

- 北京市企事业单位环境信息公开平台<http://58.30.229.134/index.do> (含有北京8个重点监控对象的污染物排放量)

- 收集遥感数据
  - 黄辉老师没有遥感数据;
- 建议:
  - 从中科院大气所、环保部、北大环科院获取;
  - 从美国或欧洲网站上下载, 如NASA和NOAA (网站数据质量可能不高);
  - 北大环科院的老师同学;

## 未来工作

- 进一步收集污染源数据:
  - 污染源排放数据可查看《2014年环境统计年报》。

## 交通组

### 阶段进展

- 改进模型, 测试算法:
  - 简化模型, 估计观测点处污染源;
  - 用人工生成数据测试算法, 收敛, 结果较为合理;

## 未来工作

- 进一步测试算法
  - 用人工生成数据对算法进行进一步测试和分析;
  - 测试实际数据:
    - 算法在实际数据上的表现还将依赖于模型的假设是否符合实际的数据;

## 天气组

### 阶段进展

- 详细解读了WRF/CHEM 的输入输出文件
  - 分析每个变量含义、单位、用法, 方便模式参数的优化、输出结果的可视化、数据分析、预测结果后处理;
- 分析目前主要的污染源数据库
  - GOCART\MOZART等数据源;
  - 分析污染源对北京雾霾预测的可用性;
  - 分析变量的含义、作用、用法;
- 总结雾霾预测的机器学习方法;
  - 来源文献和研究生数学建模竞赛;
- 完成预测系统的试运行;
  - 利用20160420-20160430气象数据和国外污染源统计数据;
  - 进行预测结果提取;
  - 进行部分可视化工作 (见邮件附件中O3变化的视频; )

## 未来工作

- 污染源数据更新修正;
- 雾霾插值方法学习, 得到初始场数据;
- 预测结果的统计后处理方法学习;

## 数据组

### 阶段进展

- 收集污染源数据:
  - 向公众环境研究中心 (IPE) 写信询问;
  - 获取污染源数据网站:
    - [http://www.mep.gov.cn/gkml/hbb/bgt/201602/t20160204\\_329897.htm](http://www.mep.gov.cn/gkml/hbb/bgt/201602/t20160204_329897.htm)
    - [http://www.mep.gov.cn/gkml/hbb/bgt/201602/t20160204\\_329897.htm](http://www.mep.gov.cn/gkml/hbb/bgt/201602/t20160204_329897.htm)

# 每周简报

北京大学雾霾计算与预测项目组

第七期(2016.05.04-2016.05.10)



## 数据库与可视化组

### 阶段进展

- 修复数据库数据:
  - 发现并解决数据中站点名字在不同时间不一致问题
- 学习更强大的爬虫方法pyspider;
- 机器学习试验:
  - 理论强化;
  - 初步试验训练;
- 熟练掌握可视化工具dimple;

## 未来工作

- 进一步优化数据库;
- 下周一前用dimple做出可视化图表;
- 试验新的机器学习模型;

---

## 未来工作

- 算法初步试验
- 先从空间场入手，不考虑时空场；

---

## 设备组

### 阶段进展

- 对模型进行理论分析：
  - 建立对粒子计数器扩散效应的数学描述；
  - 信号的区间划分；
  - 基于范数的误差分析；
  - 样本随机性分析；

### 困难问题

- 遗留理论问题：
  - 如何分析随机性，在数学上减小标准差对一致性的影响；
  - 对信号拟合和对光强拟合，哪一种得到的结果更贴合实际不同仪器之间的映射算子；
- 直接方法（根据信号和粒子分布求解光强）的困难，无法精确得到粒子分布；

### 未来工作

- 调整区间划分，进行数值实验；
- 研究更好的方法减小随机性产生的影响；
- 检验模型的方法；
  - 目前无法得到一个评判标准评价输出结果的好坏；
- 线性拟合方法可行性研究
  - 需要多少训练数据；
  - 需要什么样的训练数据；

---

## 低精度点的数据矫正问题

### 阶段进展

- 文献调研
- 用低精度点插值出高精度点，通过插出的值和实际观测值的误差，修正低精度点上的观测值；