

利用社交媒体数据模拟城市空气质量趋势面

王艳东^{1,2}, 荆彤¹, 姜伟¹, 王腾¹, 付小康¹

(1. 武汉大学测绘遥感信息工程国家重点实验室, 湖北 武汉 430079;

2. 武汉大学地球空间信息技术协同中心, 湖北 武汉 430079)

摘要: 近年来, 随着城市的发展, 空气污染日益严重。目前我国城市空气质量监测主要依靠空气质量监测站, 然而监测站数量有限, 并且空气质量在一个城市的不同区域会出现较大起伏, 单一利用监测站不易于发现城市所有位置的空气质量起伏变化。新浪微博作为中国主流社交媒体之一不断的产生涉及社交用户日常生活的海量数据, 本文基于带有地理位置信息的新浪微博数据, 提出一种城市空气质量趋势面建立的方法。这是首次利用带有地理位置信息的社交媒体数据进行空气质量趋势面建立的尝试, 并对城市细粒度空气质量进行定量的模拟。该研究为城市局部空气质量模拟研究探讨了一种新型大规模数据源的可能性, 在反映城市不同位置的空气质量分布情况以及及时发现空气相对污染区域上具有重要意义。

关键词: 社交媒体; 新浪微博; 城市空气质量; 趋势面

中图分类号: P208

Modeling Urban Air Quality Trend Surface using Social Media

WANG Yandong^{1,2}, JING Tong¹, JIANG Wei¹, WANG Teng¹, FU Xiaokang¹

(1. State Key Laboratory of Information Engineering in Surveying Mapping and Remote Sensing, Wuhan University, Wuhan, China 430079;

2. Collaborative Innovation Center for Geospatial Technology, Wuhan, China 430079)

Abstract: Air pollution is getting worse with the development of cities in recent years. Urban air quality is mainly monitored by air quality monitoring stations at present. However, the number of stations is limited and the air quality fluctuates in different urban areas. So it is unefficient to detect air quality's distribution in a city by air quality monitoring stations only. Sina Weibo is one of the most popular social media in China. It produces large amounts of data about users' daily life. This study proposes an urban air quality trend surface modeling method based on Sina Weibo data with location information. This study is the first effort to model city's air quality trend surface from Sina Weibo data with location information, and infer the finegrained urban air quality quantitatively. The findings of this study evaluate the feasibility of using a new type of large-scale data source for research on air quality estimation of any location in a city, and is of great significance when reflecting air quality distribution and finding areas where are relatively air polluted.

Key words: social media; Sina Weibo; urban air quality; trend surface

0 引言

近年来随着城市的快速发展, 我国空气质量不断恶化。以 2013 年第一季度为例, 我国遭受了异常严重且持续时间较长的雾霾污染影响^[1]。目前城市空气质量监测主要依靠空气质量监测站对空气中的常规污染因子和气象参数进行监测, 通过分析得出空气质量指数 (Air

基金项目: 高等学校博士学科点专项科研基金 (20120141110036)

作者简介: 王艳东(1972-), 男, 教师, 博士研究生导师, 从事大数据时空分析计算、地理信息系统理论与工程等研究. E-mail: ydward@whu.edu.cn

Quality Index, AQI), 用以定量描述空气质量状况。由于监测站的建设成本高昂, 一个城市中站点数量是有限的。然而受多方面因素(如地表植被、交通流量、楼房密度等)影响, 空气质量在一个城市的不同区域会出现较大起伏^[2]。研究城市中不同区域空气质量的起伏状况有助于城市空气污染的监测和防治。

人类社会正进入“大数据”时代, 特别是在互联网和信息行业的发展引起人们的广泛关注^[3-5]。大数据的兴起, 为我们理解社会经济环境带来新的机遇, 可以有效捕捉社会经济特征^[6]。近些年, Facebook、Twitter 和微博(Microblog)等社交媒体成为人类重要的沟通媒介^[7], 是重要的大数据源。国内外学者利用社交媒体数据, 在发现城市格局^[8], 揭示人类移动和活动规律^[9, 10]等方面作了大量研究。Ferrari, Rosi 等利用概率主题模型分析 twitter 数据并自动提取城市格局^[8]; 刘瑜等利用社交媒体签到数据研究城市人类移动与活动的规律^[9]。同时, 很多学者也使用社交媒体数据进行各种模拟预测, 在疾病预测^[11-13], 灾害探测^[14-16], 选举结果预测^[17, 18]等方面进行了很多研究。Nagel AC, Tsou MH 等^[11]通过分析以流感为关键词的 twitter 数据, 发现真实世界流感爆发的出现与社交媒体信息的某些趋势具有很强相关性; Sakaki T, Okazaki M 等^[14]利用 twitter 的实时特性进行灾害探测, 预测地震发生时间以及定位灾害中心以及轨迹; Tsou M-H, Yang J-A 等^[17]通过对 2012 年美国总统大选两位候选人相关的社交媒体信息趋势进行比较, 较准确的预测了大选的最终结果。

新浪微博(Sina Weibo)是我国使用较广泛的社交媒体应用之一, 它连续不断的产生涉及社交用户日常生活的海量数据, 这些数据包含时间、用户情感等多种属性。用户利用带有 GPS 的设备, 比如智能手机等发布微博, 他们可以在微博中加入发布时的位置信息^[19]。随着空气质量的不断恶化, 许多人比之前更关注空气质量, 并借助新浪微博这个平台发表自身看法^[20]。这些带有位置信息的新浪微博数据从某种程度上可反映微博发布位置真实的空气状况。

本文提出了一种利用带有地理位置信息的新浪微博数据模拟城市空气质量趋势面的方法。首先分析了新浪微博数据与城市空气质量具有显著相关性。然后我们利用麦夸特算法拟合空气质量指数 AQI 与微博数量的函数关联。基于该关联, 推测出城市各格网中的 AQI, 并以 AQI 为虚拟高程值建立城市空气质量趋势面。这是首次利用带有地理位置信息的社交媒体数据进行空气质量趋势面建立的尝试, 对城市细粒度空气质量进行定量的模拟, 为城市局部空气质量模拟研究探讨了一种新的方法。

1 空气质量趋势面建立方法

趋势面可以模拟地理要素在空间上的分布规律, 展示地理要素在地域空间上的变化趋势。本文基于带有地理位置信息的新浪微博数据, 提出了建立城市空气质量趋势面的方法。方法主要流程如图 1 所示, 主要包含 4 部分: 相关数据采集与预处理; 新浪微博数据与空气质量指数的相关性分析; 空气质量指数与新浪微博数据间的函数关系拟合与精度验证; 城市空气质量趋势面建立。

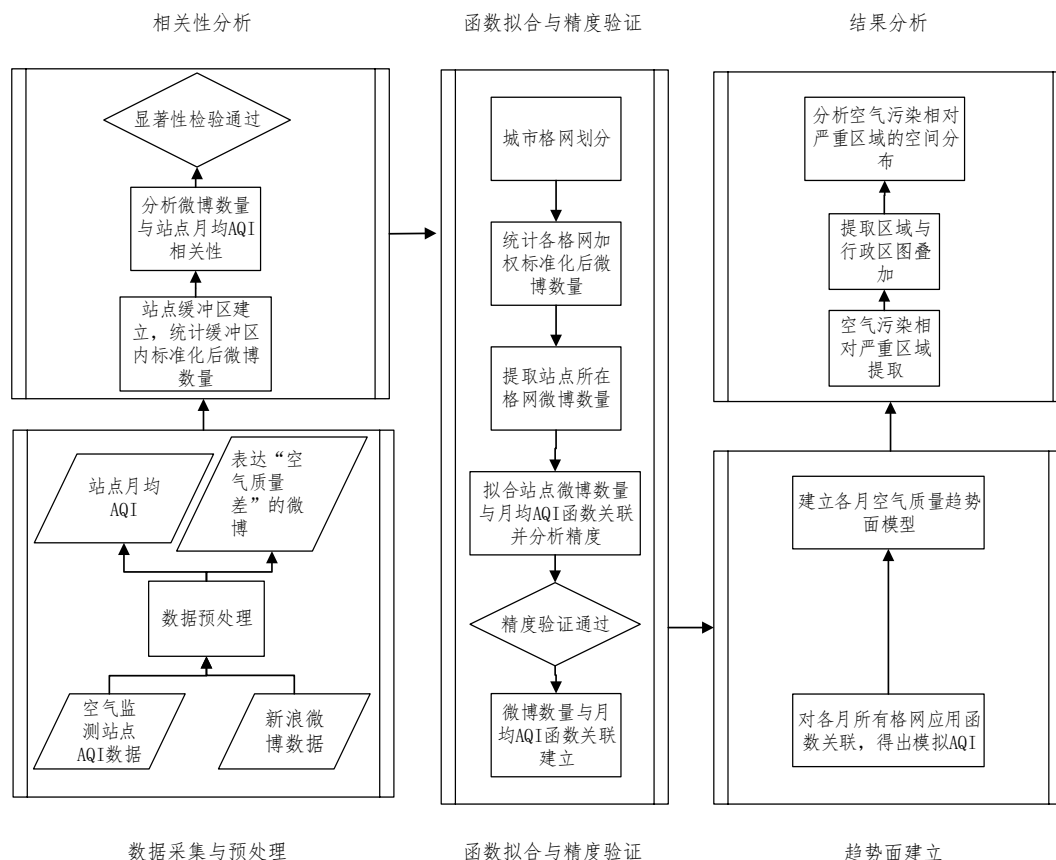


图1 城市空气质量趋势面建立与分析流程

Fig.1 The Processing of Establishing and Analysing of Urban Air Quality Trend Surface

1.1 数据采集与预处理

本文研究区域为北京市区，研究时间为2013年3月到8月。数据采集主要分为三部分：

1) PM25.in^①提供的2013年3月至8月北京市五环内7个空气质量监测站点（万柳，东四，农展馆，万寿西宫，天坛，奥体中心，官园）粒度为小时的AQI数据共29448条。对该部分数据，通过计算得出各站点每个月的平均AQI。

2) 我们事先对与空气污染相关的新浪微博数据进行人工判读，发现微博文本中较多出现“空气污染”，“口罩”，“雾霾”这些词。故使用这些词作为搜索关键词，利用网络爬虫与新浪微博API获取研究区域2013年3月至8月带有地理位置信息的新浪微博数据37409条。对数据先进行去重、去除转发微博、去除无关应用微博等预处理。然后用人工的方式进行微博情感分类预处理，具

体做法为成立一个5人小组，每个人对所有微博文本进行人工判读，只标记出表达“空气质量差”的微博。对于某条微博，如果同时被3个人及以上标记，我们就认为该条微博的确在表达“空气质量差”的含义，并将其选出作为实验数据。最终剩余数据量为9532条。约定文中该部分数据用Weibo_key表示。

3) 利用网络爬虫与新浪微博API获取研究区域2013年3~8月无关键词并带有地理位置信息的新浪微博数据，通过去重、去除转发微博、去除无关应用微博等预处理，剩余总量为7514729条。由于细粒度的人口密度数据无法获得，而人口密度在短时间内基本保持稳定，

^① PM25.in 是提供空气质量监测数据的网站，数据采自于国家环境保护部，<http://www.pm25.in/>

因此带有地理位置信息且无关键词的新浪微博数据在一定程度上可以反映人口的区域分布情况。因此本文用这部分数据来进行标准化工作,以消除人口密度带来的地区差异。文中约定用 Weibo_nokey 表示该部分数据。

1.2 相关性分析

目前大量的微博用户感受到周围恶劣的空气污染时会通过发微博来抒发不满、无奈的情绪,提出相应的建议等。这使得用户发布的关于空气污染的新浪微博数与空气污染状况之间存在一定的相关性。本文以北京市 7 个监测站点为样本,对样本周围 Weibo_key 数量与样本月均 AQI 之间的相关性进行分析。

首先以监测站点为中心,分别做出半径为 1km, 1.5km, 2km, 2.5km 的缓冲区(Buffer)。然后统计各样本点缓冲区内对应的 Weibo_key 数量与 Weibo_nokey 数量。利用相同缓冲区范围内 Weibo_nokey 数量进行标准化,以消除区域人口密度差异带来的影响。最后通过 SPSS 软件^②分析得出不同半径缓冲区内各样本 Weibo_key 数量与样本月均 AQI 的相关系数,如表 1 所示。当缓冲区半径为 2km 时,相关系数最高,达到 0.785, $p < 0.01$,说明此时 Weibo_key 数量与空气污染程度存在显著相关性。另外从图 2 所展示的 Weibo_key 数量与样本月均 AQI 的散点图可看出,Weibo_key 数量与月均空气质量指数存在较明显的正相关关系,表明本研究中使用的新浪微博数量可较好反映微博发布位置的空气质量情况。

表 1 Weibo_key 数量(标准化)与月均 AQI 相关性分析

Tab.1 The Correlation Analysis of the Standardized Quantity of Weibo_key and Monthly Average AQI

Buffer 半径	Pearson 相关性	显著性
1km	0.640**	0.000
1.5km	0.681**	0.000
2km	0.785**	0.000
2.5km	0.781**	0.000

**在 0.01 水平(双侧)上显著相关

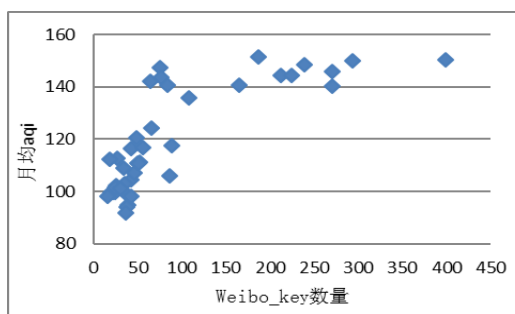


图 2 半径为 2km 的缓冲区内 Weibo_key 数量与月均空气质量指数的散点图

Fig.2 The Scatter Plot of the Quantity of Weibo_key and Monthly Average AQI in the 2km-Radius Buffer

1.3 函数拟合与精度验证

上节的相关性分析为利用新浪微博数据推测空气质量指数 AQI 打下了基础。本文设计了一种拟合月均 AQI 与 Weibo_key 数量之间函数关联的方法,根据某位置附近新浪微博数量推测该位置的空气质量指数。拟合流程如下:

- 1) 格网划分:对北京市研究区域进行格网划分(100m*100m)。

^② SPSS (Statistical Product and Service Solutions), “统计产品与服务解决方案”软件,用于统计学分析运算、数据挖掘、预测分析和决策支持。

2) 样本所在格网的微博数量获取: ①由于相关性分析研究时发现缓冲区半径为 2km 时相关性最大, 故以样本所在格网为中心, 以 4km*4km 为范围, 统计各样本所在格网中加权过后的 Weibo_key 数量与 Weibo_nokey 数量, 权重为设定范围内的各格网到样本所在格网地理距离的倒数。②采用 1.2 中的标准化处理方法对样本所在格网的 Weibo_key 数量进行标准化。③最终得到样本所在格网中经过加权标准化后的新浪微博数量, 记为 $GPS(w_norm)$ 。

计算方法如式 (1) 所示:

$$GPS(w_norm) = \frac{\sum_{i=1}^n \frac{1}{d_i} * N_i}{\frac{1}{6} * \sum_{j=1}^6 \sum_{i=1}^n \frac{1}{d_i} * M_{ji}} \quad (1)$$

其中 d_i 为 4km*4km 范围内任一格网到样本所在格网的地理距离, n 为同范围内的所有格网数, N_i 为 i 格网中的 Weibo_key 数量, M_{ji} 为第 j 个月 i 格网中的 Weibo_nokey 数量。

3) 样本月均 AQI 与 $GPS(w_norm)$ 的函数曲线拟合: 麦夸特算法是解决非线性拟合问题的有效方法, 其主要原理就是利用迭代程序进行计算残差平方和来评估是否达到最佳拟合效果, 当残差平方和达到最小值时, 迭代过程结束, 得出的即为拟合公式的最优结果^[21]。使用麦夸特算法 (Levenberg-Marquardt, LM), 需使残差平方和达到最小, 即要通过多次迭代使所求参数无限接近最佳参数。在本文中, 初始参数与函数公式选择, 以及算法迭代部分主要在 1stopt 软件^③中进行, 使用样本对应的月均 AQI 与 $GPS(w_norm)$ 数据迭代至收敛时, 拟合出的函数方程如式 (2) 所示:

$$Y = 5091.71 * X^{-0.63-69.21/X} + 99.86 \quad (2)$$

其中 Y 为月均 AQI, X 为 $GPS(w_norm)$ 。函数方程拟合优度 (R^2) 为 0.65, 该值是趋势线拟合的指标, 它表示函数关联 (2) 可解释拟合所用的 AQI 变异程度的 65%。F 检验统计量 (F-statistic) 为 263.75, 在 0.01 水平上显著, 说明拟合所用 $GPS(w_norm)$ 可较好的解释其对应的 AQI 数据。这些指标均表明该函数方程回归拟合效果较好, 各站点的函数拟合 AQI 与实际月均 AQI 相差较小。图 3 是样本的散点图及拟合函数曲线。从图中曲线可以看出, 样本的 $GPS(w_norm)$ 增大时, 其对应的月均 AQI 变化较快, 而当 $GPS(w_norm)$ 超过 400 时, 其对应的月均 AQI 变化逐渐趋于平稳。总体来说, 月均空气质量指数对于新浪微博数量的影响比较敏感, 当新浪微博数据量变化时, 其对应的月均 AQI 也会表现出相应的变化。本文从 7 站点中选出万柳与农展馆 2 个站点做出实际与推测 AQI 的对比柱状图, 如图 4 所示。从图中可以看出站点实际 AQI 与函数推测 AQI 接近, 表明函数拟合效果很好。

^③ 1stopt 是世界领先的非线性曲线拟合、综合优化分析计算软件平台。

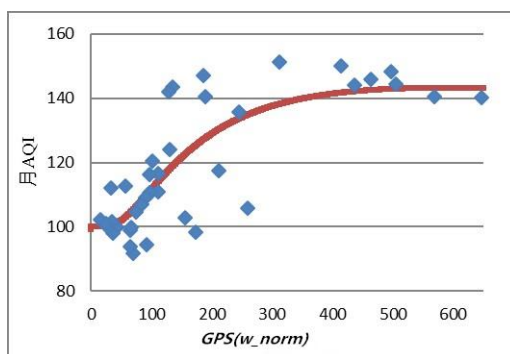


图 3 样本的散点图与拟合函数曲线

Fig.3 The Scatter Plot and the Function Fitting Curve of All Samples

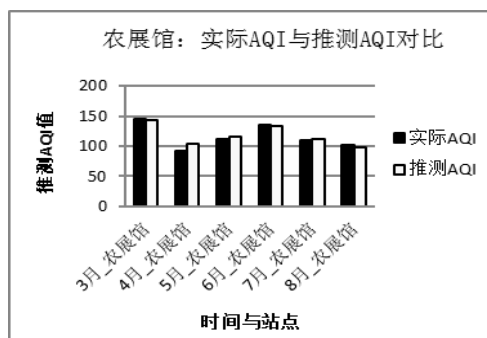


图 4 站点万柳与农展馆的实际与推测空气质量指数 (AQI) 对比

Fig.4 The Comparison of Observed and Calculated Air Quality Index(AQI) of Station Wanliu and Nongzhanguan

我们使用传统的十折验证进行精度验证, 将数据随机分为十组, 利用其中九组做拟合, 剩余一组做结果验证。经 10 次十折交叉验证得出平均的绝对平均误差 (Mean Absolute Error, MAE) 为 10.59。图 5 为 10 次交叉验证中 MAE 的分布图。从图中可以看出在 10 次验证中, MAE 均在 8~14 之间浮动, 处于较稳定状态。这些均表明本文得出的函数关联具有较好的稳定性, 可以对于北京市的各格网进行 AQI 推测。

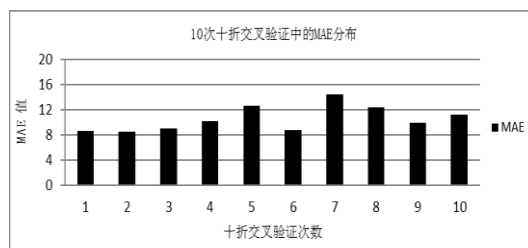


图 5 10 次十折交叉验证中的 MAE 分布

Fig.5 The MAE Distribution of 10 Times of the 10-fold Cross Validation

1.4 趋势面建立

利用上节的月均 AQI 与新浪微博数量的拟合函数可以推测出北京市所有格网的 AQI。以格网 AQI 为虚拟高程, 建立北京市研究区域内的空气质量趋势面, 进而可查询城市任意位置的空气质量。空气质量势面建立分为 3 个步骤:

1)与 1.3 中获取样本所在格网的微博数量方法相同, 得出北京市每个格网加权标准化后的 $GPS(w_norm)$ 。

2) 利用月均 AQI 与 $GPS(w_norm)$ 的拟合函数和格格网 $GPS(w_norm)$, 推测出所有格网的 AQI 值。

3) 以格格网推测 AQI 为虚拟高程, 模拟出 3-8 月分别关于 AQI 的数字高程模型(Digital Elevation Model, DEM)。此处的 DEM 并非传统的数字高程模型, 而是使用推测出的空气质量指数替代实际高程值作为虚拟高程建立起的空气质量数字高程模型, 即本文最终得出的空气质量趋势面。它可反映推测出的空气质量在空间上的起伏分布。其中图 6 为 3-8 月的北京市模拟空气质量三维趋势面在 ArcGIS 中显示的效果图, 红色凸起区域表示推测出的空气质量较差, 绿色凹陷区域表示推测出的空气质量较好。在趋势面上点击任意位置可以查询该位置的模拟 AQI, 能够清晰反映该位置的月均质量状况。

本文提出的建立城市空气质量趋势面的方法, 为展示细粒度的城市空气质量提供了有效途径。通过该方法建立的三维空气质量趋势面, 不仅可定性的表现区域空气质量的分布情况, 立体并直观的显示出空气质量相对污染与相对良好的区域, 而且与传统的密度图定性展示方法相比, 通过模拟空气质量指数定量的反映区域空气质量情况, 从具体数值上对细粒度的空气污染程度进行展现。

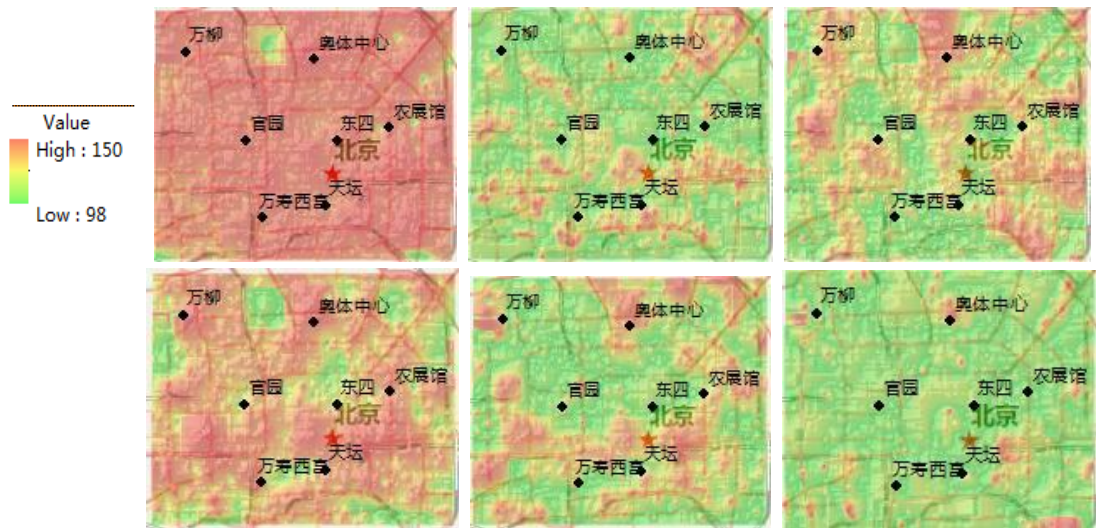


图 6 2013 年 3-8 月北京市模拟空气质量三维趋势面图
Fig.6 The Three-Dimensional Calculated Air Quality Trend Surfaces from March to August of Bei Jing in 2013

2 空气污染相对严重区域提取

由于受地形、经济、交通、人口、用地等各种因素影响, 一个城市中有些区域相对其它区域污染更为严重。提取这些区域有利于空气污染的预防与治理。本文设计的空气污染相对严重区域提取的具体方法如下: 先将各月每个格网的推测 AQI 从小到大排列, 约定每月排列中推测 AQI 大小排在后 1/3 的格网为相对污染区域。然后提取在 6 个月中出现 4 次及以上相对污染的格网, 可以认为这些格网的空气质量受时间影响较小, 受空间影响较大, 约定其为空气污染相对严重的区域。北京市 2013 年空气污染相对严重区域提取结果如图 7 所示, 红色区域表示模拟空气质量相较其它区域更差的地方。将空气污染区域图与行政图叠加后可勾画出这些较为污染的区域。例如图 7 中展示的方框 1 和方框 2 是朝阳区空气污染较差的两个区域, 1 是朝阳区奥体中心空气监测站点附近, 2 是朝阳区商务中心区附近。为了验证方法的正确性, 本文对 7 个监测站点 3~8 月每天每小时的 AQI 进行统计, 得出每个月各站点

严重污染等级 (AQI>300) 出现的次数, 并对每个月各站点出现该等级的次数进行排序, 统计各月中出现严重污染等级次数排名前三的站点。若某站点有 4 个月或以上均排名前三, 我们则认为该站点污染较严重。各月出现严重污染等级次数排名前三的站点情况如图 8 所示, 其中奥体中心站点是唯一一个出现 4 次 (3 月, 5 月, 6 月, 8 月) 的站点。所以奥体中心站点附近区域确实比其他区域空气污染更严重。



图 7 模拟空气污染行政区分布 (污染区域为红色部分)。其中框 1 为朝阳区奥林匹克公园, 框 2 为朝阳区商务中心区

Fig.7 The Administrative Distribution of Calculated Area Pollution (the red area is polluted).

The area 1 is Olympic Park, Chaoyang District, the area 2 is Central Business District, Chaoyang District

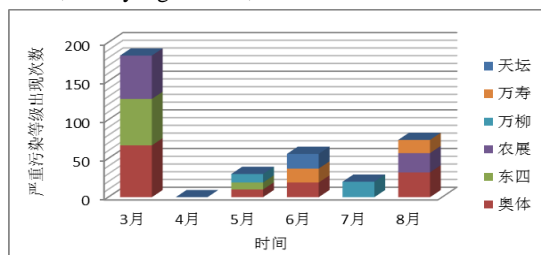


图 8 3~8 月各月中出现严重污染次数排名前三的站点

Fig.8 The Top Three Stations of High Occurrence Number of Severely Polluted Air Quality Level of Every Month

3 结论

目前城市空气质量不断恶化, 而空气监测站点个数过少, 无法进行城市空气质量精细检测。本文分析带有空气污染相关主题与地理位置信息的新浪微博数据与空气质量监测站点 AQI 数据的相关性, 并在此基础上建立两者间的函数关联, 利用该函数关联构建城市三维空气质量趋势面。本文提出方法不仅定性地表现出城市不同区域相对空气质量, 而且利用模拟的空气质量指数可定量、细粒度地展示城市空气质量情况。这对反映城市不同位置的空气质量分布情况以及及时发现空气相对污染区域上具有重要意义, 也为城市局部空气质量模拟研究探讨了一种新的方法。但研究过程中也存在一定不足, 比如无法获得城市实际空气质量趋势面, 从而无法确切的检验本文建立的空气质量趋势面的精度。

本研究为初步探索, 未来还有大量的研究工作有待完成。首先在空间上, 对于北京市的格网划分, 可尝试多种不同尺度, 分析各种尺度下的结果变化; 其次在时间上, 需要进行更细粒度的研究, 同时考虑时间的周期性变化影响, 分析城市每日的空气质量变化; 再次, 可在研究中加入更多相关的数据源, 比如气象数据等, 可提高算法的准确性。另外对趋势面进行空间分析时发现, 朝阳区商务中心区与奥体中心站点附近模拟空气质量相对较差, 可进一步对于这两个现象产生的原因进行分析。

[参考文献] (References)

- [1] Huang R-J, Zhang Y, Bozzetti C, et al. High secondary aerosol contribution to particulate pollution during haze events in China [J]. *Nature*, 2014, 514(7521): 218-222.
- [2] Zheng Y, Liu F, Hsieh H-P. U-Air: When urban air quality inference meets big data [C]. *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, 2013: 1436-1444.
- [3] Zhang Xiaoxiang. Spatial Analysis in the Era of Big Data [J]. *Geomatics and Information Science of Wuhan University*, 2014, 39(6): 655-659 (张晓祥. 大数据时代的空间分析 [J]. *武汉大学学报 o 信息科学版*, 2014, 39(6): 655-659).
- [4] Li Qingquan, Li Deren. Big Data GIS [J]. *Geomatics and Information Science of Wuhan University*, 2014, 39(6): 641-644 (李清泉, 李德仁. 大数据 GIS [J]. *武汉大学学报 o 信息科学版*, 2014, 39(6): 641-644).
- [5] Li Deren, Yao Yuan, Shao Zhenfeng. Big Data in Smart City [J]. *Geomatics and Information Science of Wuhan University*, 2014, 39(6): 631-640 (李德仁, 姚远, 邵振峰. 智慧城市中的大数据 [J]. *武汉大学学报 o 信息科学版*, 2014, 39(6): 631-640).
- [6] Liu Y, Liu X, Gao S, et al. Social Sensing: A New Approach to Understanding Our Socioeconomic Environments [J]. *Annals of the Association of American Geographers*, 2015, 105:1-19.
- [7] Huberman B A, Asur S. Predicting the Future with Social Media [C]. *Proceedings of the Web Intelligence and Intelligent Agent Technology, IEEE/WIC/ACM International Conference on IEEE*, 2010: 492-499.
- [8] Ferrari L, Rosi A, Mamei M, et al. Extracting urban patterns from location-based social networks [C]. *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks*, ACM, 2011: 9-16.
- [9] Liu Y, Sui Z, Kang C, et al. Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data [J]. *PloS one*, 2014, 9(1): e86026.
- [10] Liu Y, Kang C, Gao S, et al. Understanding intra-urban trip patterns from taxi trajectory data [J]. *Journal of geographical systems*, 2012, 14(4): 463-483.
- [11] Nagel A C, Tsou M-H, Spitzberg B H, et al. The complex relationship of realspace events and messages in cyberspace: case study of influenza and pertussis using tweets [J]. *Journal of medical Internet research*, 2013, 15(10): e237.
- [12] Salathé M, Khandelwal S. Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control [J]. *PLoS Comput Biol*, 2011, 7(10): e1002199.
- [13] Achrekar H, Gandhe A, Lazarus R, et al. Predicting flu trends using twitter data [C]. *Proceedings of the Computer Communications Workshops (INFOCOM WKSHPS)*, 2011 IEEE Conference on, IEEE, 2011: 702-707.
- [14] Sakaki T, Okazaki M, Matsuo Y. Earthquake shakes Twitter users: real-time event detection by social sensors [C]. *Proceedings of the 19th international conference on World wide web*, ACM, 2010: 851-860.
- [15] De Longueville B, Smith R S, Luraschi G. Omg, from here, i can see the flames!: a use case of mining location based social networks to acquire spatio-temporal data on forest fires [C]. *Proceedings of the 2009 international workshop on location based social networks*, ACM, 2009: 73-80.
- [16] Yates D, Paquette S. Emergency knowledge management and social media technologies: A case study of the 2010 Haitian earthquake [J]. *International Journal of Information Management*, 2011, 31(1): 6-13.
- [17] Tsou M-H, Yang J-A, Lusher D, et al. Mapping social activities and concepts with social media (Twitter) and web search engines (Yahoo and Bing): a case study in 2012 US presidential election [J]. *Cartography and Geographic Information Science*, 2013, 40(4): 337-348.
- [18] Tumasjan A, Sprenger T O, Sandner P G, et al. Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment [J]. *ICWSM*, 2010, 10:178-185.
- [19] Burton S H, Tanner K W, Giraud-Carrier C G, et al. "Right time, right place" health communication on Twitter: value and accuracy of location information [J]. *Journal of medical Internet research*, 2012, 14(6): 86-96.
- [20] Kay S, Zhao B, Sui D. Can Social Media Clear the Air? A Case Study of the Air Pollution Problem in Chinese Cities [J]. *The Professional Geographer*, 2015, 67(3): 351-363.
- [21] 程荣兰, 陈广志. 利用麦夸特理论高效求解 Van Genuchten 模型参数 [OL]. [2008-09-09]. <http://www.paper.edu.cn/releasepaper/content/200809-230.html>