

近年来由于严重的工农业污染、汽车尾气污染，中国多地频繁发生雾霾，而北京已经取代伦敦成为了新的雾都（霾都）。雾霾天的频繁发生严重的影响人们的生活和健康。“APEC蓝”及“阅兵蓝”等词汇的出现更是反映了人们对蓝天白云的渴望。雾霾的出现主要是由于PM2.5颗粒在空气中的大量积累造的。PM2.5是指大气中直径小于或等于2.5微米的颗粒物，有时也被称作入肺颗粒物，细颗粒物。它能较长时间悬浮于空气中，其在空气中含量浓度越高，就代表空气污染越严重。虽然PM2.5只是地球大气成分中含量很少的组分，但它对空气质量和能见度等有重要的影响。与较粗的大气颗粒物相比，PM2.5粒径小，面积大，活性强，易附带有毒、有害物质（例如，重金属、微生物等），且在大气中的停留时间长、输送距离远，因而对人体健康和大气环境质量的影响更大。在本项目中，我们将主要通过实际数据分析建模的方法来研究北京、上海、广州、成都及沈阳等五个城市的PM2.5情况。本项目中给出的数据主要包括以上5个城市的2008年以来每小时的PM2.5数据。注意，并非每个城市都有从2008年以来的所有数据。例如，沈阳只有2013年以来的数据。

1. 通过数据分析说明以上五个城市的PM2.5各有什么特点，包括PM2.5浓度对时间的依赖性，一年中个城市哪几个月的PM2.5最高，哪几个月的最低，一天中那个时刻PM2.5浓度最高、最低？分析可能的原因。各城市的PM2.5情况是否有所改善？哪个城市的污染更加严重？详细说明你做这样论断的依据（占比40%）
2. 对北京分别按小时和按天建立一个关于PM2.5浓度变化的时间序列模型，注意PM2.5浓度变化可能会有一定的趋势及周期，如有，你的建模中应能够反映这样的趋势及周期。利用数据分析说明你建立的模型可以较好的描述北京的实际PM2.5变化。尝试用你的模型对北京的PM2.5进行预测（按小时及按天），并设计一种办法衡量你预测的好坏（50%）
3. 你认为仅仅基于PM2.5数据是否已经足够，可以建立足够好的模型来描述PM2.5的变化并能够准确的预测PM2.5的变化？详细陈诉你的根据（10%）。如果你认为不够，尝试加入其它信息，详细说明你加入的其它信息是什么，并对北京PM2.5的变化建立新的模型，说明为什么你的新模型更好并用你的新模型对北京PM2.5进行预测。（此问为选作，选做的同学会获得0-20分的总成绩加分）。

注：所有数据均从[这里](#)下载，关于数据格式的说明可以参加[此文件](#)。