以福州为示范的城市空气质量实时精细化模拟与预报

胡清华[1]*，陆　晨[2]，胡　倩[2]，魏淑珍[1]，蒋东升[1]，黄艳艳[1]

（1.福建省环境监测中心站，福建福州 350003；2.北京思路创新科技有限公司，北京 100085）

【摘　要】本研究在已有空气质量站点的基础上，采用多功能、高效、小巧的传感设备进行监测加密，利用大数据的分析机理与空气质量模型的城市计算技术，以海峡西岸城市群中的福州市为示范，建设城市空气质量实时精细化模拟与预报系统。实时分析城市细粒度（1km×1km）的空气质量，有效提高了城市不同地理位置污染物浓度差异性的分辨率，同时对城市空气污染的发展趋势提出预测，为精细化分析污染成因提供基础数据，从而为环境管理提供决策辅助支持、市民出行提供帮助与指导。

【关键词】海西；空气质量；实时精细化；模拟预报

【中图分类号】P456.9　　【文章编号】1674-6252（2018）03-0099-06

【文献标识码】A　　【DOI】10.16868/j.cnki.1674-6252.2018.03.099

数值预报模式的研究应用始于本世纪初，至今已有10余年，出现了一大批出色的城市、区域和全球尺度的数值模式[1]。数值模式预报要求有比较详尽的污染源资料和气象资料，并且要求对影响污染物扩散的所有因子要有精确的化学、物理和数学描述，多为中大尺度，最小网格一般为3km×3km。对于城市局部地域尺度的环境空气质量无法做到精细化预测预报，同时受限于处理时效及计算机硬件设备，难以做到实时预测预报。

统计预报是在不掌握事物变化机理的情况下，通过分析事物规律来进行预测的方法，将历史上的污染物浓度监测值与前期和同期的气象条件联系起来，建立具有一定信度的统计关系，并利用该关系对未来的污染物浓度进行预报。然而，污染物浓度监测值主要依靠地面空气自动监测站点的实时监测数据来获取。由于站点建设和维护成本高昂，一个城市通常仅有有限数量的空气质量监测站点，并不能完全覆盖整个城市范围。同时，受地表植被、交通流量、人群分布、建筑物密度和气象条件等各种复杂因素影响，随地域不均匀地变化，城市中不同地域的空气质量差异显著。因此，有限站点的监测数据并不能完整、细致反映整个城市的空气污染情况，整个城市的精细空气状况也不宜用一个或有限的几个笼统数据来概括。近年来，随着感知技术和计算环境的日渐成熟，各种大数据在城市里悄然而生，如交通流、气象数据、道路网、兴趣点（POI）、移动轨迹和社交媒体等，这些数据既反映了城市中存在的问题，也可以用来解决城市所面临的挑战[2]。

本研究以海峡西岸城市群中的福州市为示范，在已有空气质量站点的基础上，采用多功能、高效、小巧的传感设备进行监测加密，利用大数据的分析机理与空气质量模型的城市计算技术，建设城市空气质量实时精细化模拟与预报平台。使用已有及新增的实时和历史空气质量监测数据，结合交通流、道路结构（路网）、POI分布、气象条件和人们流动规律等多种数据源的大数据，利用机器学习算法建立数据和空气质量的映射关系，实时分析整个城市细粒度（1km×1km）的空气质量，尝试为政府与公众提供精细化的城市空气质量服务。此外，还能对城市空气污染的发展趋势提出预测，为分析污染成因提供基础数据，从而为环境管理提供决策辅助支持，为市民出行提供帮助与指导。

1 总体设计

1.1 加密监测站点布点

为保证数据的可比性及可靠性，同时便于系统模型评估比较，在研究区域内增加布设了18个小型空气质量监测仪器，多数布设于紫阳和师大两个城市标准站点周边，其中紫阳站点周边8个、师大站点周边6个，与标准站点的距离在200～1 500m。

基金项目：福建省环保科技计划项目（2015R016）。

* 作者简介：胡清华（1979—），男，高级工程师，主要从事环境统计、空气质量监测及预报预警等研究，E-mail: huqh@fjemc.org.cn。

1.2 平台框架设计

以福州为示范的城市空气质量实时精细化模拟与预报平台分为数据感知层、数据管理层、计算支撑层、推测预测层和展示应用层等五个层次，具体的结构图见图 1。

系统分为数据感知层、数据管理层、计算支撑层、推测预测层和展示应用层等五个层次。

（1）数据感知层

数据感知层通过监测设备和传感器感知和获取平台所需数据，包括天气特征、交通流量相关特征、人类移动特征、路网结构特征和 POI 相关特征。

（2）数据管理层

利用时空索引、流数据、轨迹数据管理、图数据管理、异构数据索引等方法管理收集的城市数据。

（3）计算支撑层

提供数据分析计算的各类支撑工具，运用数学方法对获取和管理后的数据进行分析，包括数据挖掘、模式识别、机器学习和可视化等方法。

（4）推测预测层

通过加载 POI、路网数据，加载空气质量、气象数据，采集需要计算的时间点，设置模型参数、加载指定模型，代入模型计算等流程实现空气质量的推测；在空气质量推测的基础上，通过加载组装数据，考虑周边因素，计算预测特征数据等步骤完成对空气质量的推测和预测。

（5）展示应用层

包括网格分析和数据对比两部分内容，网格分析是在 GIS 地图上，对站点实时监测、网格推测、统计分析以及预测的结果进行直观的查询展示；数据对比是对加入小型监测设备后的网格、站点预测的数据进行对比与评估。

2 系统采用的关键技术

在系统建设过程中，采用了多种技术来完成系统构建和空气质量推测预测，主要包括传感器技术、物联网采集与管理技术、数据挖掘技术、城市计算优化技术、混合数据可视化技术等等。

2.1 传感器技术

本研究采用的传感器技术实现现有专业传感器（如温度传感器、位置传感器、空气质量自动监测站、空气质量小型站等）之间的互联互通，完成数据的快速收集[3]。本研究采用的空气质量监测仪器为适合于长期运行的微小型在线式仪器，直接安装在电杆等通用墙物上，采用传感器方式对环境空气中 $PM_{2.5}$、PM_{10}、SO_2、NO_2、CO、O_3、温度、湿度进行实时在线分析，数据产生的时间频率是分钟级，功率≤5W，可采用太阳能供电或外接市电供电，数据无线传输并内置存储功能。其中：① $PM_{2.5}$、PM_{10} 采用光散射法测量，测量范围 0～1 000μg/m^3，检出限分别为≤10μg/m^3、≤20μg/m^3，分辨率≤1μg/m^3。② SO_2、NO_2、O_3 采用电化学分析法测量，测量范围 0～500ppb，检出限≤5ppb，分辨率≤0.01ppb，响应时间≤45s。③ CO 采用电化学分析法测量，测量范围 0～50ppm，检出限≤0.1ppm，分辨率≤0.1ppm，响应时间≤45s。④温度测量范围 −20℃～55℃，分辨率≤±1℃。⑤湿度测量范围 5%～95%RH，分辨率≤±1%RH。

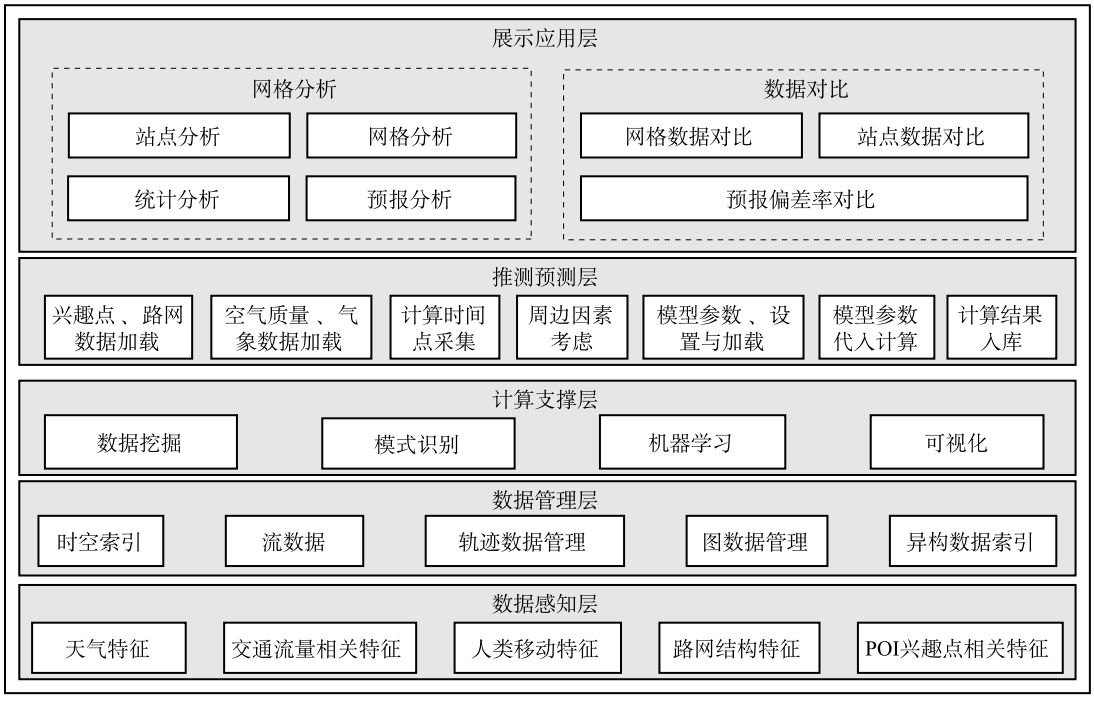


图 1 平台总体框架图

2.2 城市计算优化技术

城市计算是一个交叉学科，是计算机科学中以城市为背景，与城市规划、交通、能源、环境、社会学和经济等学科融合的新兴领域。城市计算将无处不在的感知技术、高效的数据管理和分析算法，以及新颖的可视化技术结合，致力于提高人们的生活品质、保护环境和促进城市运转效率。城市计算帮助理解各种城市现象的本质，甚至预测城市的未来[4]。

精细化网格预测是利用空气质量监测站点的实时和历史数据，结合交通流、道路结构、POI 分布、气象条件和人群流动规律等大数据，利用机器学习算法建立大数据和空气质量的映射关系，从而推断出整个城市细粒度的实时空气质量[5,6]。

模型将具有空间属性的信息和具有时序特性的信息作为输入，训练半监督的模型框架，从而得出城市细粒度的实时空气质量。整个模型计算流程如图 2 所示。

首先是数据采集与预处理。由交通工具如出租车、公交车、安装了 GPS 的私家车移动产生的空间轨迹，通过地图映射算法将这些轨迹映射到路网中，映射数据存放在轨迹数据库作为离线学习用，同时会生成地理索引以提升在线推断的效率。

然后从各个区域提取的不同数值中进一步提取特征，可以分为两类，一类是时间属性数据，特征值随时间而变化，例如温度、湿度、平均车速，这些数据可以从天气数据和空间轨迹数据中提取出来。另一类是空间属性数据，如 POI 的密度，道路的长度，这些数据可以从 POI 和路网数据中获取。如果监测站点正好是在网格内的某个位置，这个网格会被标记由该监测站点获得的 AQI 或其他数据，这个特征值就会被提取出来形成相应的标签作为训练集。但由于监测站点有限，而需要参照这一数据的地方却非常多，采用半监督的学习方法使用未加标签的数据提升推断的准确率。首先使用两个不同的分类器，通过特征集分别训练加标签的数据，一类是基于线性条件场的时序分类器，用于测算某一个位置空气质量的时序变化，另一类是基于神经元网络的空间分类器，用于测算不同位置空气质量的空间关联性。现有监测站点的 AQI 会作为空间分类器的输入。由于不同类型污染物（如 NO_2 和 PM_{10}）的影响因素不同，因此需要为每一种污染物构建模型。

最后，根据网格的影响区域计算每一个网格的特征。同时空间特征（如 POI 的分布）做离线计算，时间特征做在线计算，例如交通相关的特征基于在预处理流构建的时空索引中提取出来。对于每一个网格，将时序属性的特征放入时序分类器，空间属性的特征放入空间分类器。由于监测站点每小时发布数据，模型也每小时生成一次结果，结果包括空气质量网格推测数据和预测数据。

2.3 混合数据的可视化技术

可视化技术充分利用人们对可视模式快速识别的自然能力，将数据信息和知识转化为如图像、图形、表格等之类的视觉形式，将信息以视觉形式表现出来，利用人们视觉通道的快速感知能力去观察、识别和加工信

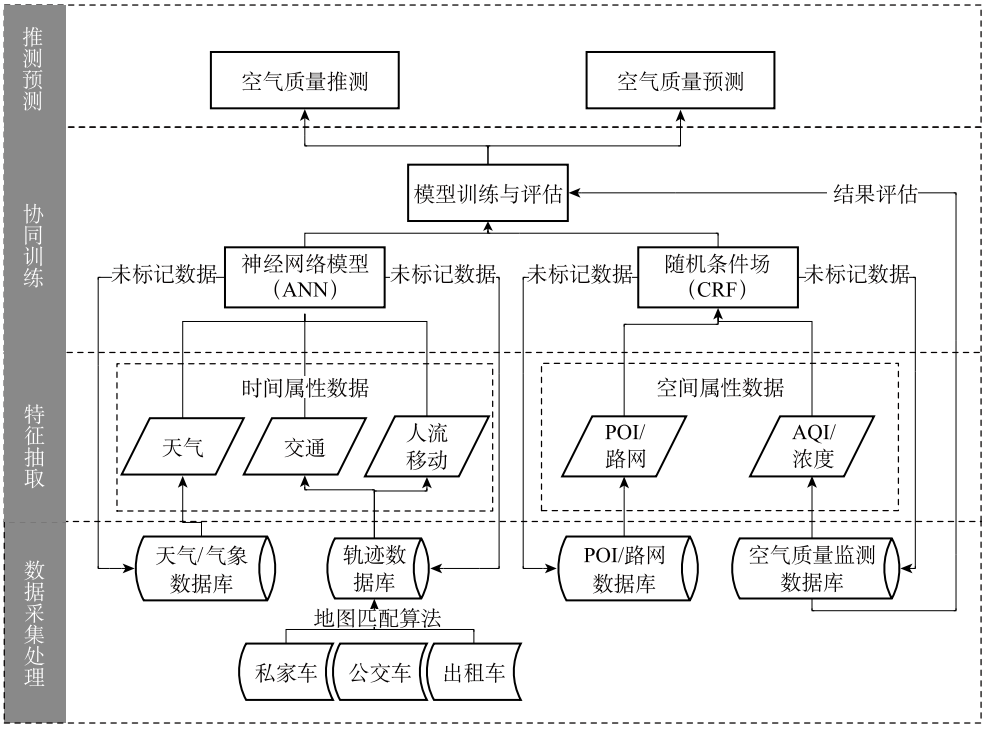


图2　城市计算模型框架图

息，以直观的方式帮助理解获取的知识和模式[7-9]。与单一数据可视化不同，城市计算中的可视化技术需要同时考虑多个维度，其中空间和时间是两个至关重要的维度。

3 平台效果及评估

3.1 平台功能实现

依托于框架设计和关键技术，本研究成功实现了城市空气质量实时精细化模拟与预报（福州示范）系统平台功能的开发。目前平台已经在福建省级环境监测部门实现了稳定运行。

运行结果达到了研发的预期目标：

（1）实现了城市细粒度级别（1km × 1km）环境空气质量的实时预报与模拟，弥补城市尺度预报粒度与实时性的不足。

（2）实现了基于小巧、高效、快速响应和节能传感设备的结合应用，结合城市环境质量监测站点更有效反映城市环境质量。

（3）实现了城市环境大数据的典型应用，包括海量异构数据的管理及协同计算；不同数据源中相互增强知识的获取及提取深度的保证；大数据的数据稀疏性问题的合理应对。

3.2 精细化实时模拟效果评估

（1）站点数据对比分析

分别选择 2017 年 11 月 1 日至 12 月 27 日紫阳站点及其周边的 1#、3#、6#、7# 四个小型站、师大站点及其周边的 1#、2#、3# 三个小型站，对小时监测数据进行统计分析，结果如图 3 所示。

紫阳和师大两个站点各项污染物均值基本一致。

紫阳站点及其周边小型站的 $PM_{2.5}$、PM_{10}、NO_2 和 CO 均值基本一致，而 SO_2 均值存在较明显的偏高、O_3 均值相对偏低，且小型站 7# 偏差较大。比较两类站点的最大值与最小值可以发现，小型站的 $PM_{2.5}$、PM_{10} 和 SO_2 最小值较高、最大值较低，NO_2 和 O_3 最大值和最小值均偏低，CO 则基本一致。

师大站点及其周边小型站的 SO_2、NO_2、O_3 和 CO 均值基本一致，而 $PM_{2.5}$、PM_{10} 均值存在较明显的偏低，且小型站 1# 偏差较大。比较两类站点的最大值与最小值可以发现，小型站的 $PM_{2.5}$、PM_{10} 最小值较高、最大值较低，SO_2 最小值较低、最大值较高，NO_2 和 O_3 最大值和最小值均偏低，CO 则基本一致。

（2）精细化实时模拟效果对比

由于紫阳站点和小型站 1#、3#、5#、6# 同在一个网格，而小型站 2#、4#、8# 在另外一个网格，为了验证本研究采用的城市计算方法推测结果的准确性，挑选小型站 8# 的实测数据与所在网格推测数据进行对比。师大站点和小型站 2#、3#、4#、5# 同在一个网格，故挑选小型站 1# 的实测数据与所在网格推测数据进行对比，对比结果如图 4 所示。从分析结果（2017 年 11 月 25 日至 12 月 25 日）可以看出，推测结果与实测结果趋势较为一致，但是由于受到参与计算的其他参数的影响，推测结果存在一定的不稳定性和跳跃性，在后续的研究中有待对模型进一步训练优化。

4 结论

本研究在已有空气质量监测站点基础上，采用多功能、高效、小巧的传感设备进行监测加密，使用已有和新增的实时和历史空气质量监测数据，结合交通流、道路结构（路网）、POI 分布、气象条件和人们流动规律等多种数据源的大数据，利用城市计算技术建立数据和空气质量的映射关系，实现了城市细粒度（1km × 1km）的空气质量推测预测，并以福州市为示范建设了海西重点城市空气质量实时精细化模拟与预报系统并应用。

通过将现有站点及其周边小型站的监测数据进行统计分析发现，不同地理位置的污染物浓度存在一定的差异性。为提高大气污染物浓度监测的分辨率和准确性，开展加密监测很有必要。

通过比较小型站实时监测数据及城市计算方法推测结果，发现两者趋势十分吻合，但是推测结果受到参与计算的其他参数的影响，存在一定的不稳定性和跳跃性，在后续研究中有待对模型进一步训练优化。

参考文献

[1] 任万辉 , 苏枞枞 , 赵宏德 . 城市环境空气污染预报研究进展 , 环境保护科学 [J]. 2010, 36(3): 9-11.

[2] 郑宇 . 城市计算概述 [J], 武汉大学学报 • 信息科学版 , 2015, 40(1): 1-13.

[3] GOLDMAN J, SHILTON K, BURKE J, et al. Participatory Sensing: A Citizen-Powered Approach to Illuminating the Patterns that Shape our World[EB/OL]. 2014. http://www.mobilizingcs.org/wp-content/up-loads/Particip atory_Sensing.pdf.

[4] ZHENG Y, CAPRA L, WOLFSON O, et al. Urban Computing: Concepts, Methodologies, and Applications[J]. ACM Transactions on Intelligent Systems and Technology, 2014, 3(5):38.

[5] ZHENG Y, LIU F R, HSIEH H P. U-Air: When Urban Air Quality Inference Meets Big Data[C]. KDD Chicago IL USA, 2013.

[6] ZHENG Y, CHEN X, JIN Q, et al. A Cloud-Based Knowledge Discovery System for Monitoring Fine-Grained Air Quality[R]. MSR-TR-2014-40, 2014.

[7] 程时伟 , 孙守迁 . 信息可视化研究综述 [J] , 中国科技论文在线 , 2008: 1-8.

[8] NAHUM D, STEPHEN G. Information visualization[J]. IEEE Computer Graphics and Applications, 1997, 17(4):29-31

[9] BEDERSON B, SHNEIDERMAN B. The craft of information visualization: readings and reflections [M]. San Francisco:Morgan Kaufmann, 2003.

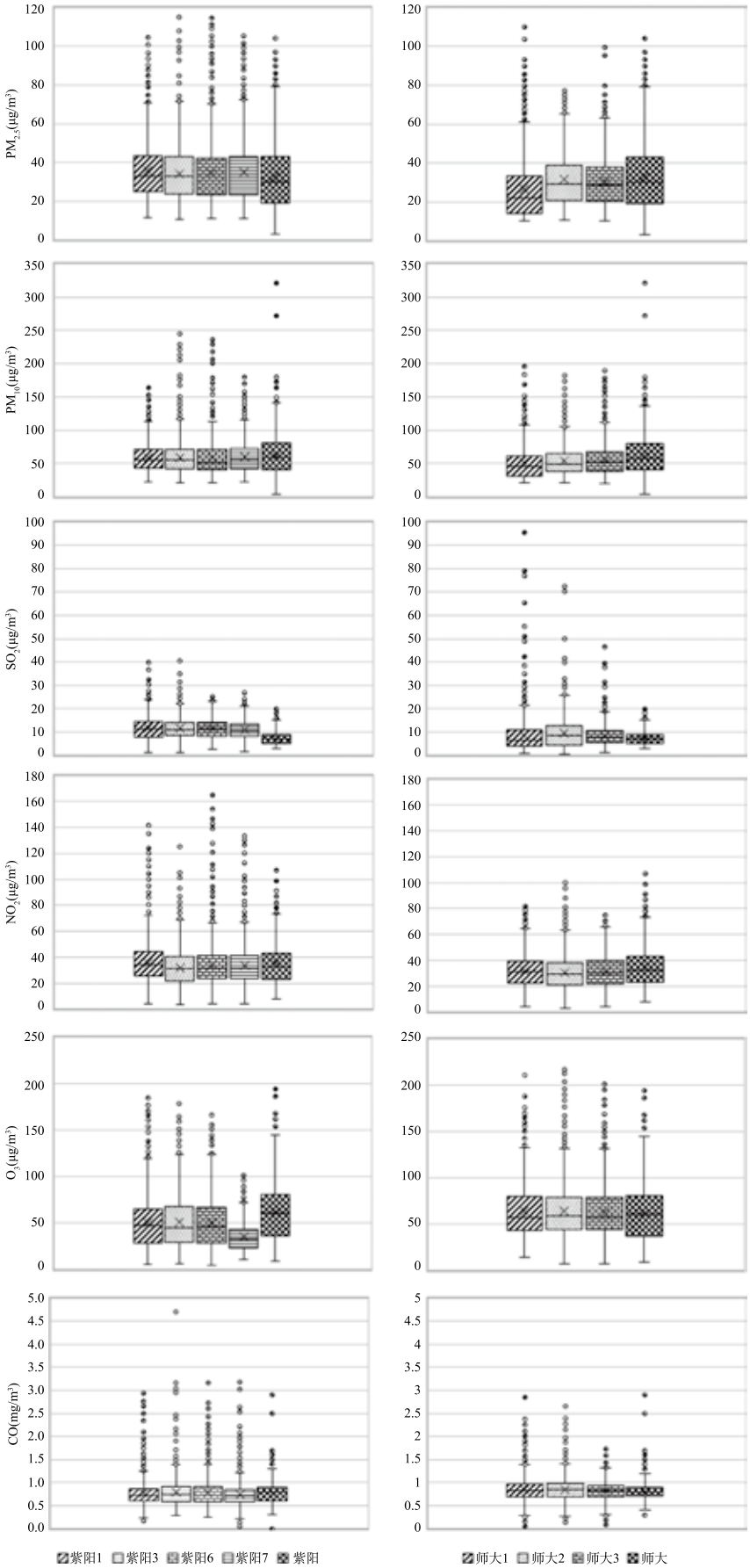


图3　紫阳（左）、师大（右）与周边小型站数据对比分析

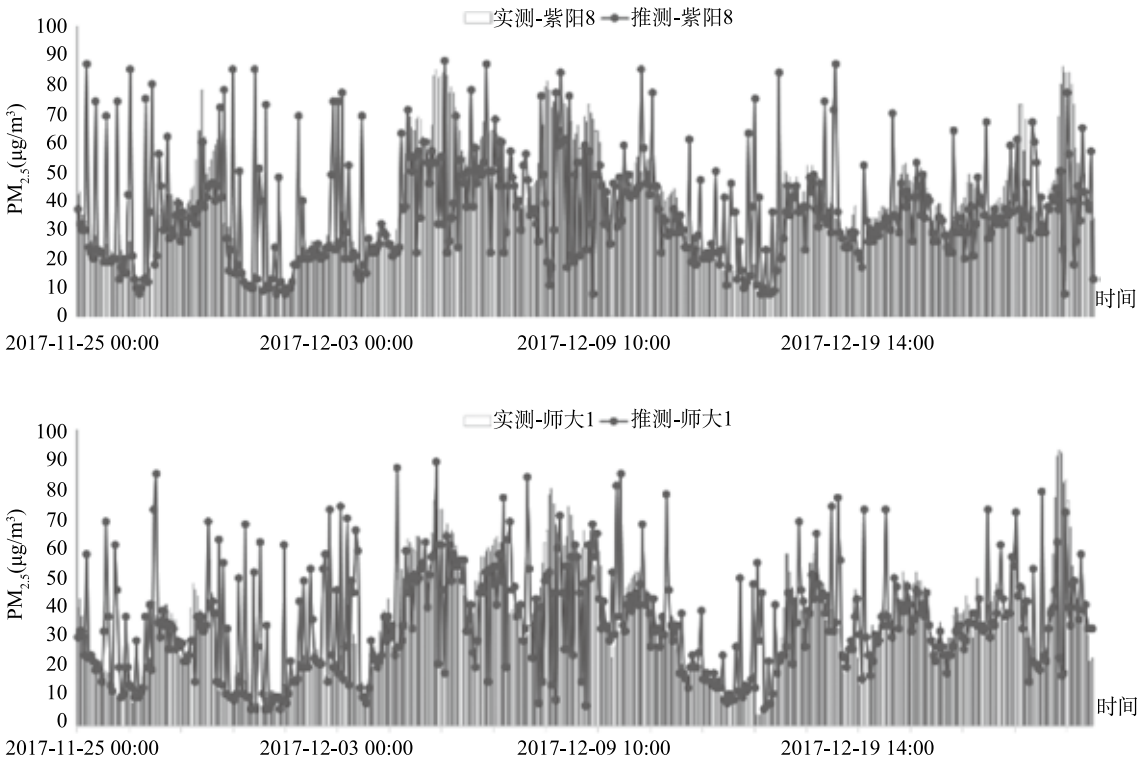


图4 站点小时预测与实际监测结果对比

The Real-Time Refined Simulation and Prediction for Air Quality in Hercynian Key Cities as Fuzhou

HU Qinghua[1*], LU Chen[2], HU Qian[2], WEI Shuzhen[1], JIANG Dongsheng[1], HUANG Yanyan[1]

(1. Fujian Environmental Monitoring Center, Fuzhou 350003, China;

2. Tshinghua-Solution Information Technology Co., Ltd., Beijing 100085, China)

Abstract: Based on existing air quality monitoring data, this study established a refined simulation and prediction system of urban air quality in Hercynian key cities as Fuzhou, using some multi-functional, efficient, small and exquisite sensors for monitoring and encrypting. The theories and mechanisms of big data and numerical air quality model were also applied in this study. Using 1 km × 1 km fine-grained grids, the result effectively improved the resolution of the Pollutants concentration difference among different locations in the same city. At the same time, this study also gave a forecast of the growing trend of urban air pollution for short time period. The result provided meaningful references for finding out the air pollution causes, to facilitate the environmental management and citizens' activities.

Keywords: Hercynian; air quality; real-time refinement; simulation and prediction