

选择必须十分谨慎，选择不同的化学机理产生的结果往往会有很大的区别。

- 本周详细解读了WRF-Chem的配置文件参数的含义和具体用法；
- 污染源数据及其处理程序配置
 - 数据获取
 - 生物排放、生物燃烧等自然污染源和部分人为污染源的统计数据；
 - 从RETRO和EDGAR等研究机构获取；
 - 预处理获取的统计数据
 - 在处理器配置污染源处理和转化程序：
prep_chem_source.exe ,convert_emmis.exe 测试成功。

低精度点的数据矫正问题

阶段进展

- 进一步理解问题
 - 问题简化：将连续场估计变成离散点场的估计；
 - 问题细化：
 - 如何衡量在未知点估计的值的精确程度；
 - 网格加密的收敛速度

未来工作

- 文献调研
 - 不同领域对该问题的提出和理解，如PDE,概率，优化等；
- Data Assimilation相关文献的查阅

- <https://www.zhihu.com/question/30600663/answer/48813715>
- <https://www.zhihu.com/question/22057894/answer/83767124>
- <https://www.zhihu.com/question/35854352/answer/67087503>
- <https://www.zhihu.com/question/20032670>
- 对相关网络进行爬虫收集数据
 - <http://aqicn.org/forecast/beijing/>
 - <http://www.bjmemc.com.cn>

数据组

阶段进展

- 增强对数据的理解
 - 国控点数据和省控点数据关系
 - 国控点包含于省控点中
 - 北京市国控点12个，省控点35个
 - 青悦的污染源数据：
 - <http://data.epmap.org/pollutions> 有数据示例和工厂例子
 - 数据缺失原因
 - 大部分由于原始数据缺失，比如通信、仪器故障；小部分源于爬虫。
 - 收集用于预测和较准的数据
 - 已将网站交与数据库组进行爬虫收集。

未来工作

- 解决国控点和省控点一致性问题；
 - 同一点的数据不一致的具体量化，误差衡量。
- 收集人口分布数据；
- 收集污染源信息数据：

天气组

阶段进展

- WRF-CHEM配置文件详解
 - WRF-Chem的配置文件与WRF具有较大不同，污染源的多样性和化学过程的复杂性使得参数的

每周简报

北京大学雾霾计算与预测项目组

第五期(2016.04.20-2016.04.26)



数据库与可视化组

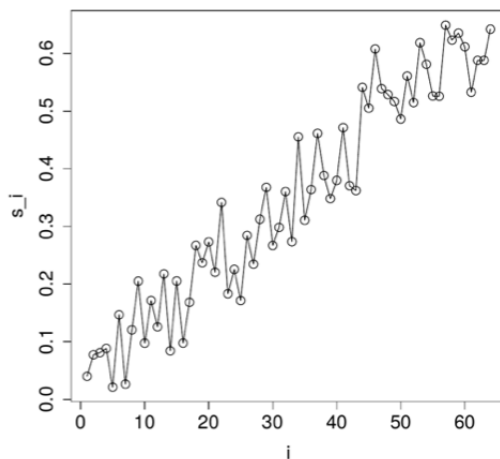
阶段进展

- 将天气数据在数据库进行了拆分，优化了存储结构
 - 5个 table: rainc, rainnc, t2, u10, v10
 - 3列: date, time, data(矩阵数据)
- 可视化方法学习
 - <https://www.zhihu.com/question/24545706>
 - <https://www.zhihu.com/question/38197892/answer/75559246>
 - <http://zhuanlan.zhihu.com/p/20641029>
 - <https://www.zhihu.com/question/21007937>
 - <https://www.zhihu.com/question/28509135>
 - <https://www.zhihu.com/question/20144795>
 - <https://www.zhihu.com/question/37636296/answer/72946536>
 - <https://www.zhihu.com/question/34495483/answer/88586049>

- Linux预处理数据文件；
- 进行数据与matlab交互；
- 算法实现，测试数据
 - 对上周信噪比低的数据测试，效果不理想；
 - 对本周获取的信噪比较高数据，正在参数调试中；

未来工作

- 数据准备工作
 - 搭建数据输入的matlab接口；
- 模型建立
 - 模拟颗粒流的信号转移行为；
- 信号处理
 - 进一步了解设备的物理机理；
 - 获取噪声性质；
 - 进行滤波研究。



(b)

Figure: Estimated s_i , Sample size = 561.

(a) Case 1 (b) Case 2

注意测试数据是根据给定模型生成的，而实际数据只能用模型去逼近，算法在实际数据上表现的好坏将很大程度上取决于模型的近似程度。

- 邮件联系清华的张强老师。

未来工作

- 将程序放入Github，便于协同工作。
- 将缺失数据补齐和模型参数学习迭代算法实现，测试其可行性。
- 用省控数据测试算法，可以先考虑无风情况下的污染源估计。
- 搞清楚风场的坐标网格参数（请求气象组的帮助）。

设备组

阶段进展

- 学习数据处理方法
 - labview读数据；

交通组

目标：

PM2.5 / 10污染源的估计；

阶段进展

- 弄清楚国控数据和省控数据的关系（与数据组合作）
- 收集整理pm10数据（省控数据中pm2.5的一小时数据没有记录，转为先分析pm10）
- 算法框架的初步实现
 - 实现belief propagation算法，用于缺失数据补全；
 - 实现pseudo-likelihood maximization 算法，从补全数据集中学习模型参数；
- 用simulated data进行初步算法校验

Case 1

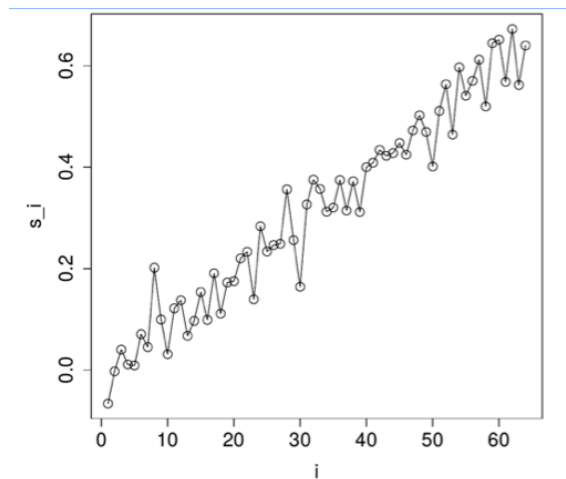
Parameters chosen: $s_i = 0.01 \times i$, $\beta = 0.01$ and $\gamma = 0.002$.

Parameters estimation: $\beta = 0.001116$ and $\gamma = 0.001866$.

Case 2

Parameters chosen: $s_i = 0.01 \times i$, $\beta = 0.05$ and $\gamma = -0.002$.

Parameters estimation: $\beta = 0.004788$ and $\gamma = -0.001987$.



(a)