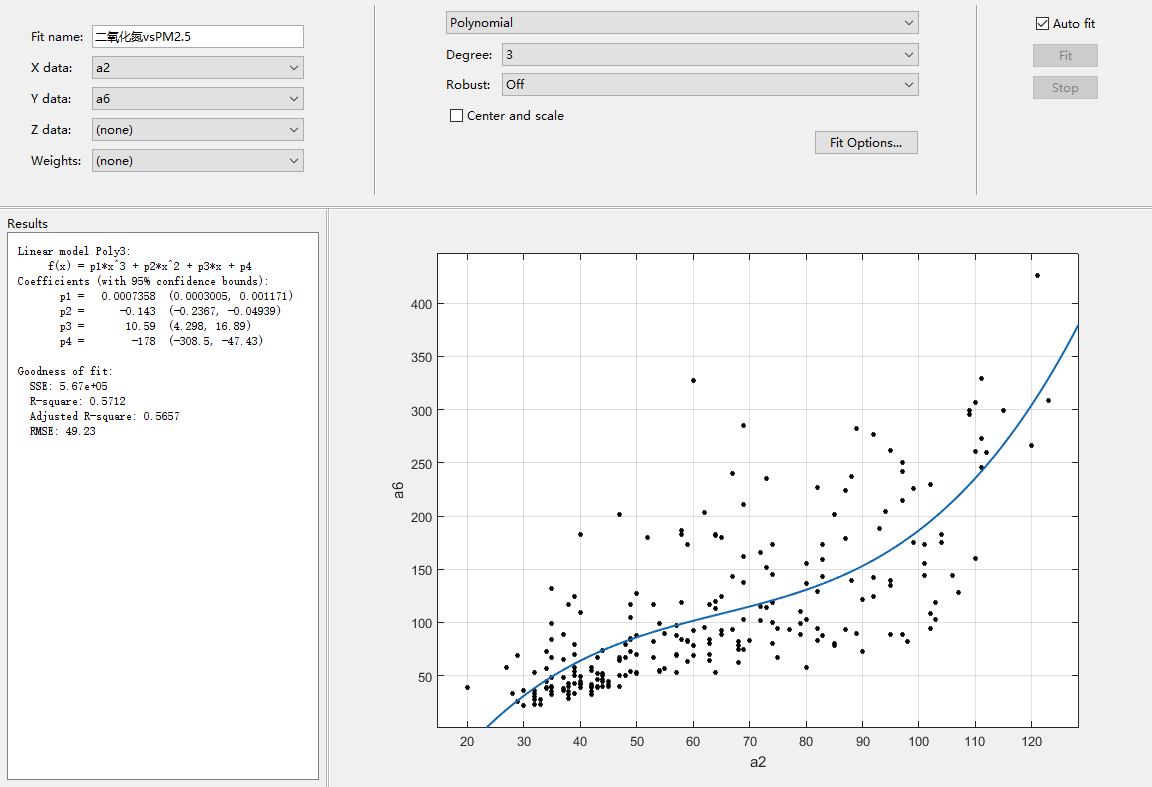
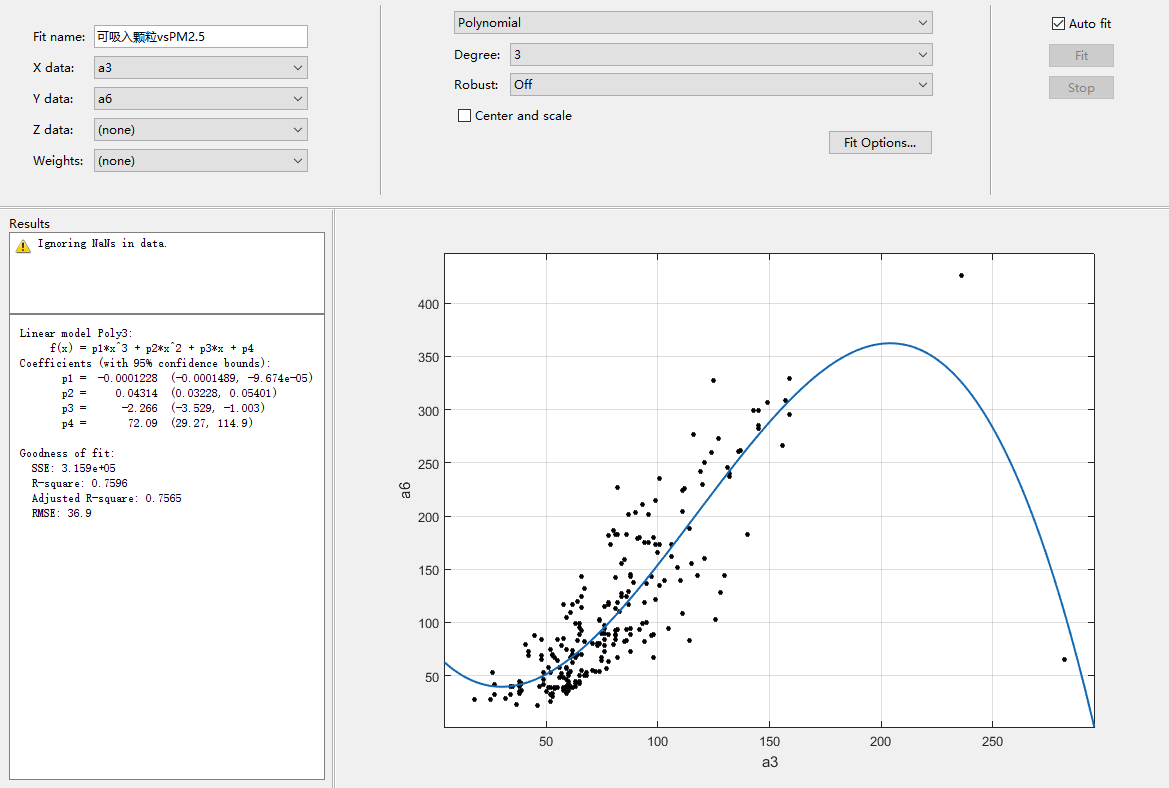
南航使用Matlab来绘制散点图，我将使用SPSS进行散点图的绘制。将数据导入SPSS中，分别作出PM2.5与另外5项指标的散点图，从散点图可以清晰看出PM2.5与二氧化硫、二氧化碳、可吸入颗粒、一氧化碳具有明显的相关性。

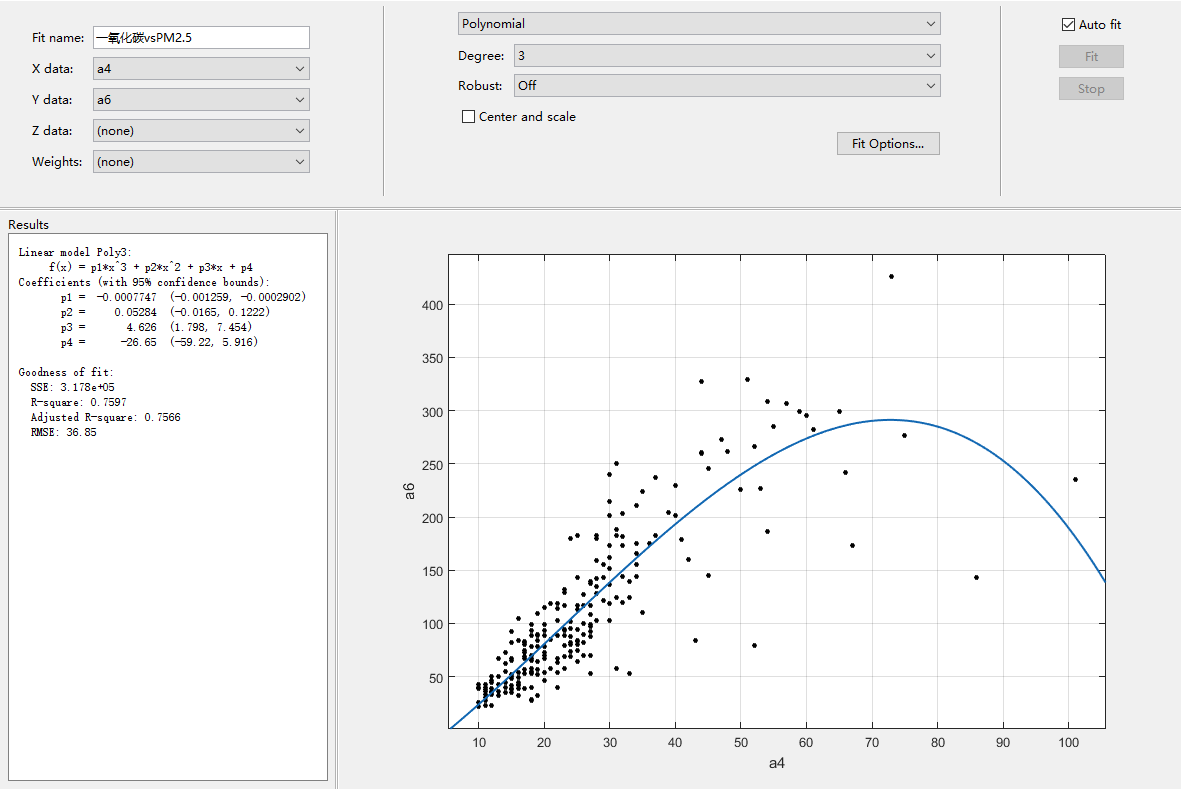
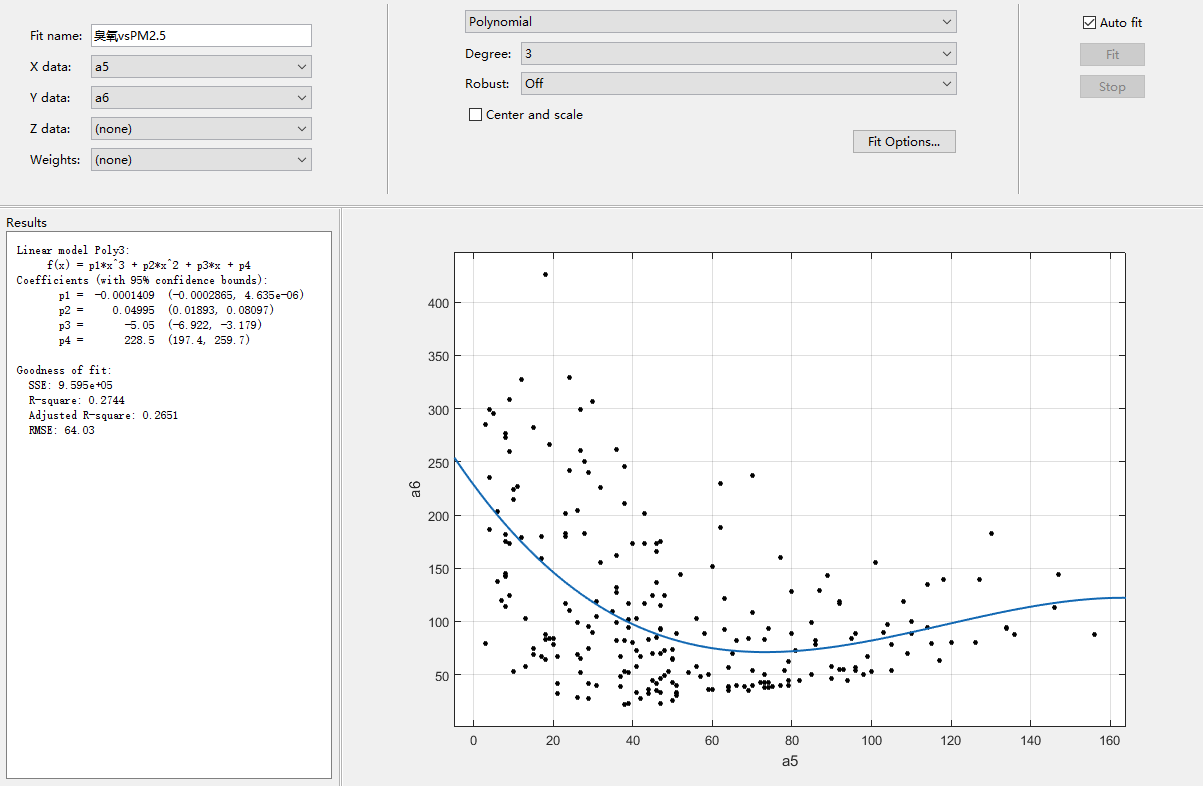


为了验证猜想与事实是否相符，使用SPSS对PM2.5与其它5项指标进行两两相关性分析，得到如下表格从表中，可以看出PM2.5与一氧化碳间相关系数为0.822，呈显著正相关，且相关系数最大；PM2.5与二氧化硫2间相关系数为0.724，呈显著正相关；PM2.5与二氧化氮间相关系数为0.732，呈显著正相关；PM2.5与可吸入颗粒物间相关系数为0.779，也呈显著正相关；PM2.5 与臭氧间相关系数为-0.355，呈负相关。

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **相关性** | | | | | | | |
|  | | 二氧化硫 | 二氧化氮 | 可吸入颗粒物 | 一氧化碳 | 臭氧 | PM2.5 |
| 二氧化硫 | Pearson 相关性 | 1 | .807\*\* | .678\*\* | .659\*\* | -.179\*\* | .726\*\* |
| 显著性 （双尾） |  | .000 | .000 | .000 | .006 | .000 |
| N | 238 | 238 | 236 | 238 | 238 | 238 |
| 二氧化氮 | Pearson 相关性 | .807\*\* | 1 | .727\*\* | .626\*\* | -.063 | .734\*\* |
| 显著性 （双尾） | .000 |  | .000 | .000 | .336 | .000 |
| N | 238 | 238 | 236 | 238 | 238 | 238 |
| 可吸入颗粒物 | Pearson 相关性 | .678\*\* | .727\*\* | 1 | .586\*\* | -.069 | .779\*\* |
| 显著性 （双尾） | .000 | .000 |  | .000 | .295 | .000 |
| N | 236 | 236 | 236 | 236 | 236 | 236 |
| 一氧化碳 | Pearson 相关性 | .659\*\* | .626\*\* | .586\*\* | 1 | -.381\*\* | .822\*\* |
| 显著性 （双尾） | .000 | .000 | .000 |  | .000 | .000 |
| N | 238 | 238 | 236 | 238 | 238 | 238 |
| 臭氧 | Pearson 相关性 | -.179\*\* | -.063 | -.069 | -.381\*\* | 1 | -.352\*\* |
| 显著性 （双尾） | .006 | .336 | .295 | .000 |  | .000 |
| N | 238 | 238 | 236 | 238 | 238 | 238 |
| PM2.5 | Pearson 相关性 | .726\*\* | .734\*\* | .779\*\* | .822\*\* | -.352\*\* | 1 |
| 显著性 （双尾） | .000 | .000 | .000 | .000 | .000 |  |
| N | 238 | 238 | 236 | 238 | 238 | 238 |
| \*\*. 在置信度（双测）为 0.01 时，相关性是显著的。 | | | | | | | |

得到相关性之后还不足以刻画PM2.5与其它5项指标间的关系，而最直接的方法就是用函数来表达。在考虑函数复杂性和实用性的情况下，使用三次多项式来进行拟合。通过使用Matlab提供的函数拟合工具箱cftool可以快速得到想要的结果，并且与原文结果一致，结果如下所示。

由于回归方程的拟合优度越接近1，拟合效果越好，对此通过以上指标对各种模型拟合结果进行对比分析后，发现PM2.5与二氧化硫、二氧化氮、可吸入颗粒物PM10、臭氧和一氧化碳等5项指标间的两两回归模型拟合效果不是很好，其中较好的为PM2.5分别PM10和一氧化碳CO的拟合，而拟合出的三次模型其，为最佳拟合模型。

因此单一变量不足以刻画PM2.5形成的原因，可对PM2.5与其它5项变量进行多元回归分析，得到最终的回归方程。原文使用Matlab进行多元回归，本文依旧使用SPSS进行多元回归，使用SPSS方法更简单且数据结果更直观。过程如下：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **模型摘要b** | | | | |
| 模型 | R | R 平方 | 调整后的 R 平方 | 标准估算的错误 |
| 1 | .986a | .973 | .972 | 12.139 |
| a. 预测变量：（常量），臭氧, 可吸入颗粒物, 二氧化硫, 一氧化碳, 二氧化氮 | | | | |
| b. 因变量：PM2.5 | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **系数a** | | | | | | |
| 模型 | | 非标准化系数 | | 标准系数 | t | 显著性 |
| B | 标准错误 | 贝塔 |
| 1 | （常量） | -47.050 | 3.243 |  | -14.507 | .000 |
| 二氧化硫 | -.174 | .098 | -.043 | -1.777 | .077 |
| 二氧化氮 | .408 | .080 | .136 | 5.114 | .000 |
| 可吸入颗粒物 | 1.268 | .068 | .534 | 18.654 | .000 |
| 一氧化碳 | 2.102 | .145 | .373 | 14.544 | .000 |
| 臭氧 | -.359 | .032 | -.168 | -11.391 | .000 |
| a. 因变量：PM2.5 | | | | | | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **观测值诊断a** | | | | |
| 个案编号 | 标准残差 | PM2.5 | 预测值 | 残差 |
| 10 | 2.001 | 260 | 235.71 | 24.290 |
| 48 | 2.170 | 173 | 146.66 | 26.338 |
| 51 | 2.094 | 143 | 117.58 | 25.422 |
| 71 | 2.053 | 40 | 15.07 | 24.925 |
| a. 因变量：PM2.5 | | | | |

从模型摘要中我们可以得到现在的多元回归方程，是较为理想的一个结果。从系数表格中可以得到多元回归方程的系数值和常量值。观测值诊断则是显示标准残差大于2个标准残差的数据，经过不断循环剔除标准残差大于2个标准残差的数据最终得到上面的多元回归方程。至此第一问的问题基本解决。

1.3点评

南航同学对于第一问的解决，思路十分清晰，从拟合到两两回归分析，再到多元回归方程的建立循序渐进，并用残差检验的方法对结果的合理性进行分析。唯一不足就是篇幅过长，显得过于拖沓，正文总共42页，到第一问的解决篇幅已经快接近一半，显得头重脚轻，散点图和拟合可以合并在一起，事实上在发表的论文中这一部分确实省略了。

# PM2.5的分布与演变及应急处理

2.1问题二第1小问模型的建立与求解

2.1.1 原文方法和思路

该问题要求描述该地区内PM2.5的时空分布及其规律，并结合环境保护部新修订的《环境空气质量标准》分区进行污染评估。

原文思路如下：

（1）绘制PM2.5浓度与时间的散点图，得出分布趋势；

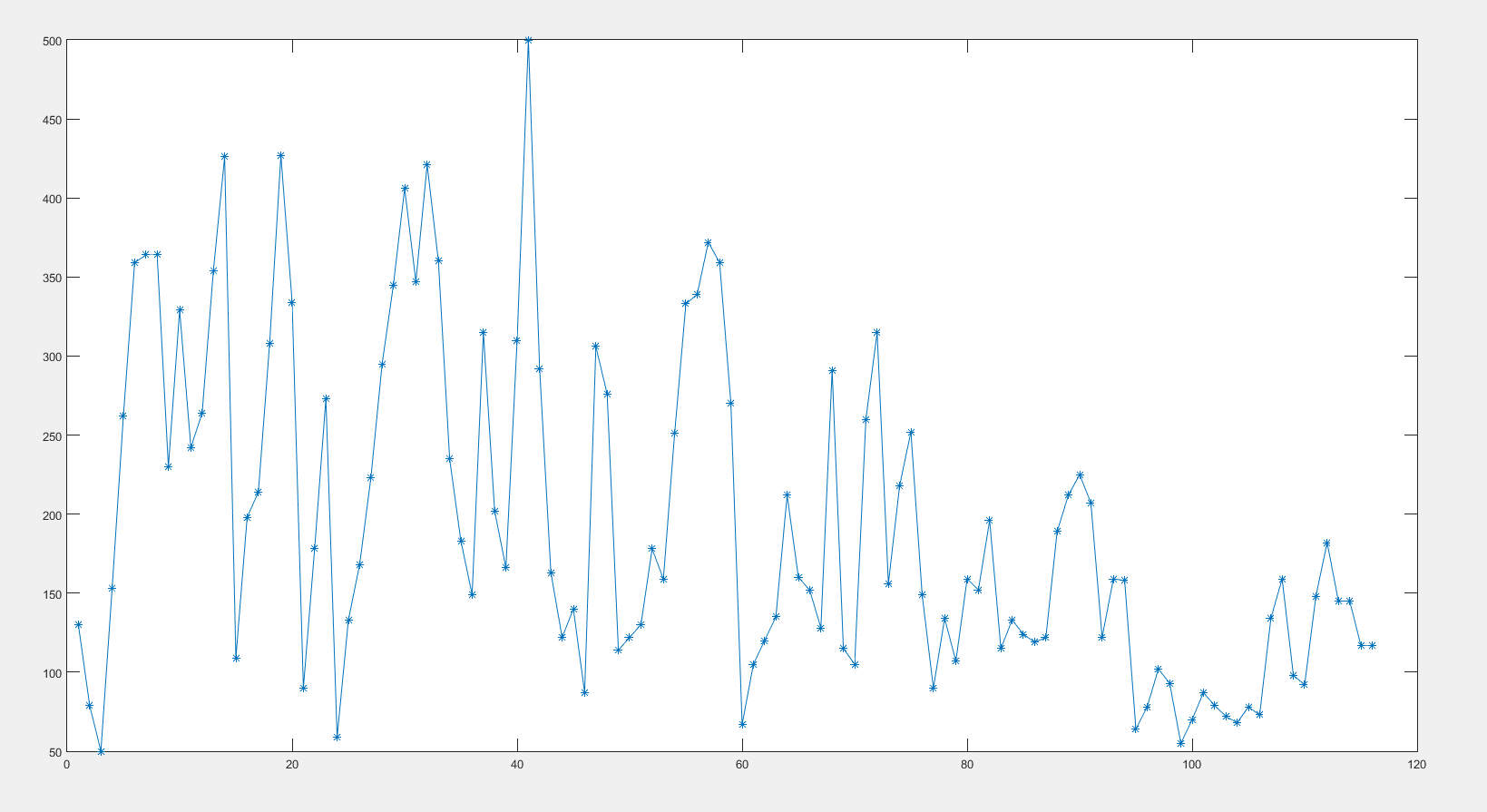
（2）以全市平均浓度代替各个监测点浓度刻画PM2.5时间分布；

（3）使用Shepard反距离插值法刻画PM2.5空间分布；

（4）污染评估。

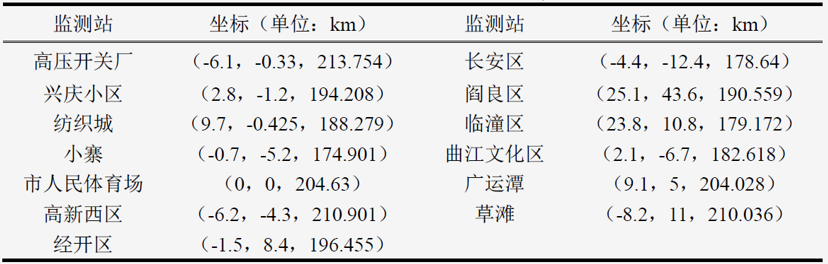
2.1.2 解题过程重现

首先作出13个监测点以及全市平均PM2.5随时间变化的散点图（太简单，此处略去）。从散点图中可以看出，西安市各个监测站PM2.5的含量随时间变化趋势大致相同，即都为随着时间PM2.5含量大致呈减少趋势，而且都与全市PM2.5平均含量大体一致，因此可用全市PM2.5平均含量随时间变化的规律描述各个监测站PM2.5含量的时间分布。进行三次样条插值得到如下全市平均PM2.5随时间变化曲线图。

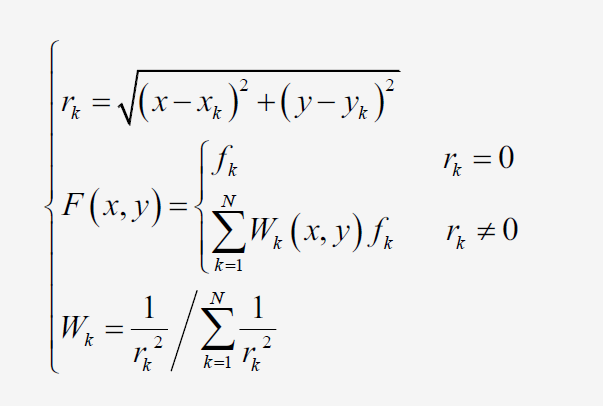


对于西安市PM2.5的空间分布，原文提出以2013年1月1日到2013年4月26日时间段内的平均值作为13个监测点的监测值，通过Shepard反距离插值法刻画西安市PM2.5的空间分布。

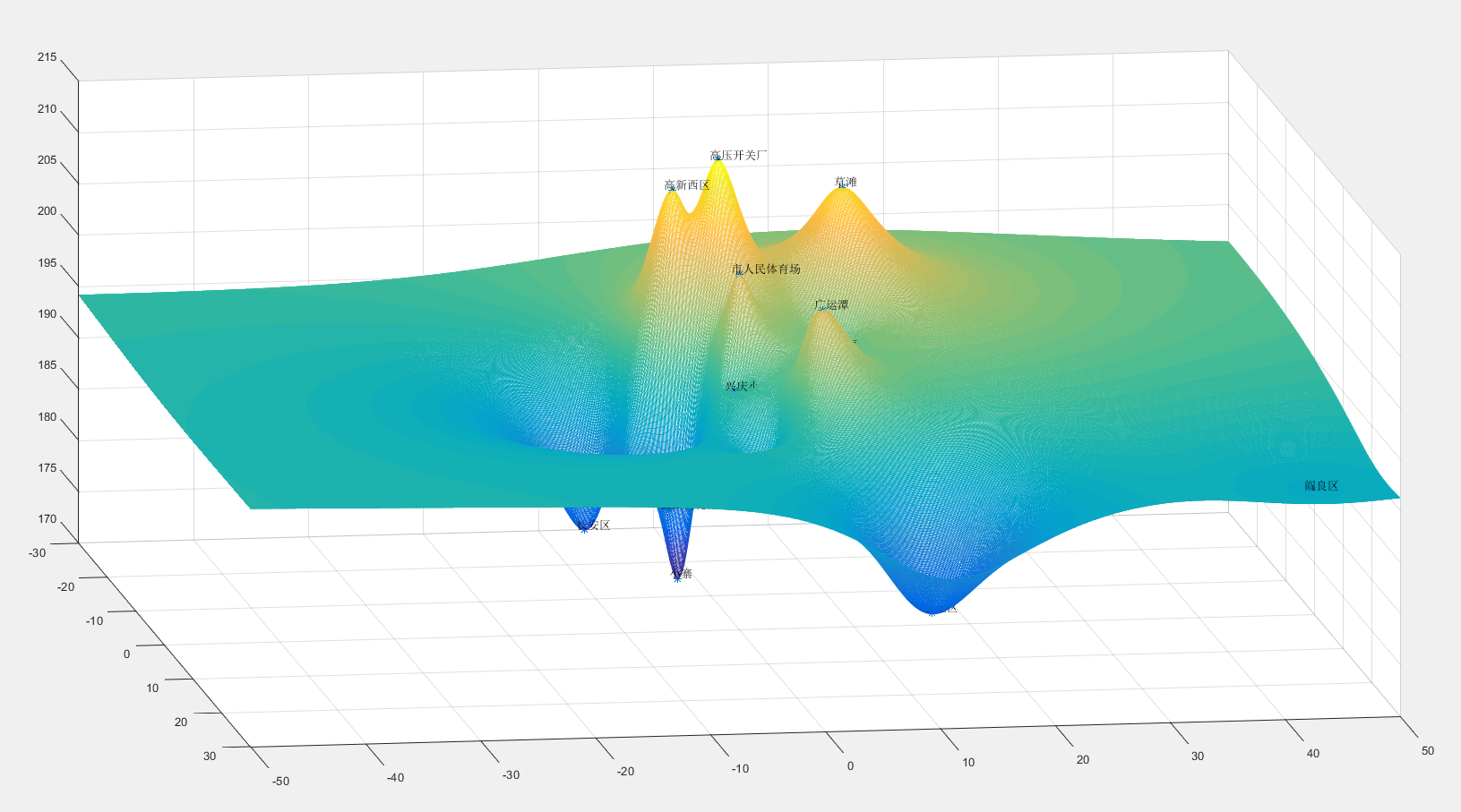
得到13个监测点数据之后以市人民广场为中心，建立平面直角坐标系，通过Google Earth可以得到各个监测点的位置坐标。



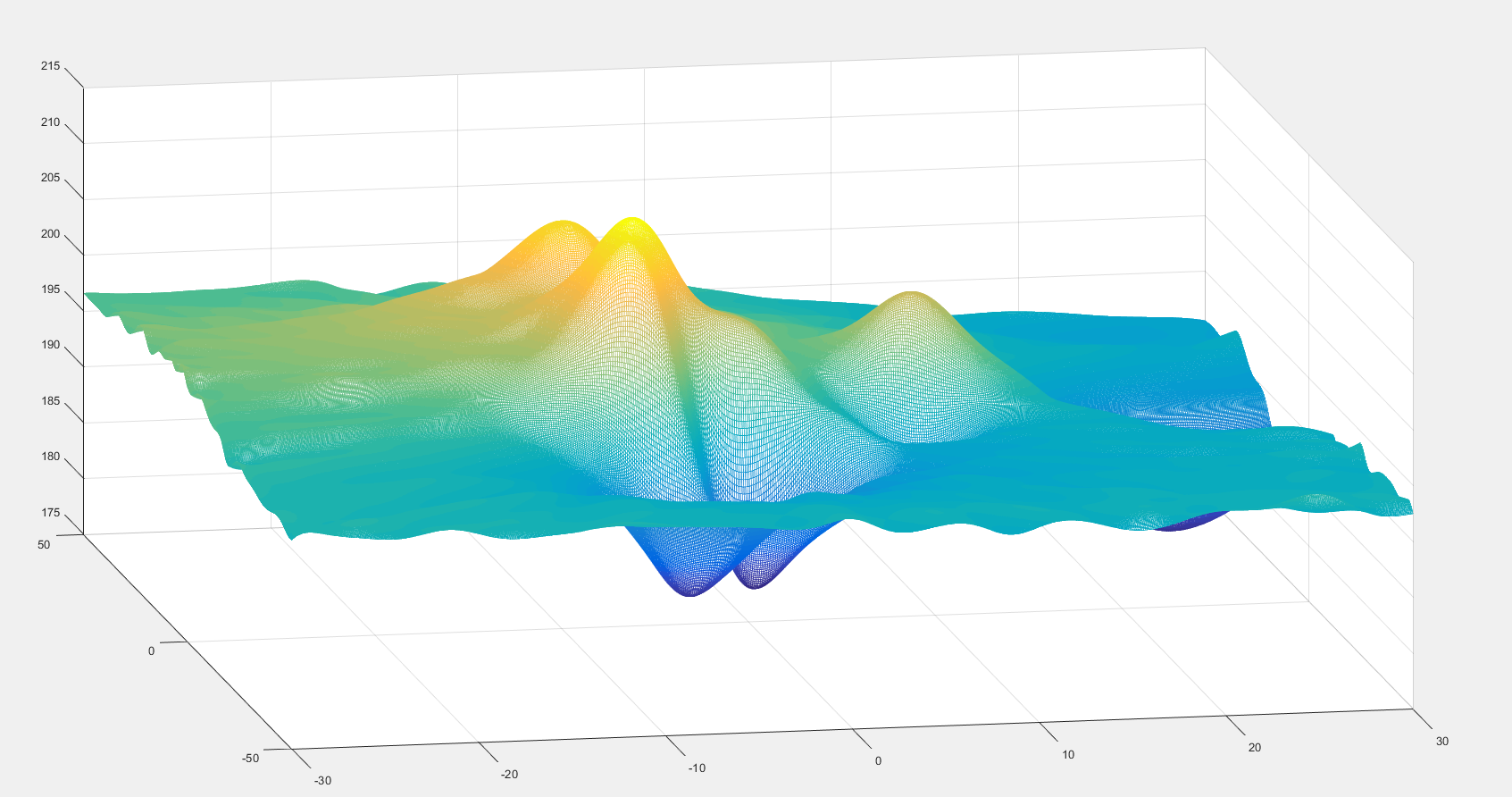
Shepard插值法原理如下所示，计算指定点与13个监测点之间的距离，以距离的反比作为权重乘以该点的PM2.5浓度，求和作为该点的PM2.5浓度。



Shepard插值法得到西安市以市人民广场为中心南北各50公里，东西各30公里范围内PM2.5的浓度分布。如下图所示。



Shepard插值法得到的PM2.5浓度空间分布比较平坦，不符合实际情况，将Shepard插值法得到的数据放入神经网络中进行训练，重新得到西安市PM2.5浓度的空间分布，如下图所示，从图可直观看出，高压开关厂的PM2.5含量是13个监测站中最大的，市人民体育场、广运潭、草滩PM2.5含量也较高；小寨、临潼区的PM2.5含量较低。



2.1.3 点评

原文思路清晰，以PM2.5全市平均浓度随时间的变化刻画西安市PM2.5浓度的时间分布；通过Shepard插值求得西安市PM2.5浓度的空间分布。但是原文程序在计算PM2.5浓度空间分布时用的却是立方插值，并没有使用上文提到的Shepard反距离插值。

2.2问题二第2小问模型的建立与求解

2.2.1 原文方法和思路

问题要求建立能够刻画该地区PM2.5的发生和演变（扩散与衰减等）规律的数学模型，合理考虑风力、湿度等天气和季节因素的影响，并利用该地区的数据进行定量与定性分析。

原文思路如下：

1. 利用该地区的数据进行PM2.5与风力、湿度等天气和季节因素间的多元回归分析，即建立PM2.5浓度的估算模型；
2. 利用大气污染物的扩散原理建立PM2.5扩散的偏微分方程模型，主要参考《城市大气污染扩散监测模型的理论与试验研究》这篇论文，本人能力有限，反正我看不懂。

2.2.2 解题过程重现

对于该问题的求解，我主要参考了国防科技大学同学的一篇优秀论文，利用经典的对流扩散方程进行建模，并使用Comsol Multiphysics对问题进行仿真实验。

对流扩散方程是一类基本的运动方程,它可以用来描述河流污染、大气污染、核废物污染中污染物质的分布,流体的流动和流体中热的传导等众多物理现象。它可以用来对流扩散问题数值计算方法的研究具有重要的理论和实际意义,可用于环境科学、能源开发、流体力学和电子科学等许多领域。这些对流扩散型偏微分方程(或方程组)具有一个共同的特点：对流的影响远大于扩散的影响,即对流占优性。对流占优性给对流扩散问题的数值求解带来了许多困难,因此,对流占优问题的有效数值解法一直是计算数学中重要的研究内容,用通常的差分法或有限元法进行求解将出现数值震荡。

如果近似认为扩散到地面的污染物不会再出现回升现象，则将整个污染扩散问题等效为一个三维自由扩散场，从而将扩散过程中的衰减项考虑到传播过程中， 因此式中可不显含衰减项。则方程可简化为：

另外，由假设分析，受目前相关气象和地理信息的数据限制，不能得到垂直方向上的逆温，层结，沉降等现象的数据，且所能得到的风向也同样不具备垂直方向的信息，因此，拟忽略各测量区的高度差异，将该扩散问题，简化为一个二维平面对流扩散问题。则方程可进一步简化为：

等号左边后两项为对流项，等号右边前两项为扩散项，第三项为源项。

颗粒物扩散系数是衡量以PM2.5 为代表的微小颗粒污染物扩散能力的重要参数，通过查阅相关研究文献，得到扩散系数的计算公式如下：

其中，d为颗粒物直径,T为绝对温度。从计算式中可以发现，扩散系数的计算主要考虑了各向同性介质中温度和颗粒大小的影响，为简化问题，可忽略了局部地形和空间温度梯度的影响，认为扩散系数近似为一个恒定的常数， 且为了简化研究，不妨取x，y方向扩散系数相同，即。

对于上述形式的非齐次抛物型方程，要求得解析解通常是不容易的，工程上常用到的是偏微分方程的数值解法，主要包括时域有限差分方法和有限元方法。

基本思想是：把连续的定解区域用有限个离散点构成的网格来代替。这些离散点称作网格的节点。把连续定解区域上的连续变量的函数用在网格上定义的离散变量函数来近似，把原方程和定解条件中的微商用差商来近似，积分用积分和来近似，于是原微分方程和定解条件就近似地代之以代数方程组，即有限差分方程组。解此方程组就可以得到原问题在离散点上的近似解。然后再利用插值方法便可以从离散解得到定解问题在整个区域上的近似解。

有限元方法的基本思想同有限差分法相似，都是通过离散化网格把原微分方程，用节点值的插值来近似表示整个有限单元内的值。区别在于，有限元方法需要找到偏微分方程相应变分问题的泛函，然后将插值关系带入到泛函中去，在每个单元上得到关于节点值为未知量的方程，将所有单元对应的方程组组合起来，得到一个大的方程组，然后根据初边值条件，适当处理方程组，并求出其解。

相比较而言，有限差分法最大的优点是比较直观、理论也比较成熟、精度可选，便于编程计算，但当区域不规则时，求解会变得困难；而有限元法的网格划分可以随意设置，能够对复杂区域进行计算，缺点是编程实现较复杂，但目前成熟的商业有限元软件很多，可以直接利用有限元软件进行数值计算，从而避免复杂的编程过程。在该问题的研究中，拟采用功能强大的多物理场仿真软件Comsol Multiphysics，借助其强大的计算内核和友好的人机界面进行仿真计算，分析PM2.5 浓度的演变规律。

2.2.3 点评

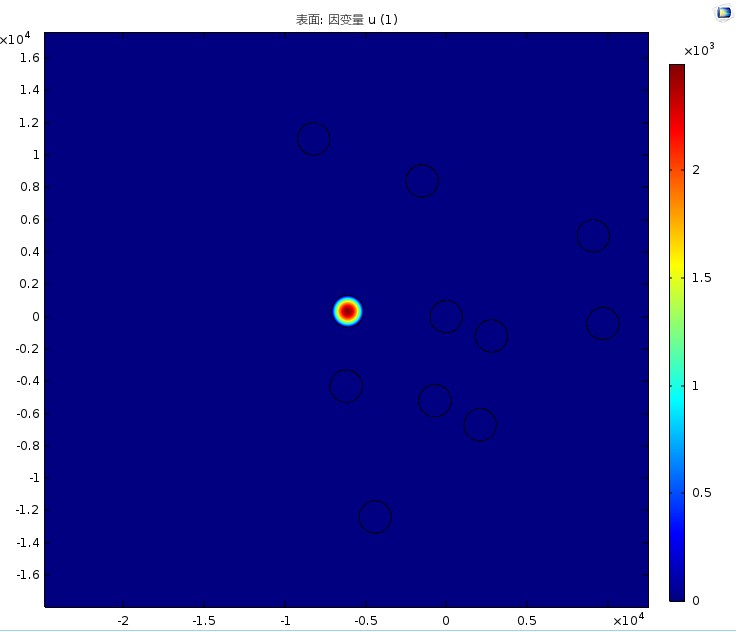
原文的方法对我而言晦涩难懂，对环境工程、污染物传播的相关知识要有良好的基础，超出了我专业能力范畴，所以我选择以国防科技大学的论文作为研究该问题的主要方法，国防科技大学的同学在论文中对所使用的方法进行了详细的阐述，通俗易懂，使我获益匪浅，并且该方法有良好的可扩展性强。南京航空航天大学的论文在进行PM2.5与风力、湿度等天气和季节因素间的多元回归分析时，提到“对于风速，根据风速与风级的对照表，由于基本都为小于等于三级风，若为平均值或等级值将为定常数不具有统计作用，对此我们另外在相关气息官网上查阅到的该地区风速情况的具体数据以及历史数据，对此为风速数据”，然而找遍全文我也没有看到风速历史数据的出处，对于一篇超过50页的科技论文而言，用一句话说明你论文中用到数据的来源应该不过分吧。

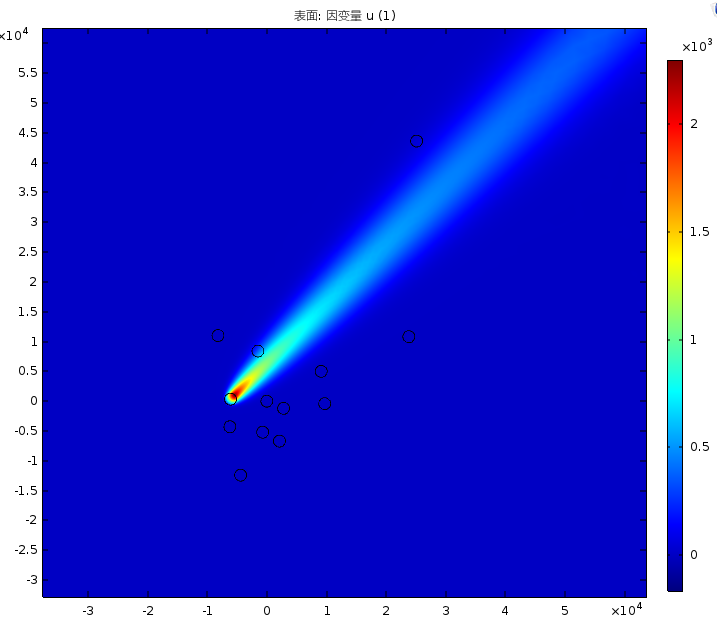
2.2问题二第3小问模型的建立与求解

原题要求以该地区PM2.5 监测数据最高的一天为例，在全地区PM2.5 浓度最高点处的浓度增至2倍，持续2小时，利用建立的模型进行预测和评估。由数据分析可知，在所测得的数据中，PM2.5 浓度最大值为2013年2月10日，浓度达到，而在这一天浓度达到最大值的监测点有多个，为了简化分析， 暂不考虑多点扩散的问题，于是选择取西安市整段观测中PM2.5 浓度一直较大的高压开关厂为污染源。并将其视为模型中心，通过谷歌地图得到其它各测量点相对高压开关厂的距离。当天的天气条件为：多云/多云，4℃/-3℃，无持续风向 ≤3 级 /无持续风向 ≤3 级。其中主要用到的是风速，将无持续风向近似认为是紊流场条件，该因素认为已考虑在扩散系数D 中，故不再加入定向风项，风速取为3级风的最大值。

由上述的扩散规律分析可知，扩散系数的确定对污染源的扩散分布有很大的影响，因此要利用模型进行准确的扩散预测与评估，需要得到较为准确的扩散系数D。

在无风条件下，PM2.5颗粒主要发生沉积现象，而在有风条件下，对流占主导作用，PM2.5在风向主导方向上发生对流扩散现象。可视化图像如下图所示：





3.总结

自己用具体方法实现可能还可以，但具体的数学逻辑推导可能就不能表达得很清楚，在数学逻辑写作方面需要向两篇优秀论文学习。

南京航空航天大学同学论文的问题：

1. 对于风速数据的出处没有说明；
2. 论文中求PM2.5的空间分布使用的是Shepard插值法，但是程序中使用的仅仅是立方插值；
3. Lingo程序与原文逻辑不一致；
4. 没有留下联系方式