

- 结果显示, WRF-CHEM的计算量远大于WRF, 单层133x133的网格, 预测时长48小时, 利用32线程计算, WRF-CHEM耗时7小时。并行效率为80%~90%。

未来工作

- 实时预测系统需要实时测量数据进行数据同化
 - 实时预测系统的驱动程序将集成数据组的测量数据自动抓取程序, 和基于测量数据的数据同化程序。
- 程序移植cluster解决计算量问题
 - 由于计算量较大, 在工作站运行无法满足实时预测需求, 以后将移植到Cluster进行运行。
- 更新数据源
 - 目前数据源来自于2000年和2005年的分析结果。
- 初步结果展示
 - 预测系统将暂时配置为单层网格进行实时预测, 需要与网站组合作进行预测结果处理与展示。

设备组

阶段进展

- 实现信号检测算法;
- 对两天一共14.5小时的数据进行简单的噪声分析;
- 采用时间平均下的光滑线性模型, 对已有数据进行拟合;
- IBM的目标是让各设备在1分钟内的测量达到一致性, 而每分钟测量的样本得到的粒子数波动过大, 于是将原先的简单非光滑线性模型改成时间平均下的光滑线性模型;

- 需要爬虫的网站:
 - 环保部环境监测总站空气质量实时发布系统(国控): <http://106.37.208.233:20035>
 - 省控数据(省控): <http://zx.bjmemc.com.cn>
- 数据修正与理解
 - 已与公司联系获取完好的数据;
- 数据解释:
 - pm2.5: 颗粒物(粒径小于等于2.5微米) 1小时平均;
 - pm2.5_24: 颗粒物(粒径小于等于2.5微米) 24小时滑动平均;
 - index: 污染物浓度对应的空气质量指数;

未来工作

- 收集污染源数据, 包括每个点每种污染物的日均排放量;
- 向黄辉老师收集遥感数据;

天气组

阶段进展

- 污染数据与WRF-CHEM程序的接口衔接
 - 分析来自RETRO, EDGAR等来源的污染源背景场、人为污染、生物排放、生物燃烧等污染数据, 完成污染数据的提取程序, 污染数据与WRF-CHEM程序的接口, 数据格式转换等工作, 使污染数据与污染源准确衔接。
- 完成北京地区预测系统运行配置
 - 完成运行参数设置, 预测系统的自动运行配置与驱动集成程序, 包括: 气象数据的自动下载、WRF-CHEM运行参数的自动配置、污染源数据的自动处理与转化、WRF-CHEM的自动运行、运行结果的提取与处理。
- 针对北京地区运行测试

每周简报

北京大学雾霾计算与预测项目组

第六期(2016.04.27-2016.05.03)



数据库与可视化组

阶段进展

- 数据爬虫
 - 监控空气污染物实时测量数据;
- 学习试验可视化方法
 - 基本方法已经掌握, 下周会给出图表;
- 机器学习方法试验
 - 由于输入变量过于简单, 结果不是很理想;

未来工作

- 给出初步的可视化效果;
- 进行进一步机器学习的尝试;

数据组

阶段进展

- 通过数据支持设备组对公司仪器较准的需求;
- 与数据库组合作, 收集每小时pm2.5和pm10数据;

- 省控逐小时pm10数据不一致，时间序列中出现较大间断；

未来工作

- 改进模型，测试算法
 - 用简单测试数据
 - 修改参数
 - 修改初值

未来工作

- 核实粒子变化与噪声变化趋势一致的原因
- 之后主要对不同空气成分进行拟合，需要继续采集样本模拟；
 - 目前仅两天的数据中，空气中颗粒物各成分的变化并不明显；
- 研究异常出现原因
 - 与公司沟通，了解设备；
 - 探究能否自动检测异常；
- 标准粒子分布函数
 - 针对0.3,0.5,1.0,2.5,5.0 微米的标准粒子，研究其脉冲分布函数；
 - 探寻更合适的粒子区间划分方法；
 - 由于用标准粒子对仪器标定存在困难，分布函数的日后参考价值有待确定；
 - 尝试更优的信号检测算法；

交通组

阶段进展

- 实现缺失数据补全和模型参数学习的迭代算法
 - 结果不是很理想；
 - 下一步将用测试数据进行检验程序；
 - 参数有待进一步调整；
- 将程序放入Github中，便于协同工作；
- 分析省控pm10数据
 - 国控数据点偏少，pm2.5只有8个；
 - 省控数据点不全；
 - 省控数据点时空不统一
 - 省控逐小时 PM10 数据中,北京在 2015 年 12 月之前有 35 个观测点,之后很多时候整个北京只有几个观测点,而且大部分河北地区的记录都是 0。