

- 存在问题是得到的结果，相邻区域之间的雾霾水平没有相关性，需要进一步分析和修正；

## 未来工作

- 考虑风的影响，用人工生成的数据测试；

## 天气组

### 阶段进展

- 建立国控站点监测数据的实时下载和预处理系统；
  - 包括北京、天津、河北、山东、山西、辽宁、内蒙古等北京周边省份；
- 初步预处理，得到监测数据和站点坐标之间的准确对应关系；
- 系统每小时进行一次下载和预处理，储存在Haze01/AirMeasData中；
- 为实时预报系统的建立做好数据准备；
- 研究处理数据缺失和雾霾场重构方法
  - 基于Gappy-POD方法；
- 利用北京及周边地区的雾霾监测数据以及WRF/Chem输出的样本数据；
- 收集清华MEIC团队发布的MIX污染源数据，目前正在联系MEIC团队申请使用进一步的MEIC数据权限；

## 未来工作

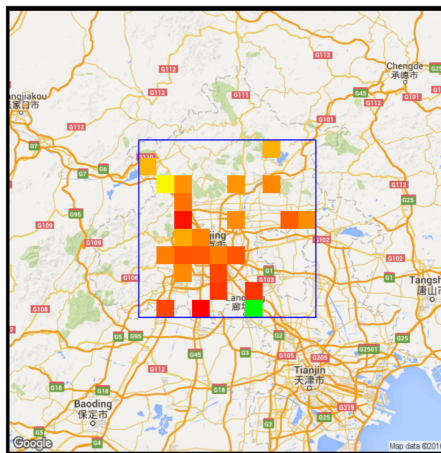
- 进一步研究和完善雾霾的初始场信息；
- 将MEIC污染源数据加入预测系统进行预测；
- 联系南京大学团队了解更多的WRF/Chem运行的设置方法和数据处理方法。

- pm2.5主要来源于二次污染（唐院士）
- 源解析问题（陶）：PKU-FUEL 2007全球燃料源，网站是[www.inventory.pku.edu.cn](http://www.inventory.pku.edu.cn)其上数据较为有用。
- 模型可考虑加入湿度的影响；
- 考虑站点局部污染源，大范围的污染源意义不大（陈）；
- pm10数据质量不高（尤其省控数据），人为干涉现象明显；
- “中国－英国大气污染国际合作研究计划”已经布局一个北京市来源辨析和释放清单研究重大项目；
- 文献：<http://www.fxyqpx.org/sciD/15zd0227.htm>基于实测PM2.5、能见度和相对湿度分辨雾霾的新方法

## 交通组

### 阶段进展

- 用矫正的北京省控pm10进行模型测试
  - 不考虑风的影响，只考虑雾霾往四周扩散；
  - 计算结果收敛；
  - 初步污染源强度见图：



# 每周简报

北京大学雾霾计算与预测项目组

第八期(2016.05.11-2016.05.17)



## 数据库与可视化组

### 阶段进展

- 维护数据库；
- 进行数据可视化方法操作；
- 机器学习方法尝试；
- 爬虫数据收集；

## 未来工作

- 网站建立；

## 数据组

### 阶段进展

- 数据矫正修复；
- 5.14会议总结：
  - 会议视屏：链接：<http://pan.baidu.com/s/1eSuKk7K> 密码：0b9q

## 数据精度研究问题

- 单点：在一点处放置  $n$  个相同的测量设备,能够将测量值的方差降为原来的  $1/n$ ,标准差降为原来的  $1/\sqrt{n}$ 。

【定理】  $x_1, x_2, \dots, x_n$  为相互独立、存在期望和方差的随机变量，则有：

$$Var(\frac{x_1 + x_2 + \dots + x_n}{n}) = \frac{1}{n^2} \sum_{i=1}^n Var(x_i)$$

特别地，当  $x_1, x_2, \dots, x_n$  服从同方差的正态分布，即  $x_i \sim N(\mu_i, \sigma^2)$ ,  $1 \leq i \leq n$  时，有

$$Var(\frac{x_1 + x_2 + \dots + x_n}{n}) = \frac{\sigma^2}{n}$$

- 利用空间信息

设在闭区域  $D$  上原有  $n$  个（可重）观测点  $A = \{A_1, A_2, \dots, A_n\}$ ，每点  $A_i$  的测量方差为  $\sigma_i^2$ 。我们重复以下步骤：

1. 通过  $A$  上的观测值，插值得到  $D$  上的场数据  $f(D)$ 。
2.  $\forall X \in D$ ，计算插值方法下  $X$  点数据的方差  $Var(f(X))$ 。
3. 找出使  $Var(f(X))$  达到最大的  $X \in D$ ，在  $X$  处放置 1 台观测设备，即将  $X$  加入  $A$  中。

以上算法希望在每次迭代后使得  $\max(Var(f))$  尽可能小，其中第三步里也可以放  $n$  台设备；

- 难点在于：
  - 插值方法  $D \rightarrow f(D)$  的选择；
  - $Var(f(X))$  的计算；
  - 算法的收敛性分析；
  - 算法的精度分析；

## 设备组

### 阶段进展

- 加深对光电感应器的理解
  - 电信号和接收到的光强非线性关系，但不影响对信号的区间划分以及再合并；
- 对新数据预处理；
- 尝试新的拟合方法
  - 出于模型卷机的本质考虑fft模型；但时域信息不完整，需对传统fft方法加以修正；

### 未来工作

- 检验模型
  - 利用现有数据比较训练出的拟合矩阵；
  - 结合当天空气质量判断训练数据是否足够；
- 尝试新的信号检测方法；
- 设计  $N$  台设备累积误差最小的训练方法：
  - 尽可能利用网上的其他数据 (例如 PM2.5, PM10 的数据)。
  - 设计多台设备的交叉训练方案, 使得累积误差最小。