

空气质量预测模拟技术演变与发展研究

卢亚灵^{1,2,3}, 李勃^{1,2}, 范朝阳⁴, 王建童^{1,2}, 张鸿宇^{1,2,3}, 蒋洪强^{1,2*}

(1. 生态环境部环境规划院国家环境保护环境规划与政策模拟重点实验室, 北京 100012;
2. 生态环境部环境规划院京津冀区域环境联合研究中心, 北京 100012; 3. 天津大学环境科学与工程学院,
天津 300072; 4. 南开大学环境科学与工程学院, 天津 300350)

【摘要】 近年来我国的环境问题日益引人注目, 大气污染问题显得尤为突出。空气质量预测技术方法能够提前预测区域空间内的大气污染物浓度, 其发展十分迅速。本文介绍了一些国内外常用的空气质量预测技术方法的原理及案例, 对其结果与局限性进行归纳总结, 并对其未来发展提出展望。空气质量预测技术方法分为两类: 数值模拟和统计学习, 数值模拟通常可以分为第一代、第二代和第三代空气质量模型, 统计学习可以分为简单经验统计和机器学习。两类方法的目的都是尽可能真实、准确地实现特定时间、特定区域范围内大气污染物浓度的预测, 但两类方法的原理算法差异较大。该研究系统梳理了空气质量预测技术方法演变的历程和发展的现状, 展望了其发展趋势, 分析了建立、健全空气质量预测技术标准规范体系的重要性, 提出了相关参考建议。

【关键词】 空气质量; 浓度预测技术; 演变; 发展

【中图分类号】 X169

【文献标识码】 A

【文章编号】 1674-6252 (2021) 04-0084-09

【DOI】 10.16868/j.cnki.1674-6252.2021.04.084

引言

大气污染对生态系统和人类正常生存与发展造成严重危害^[1,2]。大气污染物主要包括气态物质和颗粒物两类^[3], 大量的研究表明, 短时间暴露在高浓度大气污染物中会引起眼部不适、呼吸困难、心肺疾病和早产婴儿夭折等严重的短期健康风险^[4,5], 长时间暴露极有可能引发癌症, 会给身体免疫系统、神经系统和呼吸系统带来永久健康风险^[6,7]。随着工业化、城镇化的快速发展, 大气污染治理面临严峻挑战, 近年来, 政府和公众对空气质量实时、准确和快速预测的需求日益增加。空气质量预测技术对大气污染物在给定时间空间范围内的浓度进行科学预测, 以支撑开展工程减排、结构减排和管理减排等措施, 达到降低大气污染物浓度的目的; 实时的空气质量预测技术可以通过减少污染物排放和提升健康效应, 为个人、组织、社区带来巨大的社会经济价值^[8]。空气质量预测技术主要可以分为两类——数值模拟和统计学习, 数值模拟通常可以分为第一代、第二代和第三代空气质量模型, 统计学习可以分为简单经验统计模型和机器学习。两类方法的目的都是尽可能真实、准确地实现特定时间、特定区域范围内大气污染物浓度的预测, 但

两类方法的原理算法差异较大。数值模拟从大气环境基础理论出发, 以高斯模型、拉格朗日模型和欧拉模型等为内核, 以气象场为驱动, 构建模拟真实大气环境变化的系统, 从而预测大气污染物浓度的时空变化规律; 统计学习从统计理论出发, 以回归、分类、拟合和过滤等算法为内核, 以历史数据为驱动, 寻找多源异构数据变化的规律, 从而预测大气污染物浓度的时空变化规律。一直以来, 数值模拟方法使用更广泛, 但近年来随着人工智能技术的崛起, 机器学习、深度学习在空气质量预测技术领域的应用已逐渐成为研究热点。

1 空气质量预测技术发展历程

1.1 国际发展历程

1955年6月, 为防止大面积空气污染事件的发生, 洛杉矶地区空气质量管理局(LACAPCD)针对洛杉矶重度烟雾污染建立了O₃和其他3项污染物的三级烟雾预测预警, 这被认为是实时空气质量预测技术的开端^[9]。20世纪60年代, 美、日、德等发达国家均开始了关于空气质量预测技术的理论和应用研究, 并在随后的几十年中衍生出数十种空气质量预测模型,

资助项目: 国家自然科学基金项目(91846301); 环境规划院青年基金项目(40050701)。

作者简介: 卢亚灵(1984—), 女, 硕士, 副研究员, 研究方向为环境规划与大气污染防治, E-mail: luyi@caep.org.cn。

*** 责任作者:** 蒋洪强(1975—), 男, 博士, 研究员, 研究方向为环境规划与生态环境大数据治理, E-mail: jianghongqiang@caep.org.cn。

多以高斯模型、拉格朗日模型或欧拉模型为理论基础,形成了一套完整技术体系暨嵌套气象场模拟—网格化排放清单制作—空气质量浓度模拟,并完成了三个代际的迭代,预测范围、预测时序和预测精度也随之逐步提高。另外,统计学习理论在空气质量预测领域也得到广泛应用,逐步成为空气质量模型技术体系中的第2条分支。近年来以数值模拟为基础,结合统计学习的集成模型逐渐成为研究热点,成为空气质量模型技术体系中第3条分支。据此,该领域形成了以数值模拟为主,统计学习和集成模型并驾齐驱的空气质量模型技术体系。

(1) 20世纪60—70年代。模型发展初期,第一代城市/区域尺度空气质量模型如ISC、AERMOD、ADMS等高斯模型初步建立并发展,此后研发的ADMS、AERMOD、CALPUFF等模型的理论内核也是这一阶段的成果。20世纪60年代,美国天气管理局(USWB)第一次通过运用数值天气预报模型(NWP)对大气停滞和大气污染潜发性进行了预测。随后美国国家海洋与大气管理局(NOAA)在1965年成立,第二年25个国家代表齐聚伦敦参加第一届世界清洁空气代表大会,商讨关于大气污染和清洁空气等议题^[8],空气污染问题逐步走进公众视野,空气质量预测技术的研究热度逐渐上升。随着美国国家环境局(U.S. EPA)等一系列机构的成立,一些基于历史空气质量数据构建的空气质量预测经验统计模型也得到发展^[9-11],其中20世纪70年代早期诞生了第一代城市/区域尺度空气质量模型,20世纪70年代晚期诞生了第一代全球尺度空气质量模型。第一代模型主要以扩散参数曲线为基础,采用简单的、参数化的线性机制描述复杂的大气物理过程,对于模拟惰性污染物的长期平均浓度有较好的效果^[12]。

(2) 20世纪80—90年代。20世纪80年代开始,随着大气边界层中非线性光化学反应研究的推进^[13,14],对于空气质量预测精度的更高需求推动了第二代短期和长期全尺度(城市—区域—全球)大气污染物浓度3D数值模拟模式的发展,并在20世纪90年代早期成功应用于短期O₃的模拟预测。然而很多国家对于空气质量预测的需求不断增加,因此在建立覆盖全国的大气污染物浓度监测、预测网络的推动下,空气质量预测模型在20世纪90年代中期进一步更新迭代,形成了第三代短期和长期全尺度3D数值模拟模型。在相关国家及权威机构的支持下,大气污染物预测技术得到更加广泛的应用,例如U.S. EPA修改了AQI指数并建立了AIRNow项目,向公众提供实时空气质

量预测结果,随后还发布了关于开展O₃浓度预测的指南;加拿大气象服务局(MSC)针对加拿大东部地区开展了空气质量预测;剑桥大学首次使用大气污染浓度预测模型支撑了两个平流层局地观测实验的规划应用。

(3) 21世纪至今。进入21世纪,美国通过相关法案与国会授权等政策手段大力推动了第三代3D数值模拟模式的发展,并通过NOAA和U.S. EPA的科研活动进一步推广模型的应用。2002年,NOAA在美国能源政策法案(the Energy Policy Act)和国会的授权支持下与U.S. EPA合作开发覆盖全美的3D数值模拟模式,并把预测结果发布给各州和其所属的预测机构^[15-22];2003年,U.S. EPA发布了关于开展PM_{2.5}浓度预测技术的指南;2004年NOAA资助了新英格兰地区开展O₃和PM_{2.5}浓度预测集成模拟^[18,19],同年由NOAA和U.S. EPA联合开发的国家实时空气质量预测系统(NRT-AQFC)正式部署运行,该系统是第一个由CMAQ模型构成的大气污染浓度预测业务化平台^[20];2005年欧盟开发了FUMAPEX UAQIFS模型,该模型耦合了城市气象、空气质量和人口暴露模型,并在欧洲6个城市得到业务化运用;2006年有学者开发了基于神经网络的PM₁₀预测模型^[23];2007年科隆大学首次把大尺度反演模型和复杂3D数值模拟模型集成起来进行大气污染浓度预测。

1.2 国内发展历程

相较于国际而言,国内空气质量模型技术的研究起步较晚,直到20世纪90年代才开始引进空气质量预测技术等相关概念,学习使用第一代、第二代、第三代大气污染物浓度数值模拟模型,但国内空气质量模型应用研究发展迅速,中国科学院大气物理研究所、清华大学和南京大学等各类型研究、教学机构都开发了自主的空气质量预测模型系统,并通过大量实践和研究开展了模型技术文件和参数化方案的本地化,逐步开展法规化空气质量模型基础数据、技术文件、参数化方案等标准化研究,逐步形成我国自主的空气质量模型技术体系。

(1) 20世纪60—80年代。1973年我国召开第一次全国环境保护会议,随后成立了原国务院环境保护领导小组,大气污染问题逐步得到重视^[21]。因环境管理和决策的重大需求,大气环境容量^[22]与空气质量预测^[12]研究逐步发展起来。这个阶段我国大气环境领域主要针对大气环境容量、光化学污染机理、酸雨污染机理和消除烟尘污染技术等几个方面开展了大量

理论和实证研究,为日后开展空气质量预测奠定了良好的理论基础。

(2) 20世纪90年代至今。90年代我国煤炭消费量大幅上升,SO₂排放量和大气中SO₂浓度随之快速上升,长江以南出现大面积酸雨区,大气污染防治主要对象为SO₂和悬浮颗粒物^[21]。大气环境问题的加重促进了相关标准的制定及污染物监测水平的提升,1996年对《大气环境质量标准》进行第一次修订,将可吸入颗粒物(PM₁₀)纳入常规污染物^[12]。随着监测手段的完善,精确的监测数据保证了我国空气预测模型得到进一步发展,针对模型普适性检验的大量研究为之后的模型改进打下了基础。如孙大伟^[24]应用ADMS模型计算了朝阳市在不同气象条件下的SO₂日均浓度,结果显示ADMS模型可以得到与实测浓度一致的结果;杨洪斌等^[25]在沈阳开展了AERMOD空气扩散模型的应用和验证。此外,不同地区的科研人员也致力于方法和模型改良,根据当地气象特点或污染物种类的选择改进模拟方法和模型,使模拟更具针对性,区域模型的精度不断提高。梁秀婷等^[26]根据呼和浩特市1994—1998年5年间每天的空气质量级别和气象历史资料创立了一个半定量的大气污染潜式预报模型,通过与同时期5个监测点TSP和SO₂平均浓度对比,准确率高达82.7%。这种预测模式的优点是计算量小,对特殊天气的预报精度较高,但由于在建模的时候不会考虑污染源的位置及污染物排放量,可能会导致对具有潜在高污染的地区预测误差率较高。

早期的统计预测模式探索了污染物与污染物、气象要素之间的相关性,滕浩宇^[27]根据南京市和吉林市2014年的日均观测数据,建立了基于统计方法的PM_{2.5}与温度、风力、气象因素及常规污染物之间的多元回归模型,通过模拟预测发现PM_{2.5}与风力、天气、温度、CO、NO₂、PM₁₀、SO₂等因素具有相关性,其中CO浓度与PM_{2.5}的浓度呈现极大的相关性。大尺度的污染源解析也是空气质量模型领域的一个热点,Hao等^[28]利用ISC3模式确定了北京市机动车排放对空气质量的贡献,薛志刚等^[29]利用ISC3模式分析了电厂脱硫后的大气环境影响。高怡等^[30]、朱凌云等^[31]和王书肖等^[32]应用Model-3/CMAQ模型对不同区域、不同时期和不同减排情景下,不同污染物的浓度模拟进行了应用和研究,取得较好的效果。韩素芹等^[33]、周广强等^[34]和庞杨等^[35]应用WRF-Chem模型对京津冀、长三角等地大气污染物浓度进行数值模拟。

进入21世纪我国结合独特地域条件,创立多个

数值预测模型,如气象科学院研发的CAPPS和中国科学院建立的基于RADM模型的高分辨率对流层化学模式HRCM。生态环境部环境规划院区域空气质量实时模拟与管控平台集成了WRF气象模型和CMAQ、CAMx空气质量模型,对全国338个城市空气质量进行预测预报。中国科学院大气物理研究所研发了RAQM和NAQPMS,目前上海、北京、深圳等城市已经开始应用以NAQPMS为代表的嵌套网格空气质量预报模式,并对中国中东部地区雾霾和东亚酸雨中和效应的时空分布进行模拟分析^[36,37]。南京大学开发了城市空气质量数值预报模式系统(NJU-CAQPS),并引入气溶胶模块,实现冬夏两季PM_{2.5}浓度时空分布特征的模拟预测,并依托该模型以南京市为研究对象,对城市尺度空气质量和灰霾预测开展大量研究^[38,39]。清华大学、华南理工大学受EPA委托,开发了空气质量达标评估工具(SMAT-CE)和空气污染控制健康效益评估工具(BenMAP-CE),可以对不同污染减排情景下的空气质量点位是否满足空气质量达标要求和健康效应进行量化评估。中山大学依托大气科学、海洋科学领域的学科优势对局部地区不同污染源对空气质量的贡献与影响机制开展了相关研究^[40,41]。

2 空气质量预测技术方法特点

2.1 数值模拟

(1) 第一代模型。第一代模型诞生于20世纪70年代早期,并在70年代晚期从城市/区域尺度发展到全球尺度。第一代模型主要是基于质量守恒定律的箱式模型、基于湍流扩散统计理论的高斯模型和拉格朗日轨迹模型^[12],其中高斯模型有ISC、AERMOD和ADMS等,拉格朗日轨迹模型有EMKA和CALPUFF等。第一代模型具有运算速度快、算力要求小、易于操作和长期模拟浓度高的特点,现在仍然得到广泛的运用;但其结构简单,对于大气环境中大气化学、大气物理过程的模拟效果较差,无法满足短期、高精度大气污染物浓度模拟的需求。

(2) 第二代模型。第二代模型诞生于20世纪80—90年代,其主要是基于梯度输送理论的欧拉模型,使用了网格系统把模拟区域分为三维网格单元,针对每一个网格单元进行模拟与计算,主要有CIT、UAM、RADM、ROM和ADMO等。第二代模型相比第一代实现了更精细化的空气质量预测,针对每个三维网格单元考虑了复杂的大气化学、大气物理过程,模拟结果也更准确;但其结构更复杂,并且只考虑大

气环境中单一的大气化学、大气物理过程,没有考虑对于不同大气污染物本身之间的产生、反应和消除机制,仍有一定局限性。

(3) 第三代模型。第三代全尺度模型于 20 世纪 90 年代诞生,其主要是在第二代模型的基础上构建了基于“一个大气”理念的多模块集成、多尺度网格嵌套三维欧拉模型^[12],模型突破了单一物种、单一模块的限制,针对多尺度网格嵌套单元进行模拟与计算,主要有 Model-3/CMAQ、CAMx、WRF-Chem 和 NAQPMS 等。在应用过程中,4 阶离散变分方法、卡尔曼滤波和集成模拟等方法被广泛地与模型进行耦合来提升大气污染浓度预测的效率。Elbern 等首次使用 4 阶离散变分方法对 O₃ 和 NO₂ 观测数据进行资料同化,并用作大气污染浓度预测模型输入^[42]; Vautard 等首次把 CHIMERE 和 NWP 模型集成起来进行 O₃ 浓度预测^[43]; Doraiswamy 等使用 CMAQ 集成耦合 WRF 和 MM5 模型对纽约开展 O₃ 和 PM_{2.5} 浓度预测;取得了较好效果,并发现采用权重和偏差校正后可以显著提升预测结果的准确性^[44]; San José 等尝试耦合计算流体力学模型与大气污染浓度预测模型,并在水平方向上建立分辨率为 1~10m 的模拟网格^[45,46],很多研究开始关注于耦合大气污染浓度预测数值模型与统计模型,以提高大气污染浓度预测的准确性^[47-50]。第三代模型精细化程度进一步提高,针对每个三维网格单元考虑其复杂的大气化学、大气物理过程,可以实现特定时间、特定区域实时大气污染物浓度模拟,但其结构复杂,算力要求高,操作难度大。

2.2 统计学习

自 20 世纪 70 年代以来,统计模型一直被运用在空气质量预测领域,包括多元线性回归、分类树、人工神经网络、模糊逻辑和卡尔曼滤波器等多种算法,统计模型一般需要大量的历史数据作为模型输入,不同的算法可以解决不同的空气质量预测问题。在统计学习中一类是简单经验统计,另一类是机器学习。机器学习是在简单经验统计的基础上,通过算法实现计算机模拟人类学习行为,具有重新组织已有知识结构使之不断改善自身的性能,相比简单经验统计,机器学习具有自学习的特点,在空气质量预测技术领域,通过大量历史数据的输入,训练计算机找到历史数据特定区域、特定时间内的变化规律,并以此作为依据对未来空气质量进行预测。

2.2.1 简单经验统计

简单经验统计方法主要有 3 种:持续模型(persistence)、气候学模型(climatology)和经验模型(empiricism)。持续模型基于一个假设,即今日大气污染物浓度监测值等于明天的预测值^[49],模型输入需要前一天的大气污染物浓度监测值,持续模型具有计算快、大气环境稳态下精度高、使用便捷和成本低等优点,但无法适用于天气、排放和空气质量等因素突变下的预测情景,一般只作为其他预测方法的参考使用。气候学模型与持续模型很相似,但主要基于大气污染物浓度与气候学特征高度相关的假设,该模型需要数年大气污染物浓度和气候学历史数据,优缺点与持续模型类似,无法对短期、突变的情景进行有效预测。经验模型是在气候学模型的基础上,进一步建立气象要素的临界值与大气污染物浓度之间的关系,即达到某一气象要素的某一临界值表示某一大气污染物到达某一浓度,需要气象要素的历史数据和预报数据。这种模型具备其他简单经验统计方法的优点,应用十分广泛,如 Wolff 等^[21]在 1978 年开发了经验统计模型用于预测美国东北地区臭氧最大日浓度,而且准确度有所提高,是一种有效预测重污染天气的方法。但该模型无法预测大气污染物浓度的具体数值,无法预测多种复杂气象条件下高敏感度污染物的浓度。

2.2.2 机器学习

(1) 分类回归树(CART)。CART 是决策树的一种实现,是一种二分递归分割技术,把当前样本划分为两个子样本,使得生成的每个非叶结点都有两个分支,因此 CART 算法生成的决策树是结构简洁的二叉树。CART 算法主要分为两个步骤,包括将样本递归划分进行建树过程,以及生成决策树后用验证数据进行剪枝的过程。在空气质量预测中,通过选取特征,计算每个特征的基尼系数,选择基尼系数最小的特征及其对应的切分点作为最优特征与最优切分点,然后再重复以上计算步骤,直到叶节点满足停止条件。Burrows 等^[22]开发了基于决策树的夏季地表臭氧浓度统计预测模型,通过输入多年气象站点数据,决策树模型找到近地层中与夏季地表臭氧浓度具有高度相关性的气象因子,以达到准确预测夏季地表臭氧浓度是否超过最大地表臭氧浓度限值标准,此模型在加拿大多地得到实证应用。分类回归树具有计算快、分类准确、算力要求低等优点,也能保证一定的准确度,但无法准确预测大气污染物浓度极值,在观测数据受限

的情况下模型使用也同样受限。

(2) 回归模型 (Regression)。回归模型从一组数据出发, 确定某些变量之间的定量关系式, 建立数学模型并估计未知参数, 回归的目的是预测数值型的目标值。它的目标是接受连续数据, 寻找最适合数据的方程, 并能够对特定值进行预测。常见的回归模型包括线性回归、局部加权线性回归、岭回归和逻辑回归等。Coburn 等^[51]基于多元非线性回归开发了臭氧浓度预测模型, 使用五年历史数据对模型进行校验以后, 模型对日 1 小时最大臭氧浓度的预测效果得到较好提升, 平均绝对误差 (MAE) 稳定在 21.7~12.1 ppb^①。回归模型具有计算快、易操作、运行成本低等优点。但空气质量预测中存在大量简单线性回归模型无法模拟的非线性问题, 且回归模型通常需要大量、连续的历史数据, 对数据数量、质量要求较高。

(3) 人工神经网络 (ANNs)。人工神经网络是基于生物学中神经网络的基本原理, 在理解和抽象了人脑结构和外界刺激响应机制后, 以网络拓扑知识为理论基础, 模拟人脑的神经系统对复杂信息的处理机制的一种数学模型, 也被称作深度学习模型。它实际上是一个由大量简单元件相互连接而成的复杂网络, 具有高度的非线性, 能够进行复杂的逻辑操作和非线性关系实现的系统。神经网络系统分为三层结构: 输入层、隐含层和输出层, 其主要的处理过程是在输入层输入样本数据, 在隐含层处理数据, 在输出层得到结果, 如果误差过大则需要调整神经网络的权值返还到隐含层里继续处理, 直至结果符合预期。在空气质量预测中, 人工神经网络模型与传统的预测模型相比具有自适应学习能力, 可以有效处理大气环境中的非线性问题, 而且其计算成本较数值模拟更低; 但人工神经网络对于污染物浓度极值模拟效果较差, 其模拟准确度直接与计算成本相关, 计算成本投入越多准确度越高, 而且在建立模型的过程中会出现无法收敛的问题。其中基于长短期记忆循环神经网络 (LSTM) 的时间序列预测是一类重要应用, LSTM 是一类特殊的递归神经网络, 其核心特点是用上一时刻学习到的信息进行当前时刻的学习, 因此对于空气质量数据的时间序列预测效果较好。近年来相关研究逐渐增多, 如 Perez 等^[23]开发了一种基于人工神经网络的统计预测模型, 对 PM₁₀ 浓度进行了预测, 并在智利得到实证应用; VIOTTI 等^[52]和 Sang^[53]等利用 BP 神经网络预

测模型对空气中的臭氧进行浓度预测, 结果表明这种模型对空间浓度的变化趋势预报准确。随着人工智能技术的崛起, 近年来基于人工神经网络的空气质量预测再度成为研究焦点, 其可能成为解决“情景减排—浓度预测—来源解析—响应措施—减排落地”分钟级响应的关键技术。

(4) 模糊逻辑 (FL)。模糊逻辑模型是一种让人工智能按照人的思考方式去考虑推理模糊概念问题的模型。通过模糊集合, 使得一个变量可能同时属于多个集合, 每个集合部分占有这个变量, 将一个具体的离散值模糊化, 再通过许多不同角度的模糊规则去解释模糊集合, 使模糊集合中的变量计算得到一个确定的值, 达到去模糊化的目的。模糊逻辑模型在空气质量预测中运用较少, 部分国外学者通过引用模糊克里金模型中的半统计隶属函数对空气质量进行预测, 并结合进化遗传算法 (GAs) 匹配不同污染物与不同函数关系, 以达到提高模型预测效率的目的^[54,55]。该模型与人工神经网络特点类似。

(5) 卡曼滤波 (KF)。卡曼滤波是一种对含有不确定信息的动态系统走向做出有根据的预测的模型。在连续变化的系统中使用卡尔曼滤波是非常理想的, 它具有占用内存小的优点, 并且速度很快, 很适合应用于实时问题和嵌入式系统。Chenevez 等^[56]在丹麦气象学院的资助下开发了基于卡尔曼滤波器的臭氧浓度预测模型, 并于丹麦气象学院数值模型 (DMI-HIRLAM) 结合实现空气质量预测, 该模型的回归方程通过 6 个空气质量监测站点数据校验, 形成了 4 类回归方程组合, 预测效果较好, 并在哥本哈根得到应用。卡曼滤波模型可以在考虑不确定性的基础上给出准确的预测, 特别是基于时间序列的空气质量预测, 其缺点是对于高度非线性的系统模拟效果较差。

2.3 集成模型

在数值模拟和统计学习的基础上, 有学者尝试把两类方法结合, 以实现快速、准确的实时大气污染物浓度模拟。这类模型主要是通过统计学习的方法对数值模拟结果进行评估、修正, 大量的数值模拟历史数据可以作为统计模型的输入。

国内相关研究从 20 世纪 90 年代开始逐渐起步, 王芳等^[57]构建的北京市 PM₁₀ 的遗传神经网络预测模型的预测精度与美国第三代空气质量模型 CMAQ 进行比较, 结果表明两者预测结果精度相当; Zhu^[58]等

① ppb (part per billion), 1ppb=10⁻⁹, 即十亿分之一, 是一个无量纲量。

使用 ARMA 预测 $PM_{2.5}$ 的浓度, 并使用 BP 神经网络模型学习 ARMA 预测的误差, 最终 $PM_{2.5}$ 浓度的预测值即为二者加和。这些基于数值预测构建的模型科学性极强, 能够对空间污染物浓度的时间和空间进行定量分析, 但是需要很多相关数据, 且计算困难, 同时海量数据的输入也需要大量的时间。经过一段时间的耦合, 可以通过统计模型替代一部分数值模拟运算, 以达到减少数值模拟次数、降低运算成本的效果。清华大学王书肖教授课题组^[59]借助统计方法对“排放—浓度”非线性关系进行了显示表征, 建立了高阶多项式的响应曲面模型 (pf-RSM), 将化学传输模式 (CTM) 仿真次数减少 60%。基于该成果, 王书肖课题组与微软亚太研究院^[59]合作采用卷积神经网络的方法 (CNN) 建立了基于深度学习方法的响应曲面模型 (deepRSM), 该研究基于两个 CTM 模拟的浓度空间分布建立各网格点“排放—浓度”响应函数, 大大提高了计算效率, 扩展了该模型的应用范围。

3 存在问题分析

(1) 空气质量预测技术发展迅速, 技术门槛不断提高, 使用成本不断增加。数值模拟是空气质量预测技术的主流方法, 其经过三代模型迭代, 已形成成熟的技术体系, 即嵌套气象场模拟—网格化排放清单制作—空气质量模拟, 气象场是模型的驱动场, 网格化排放清单把各类污染源排放按照一定机制分配到网格中, 污染物浓度模拟是使用模型对大气环境中的复杂大气物理、大气化学反应进行模拟。随着计算机技术的发展, 空气质量预测技术门槛不断提高, 构建一套大气污染物浓度数值模拟与预测技术体系需要完成模型集成部署、代码编译、参数化方法调试、模型运行和结果验证等步骤, 以上均需要在基于 Linux 系统的高性能计算服务器上开展, 是一个计算机科学、大气科学和环境科学等多学科交叉的领域, 对于模型使用者的综合能力要求较高; 基于统计理论的机器学习等方法, 对于大气环境的复杂反应机理知识要求不高, 但要求模型使用者可以有机地结合统计与算法理论, 正确、准确地使用、调试各类机器学习算法, 从数据的角度发现历史数据分布、发展、运行的规律, 从而对未来进行模拟预测。机器学习模型的性能与超参数直接相关, 超参数是在建立模型时用于控制算法行为的参数, 超参数调优的越好, 得到的模型就越好, 超参数调整的过程不确定性较大, 不同参数组合下的模型运行结果往往无法重复。另外, 决策机构和管理机构对于空气质量预测的速度和准确度需求不断提高,

对于数值模拟来说是投入更多、更快的高性能计算机服务器, 对于统计学习来说是输入大量的历史数据、使用更复杂的迭代算法, 以上都会极大地增加空气质量预测的成本。

(2) 原始数据多源异构, 模型输入数据欠缺标准化。对数值模拟和统计学习方法来说, 多源异构的原始数据都会造成模拟结果不具有可比性。数值模拟方法中网格化清单的制作需要收集区域、行业、部门的大气污染物排放量数据, 其网格化过程中需要收集交通、能源、产业和自然资源等社会经济数据。目前排放量数据和社会经济数据数出多门, 原始数据多源异构, 根据原始数据制作的网格化排放清单也就有所差异。例如, 中国环境科学研究院、清华大学和中国科学院大气物理研究所等机构均有独立的排放清单数据, 同一研究区域内使用不同排放清单数据所得到的大气污染物浓度模拟结果往往差异较大。对于统计学习方法来说, 统计学习通常把长周期气象数据和大气污染物监测数据作为输入数据, 多源异构的原始数据作为输入数据投喂给模型进行训练所引起的模型预测结果波动大, 同一学习拟合过程难以重复。

(3) 模型参数化方案缺少技术规范指导, 预测模拟结果可比性差。数值模型方法需要针对给定区域和给定时间的模型选取不同的参数化方案, 目前国内很多研究在开展空气质量预测时对于参数化方案的选择仅限于通用方案。我国大气污染物浓度监测网络的建设近年取得显著成果, 但 $PM_{2.5}$ 、 PM_{10} 和 O_3 等主要大气污染物的历史监测数据还比较欠缺, 缺乏长周期的观测数据对于参数本地化造成较大困难。参数本地化的工作开展也较少, 生态环境管理部门也未针对参数化方案的选取给出相应的技术规范指导。通过数值模拟方法开展空气质量预测还处在“百家争鸣”的阶段。统计学习算法模型中有大量参数需要调试, 我国目前对于使用机器学习、深度学习等经典算法开展空气质量预测的核心参数率定的研究还较少, 大量研究的参数方案无法复制。

(4) 模拟结果评估指标和方法不一, 模拟结果业务化程度低。针对数值模拟和统计学习两类方法的预测和模拟均可以采用离散统计和分类统计来进行统计学评估, 以达到理想的模拟预测效果。目前国内研究所采用的统计评估指标多达数十种, 离散统计包括相关系数、平均偏离 (MB)、平均绝对总误差 (MAGE), 分类统计包括准确率、偏离率、虚警率等。评估指标是决定模拟预测结果的重要手段, 直接决定预测模拟结果是否科学、有效。不同统计指标评

估的预测结果难以具有可比性,导致模拟结果业务化的程度较低。

(5) 模型技术文件与工具解读不一,缺乏本土化、标准化模型辅助技术文件。模型技术文件与工具包是模型使用最重要的参考资料,目前针对空气质量模型的技术文件和工具包大多是基于国外原版文件的翻译,对于模型本土化的应用缺乏本土化、标准化的技术文件和工具包支撑,某种程度会导致模型“水土不服”、模拟结果误差较大和结果难以重复等问题。另外,不同科研工作者对于模型技术文件和工具包的使用解读也存在偏差,造成模型在不同领域、不同专业的使用和推广有所差异,进而出现模型使用“百家争鸣”的局面。

4 展望与建议

(1) 推动机器学习等人工智能方法在空气质量预测领域的应用。统计学习方法一直在空气质量预测领域有所应用,近年来机器学习等人工智能方法在该领域的应用重新成为热点。机器学习和深度学习被看作满足分钟级快速、准确空气质量预测的解决方案之一。机器学习从统计理论和数据科学的角度出发,不考虑大气环境中大气化学等复杂反应机理,而是通过产品产量、资源消耗和气象要素大数据输入,通过算法寻找输入数据的分布规律,并对未来大气污染物浓度进行预测。因此计算速度大幅提升,计算成本大幅下降。机器学习方法的深化应用,是丰富空气质量预测方法的重要方向,也是服务生态环境管理部门科学、精准治污需求的重要手段。

(2) 推进搭建数值模拟和统计学习两类方法集成的预测技术体系。随着管理决策部门对于浓度预测更加精细化的需求,构建结合数值模拟与统计学习的空气质量预测技术方法体系是未来的发展方向。通过数值模拟实现对大气环境中大气环境复杂反应机制的模拟,以机器学习等人工智能方法进行评估和检验,缩短数值计算时间,降低计算成本。目前已经有很多学者开展了两类方法集成预测的技术方法,建议针对大气污染物浓度模拟的新形势,推进相关集成预测技术体系业务化落地。由于收集特定范围尺度内具有时间序列的数据是一大难点,因此如何将多源异构数据在空间、时间上进行融合并构建标准化排放清单也是未来研究需要解决的问题。

(3) 建立全面的空气质量预测方法技术规范体系。2020年生态环境部发布《环境空气质量数值预报技术规范》(HJ 1130—2020),该标准对环境空气质

量数值预报模式的基本要求、运算处理、效果评估等内容进行了规定,是该领域首次发布相关技术规范标准,是我国空气质量预测模拟业务化标准体系建设的重要开端。目前该标准仅针对数值模拟方法提出了相应的技术规范和指导,机器学习等新一代人工智能方法相关技术规范还是空缺,建议加快推进空气质量预测方法技术规范体系的构建,把数据输入、模型训练和结果评估等模拟预测过程标准化。

参考文献

- [1] 王文兴,柴发合,任阵海,等.新中国成立70年来我国大气污染防治历程、成就与经验[J].环境科学研究,2019,32(10):1621-1635.
- [2] 黄顺祥.大气污染防治的过去、现在及未来[J].科学通报,2018,63(10):895-919.
- [3] ZHANG Y, BOCQUET M, MALLET V, et al. Real-time air quality forecasting, part I: History, techniques, and current status[J]. Atmospheric environment, 2012, 60: 632-655.
- [4] GREENBAUM D S, BACHMANN J D, KREWSKI D, et al. Particulate air pollution standards and morbidity and mortality: case study[J]. American journal of epidemiology, 2001, 154(12 S): S78-S90.
- [5] World Health Organization. Health aspects of air pollution: results from the WHO project "Systematic review of health aspects of air pollution in Europe" [Z]. Copenhagen: WHO Regional Office for Europe, 2004.
- [6] GEORGOPOULOS P, ISUKAPALLI S, BURKE J M, et al. Air quality modeling needs for exposure assessment from the source-to-outcome perspective[Z]. Air & Waste Management Association. 2009: 26-35.
- [7] PHALEN R F, PHALEN R N. Introduction to Air Pollution Science: A Public Health Perspective[M]. Burlington, MA: Jones & Bartlett Learning, 2011.
- [8] GROSS E. The National Air Pollution Potential Forecast Program[R]. Suitland, MD: National Meteorological Center, 1970.
- [9] MCCOLLISTER G M, WILSON K R. Linear stochastic models for forecasting daily maxima and hourly concentrations of air pollutants[J]. Atmospheric environment, 1975, 9(4): 417-423.
- [10] Revlett G H. Ozone forecasting using empirical modeling[J]. Journal of the Air Pollution Control Association. 1978, 28(4): 338-343.
- [11] ARON R. Forecasting high level oxidant concentrations in the Los Angeles basin[J]. Journal of the air pollution control association, 1980, 30(11): 1227-1228.
- [12] 薛文博,王金南,杨金田,等.国内外空气质量模型研究进展[J].环境与可持续发展,2013,38(3):14-20.
- [13] RYAN W F. Forecasting severe ozone episodes in the Baltimore metropolitan area[J]. Atmospheric environment, 1995, 29(17): 2387-2398.

- [14] RYAN W F, PETTY C A, LUEBEHUSEN E D. Air quality forecasts in the mid-Atlantic region: current practice and benchmark skill[J]. Weather and forecasting, 2000, 15(1): 46-60.
- [15] STOCKWELL W R, ARTZ R S, MEAGHER J F, et al. The scientific basis of NOAA's air quality forecasting program[Z]. Environmental Managers. 2000: 20-27.
- [16] WAYLAND R A, WHITE J E, DICKERSON P G, et al. Communicating real-time and forecasted air quality to the public[Z]. EM-Pittsburgh-Air and Waste Management Association, 2002: 28-36.
- [17] DABBERDT W F, CARROLL M A, BAUMGARDNER D, et al. Meteorological research needs for improved air quality forecasting: report of the 11th prospectus development team of the U.S. Weather Research Program[J]. Bulletin of the American meteorological society, 2004, 85(4): 563-586.
- [18] MCKEEN S, WILCZAK J, GRELL G, et al. Assessment of an ensemble of seven real-time ozone forecasts over eastern North America during the summer of 2004[J]. Journal of geophysical research: atmospheres, 2005, 110: D21307.
- [19] MCKEEN S, CHUNG S H, WILCZAK J, et al. Evaluation of several PM_{2.5} forecast models using data collected during the ICARTT/NEAQS 2004 field study[J]. Journal of geophysical research: atmospheres, 2007, 112: D10S20.
- [20] OTTE T L, POULIOT G, PLEIM J E, et al. Linking the Eta model with the community multiscale air quality (CMAQ) modeling system to build a national air quality forecasting system[J]. Weather and forecasting, 2005, 20(3): 367-384.
- [21] WOLFF G T, LIOY P J. An empirical model for forecasting maximum daily ozone levels in the northeastern U.S.[J]. Journal of the air pollution control association, 1978, 28(10): 1034-1038.
- [22] BURROWS W R, BENJAMIN M, BEAUCHAMP S, et al. CART decision-tree statistical analysis and prediction of summer season maximum surface ozone for the Vancouver, Montreal, and Atlantic regions of Canada[J]. Journal of applied meteorology, 1995, 34(8): 1848-1862.
- [23] PEREZ P, REYES J. An integrated neural network model for PM₁₀ forecasting[J]. Atmospheric environment, 2006, 40(16): 2845-2851.
- [24] 孙大伟. 新一代大气扩散模型 (ADMS) 应用研究 [J]. 环境保护科学, 2004, 30(1): 67-69.
- [25] 杨洪斌, 张云海, 邹旭东, 等. AERMOD 空气扩散模型在沈阳的应用和验证 [J]. 气象与环境学报, 2006, 22(1): 58-60.
- [26] 梁秀婷, 宋进华, 高春香, 等. 呼和浩特市空气污染潜势预报研究 [J]. 内蒙古气象, 2002(2): 36-38.
- [27] 滕浩宇. 基于常规大气监测数据对 PM_{2.5} 的评估和预测 [D]. 哈尔滨: 东北林业大学, 2015.
- [28] HAO J M, HE D Q, WU Y, et al. A study of the emission and concentration distribution of vehicular pollutants in the urban area of Beijing[J]. Atmospheric environment, 2000, 34(3): 453-465.
- [29] 薛志钢, 柴发合, 段宁, 等. 运用 ISC3 模型模拟电厂脱硫后的大气环境影响 [J]. 环境科学研究, 2003, 16(5): 62-64.
- [30] 高怡, 张美根, 朱凌云, 等. 2008 年奥运会期间北京地区大气 O₃ 浓度模拟分析 [J]. 气候与环境研究, 2010, 15(5): 643-651.
- [31] 朱凌云, 张美根, 高丽洁, 等. 东亚地区硝酸盐湿沉降的数值模拟 [C]// 第 27 届中国气象学会年会. 北京: 中国气象学会, 2010: 1-5.
- [32] 王书肖, 陈瑶晟, 许嘉钰, 等. 北京市燃煤的空气质量影响及其控制研究 [J]. 环境工程学报, 2010, 4(1): 151-158.
- [33] 韩素芹, 冯银厂, 边海, 等. 天津大气污染物日变化特征的 WRF-Chem 数值模拟 [J]. 中国环境科学, 2008, 28(9): 828-832.
- [34] 周广强, 高伟, 谷怡萱, 等. WRF-Chem 模式降水对上海 PM_{2.5} 预报的影响 [J]. 环境科学学报, 2017, 37(12): 4476-4482.
- [35] 庞杨, 韩志伟, 朱彬, 等. 利用 WRF-Chem 模拟研究京津冀地区夏季大气污染物的分布和演变 [J]. 大气科学学报, 2013, 36(6): 674-682.
- [36] WANG Z F, LI J, WANG Z, et al. Modeling study of regional severe hazes over mid-eastern China in January 2013 and its implications on pollution prevention and control[J]. Science China earth sciences, 2014, 57(1): 3-13.
- [37] TERADA H, UEDA H, WANG Z F. Trend of acid rain and neutralization by yellow sand in east Asia—a numerical study[J]. Atmospheric environment, 2002, 36(3): 503-509.
- [38] 吕梦瑶, 刘红年, 张宁, 等. 南京市灰霾影响因子的数值模拟 [J]. 高原气象, 2011, 30(4): 929-941.
- [39] 欧阳琰, 蒋维楣, 刘红年. 城市空气质量数值预报系统对 PM_{2.5} 的数值模拟研究 [J]. 环境科学学报, 2007, 27(5): 838-845.
- [40] WANG X M, CHEN W H, CHEN D H, et al. Long-term trends of fine particulate matter and chemical composition in the Pearl River Delta Economic Zone (PRDEZ), China[J]. Frontiers of environmental science & engineering, 2016, 10(1): 53-62.
- [41] WANG X M, CARMICHAEL G, CHEN D L, et al. Impacts of different emission sources on air quality during March 2001 in the Pearl River Delta (PRD) region[J]. Atmospheric environment, 2005, 39(29): 5227-5241.
- [42] ELBERN H, SCHMIDT H. Ozone episode analysis by four-dimensional variational chemistry data assimilation[J]. Journal of geophysical research: atmospheres, 2001, 106(D4): 3569-3590.
- [43] VAUTARD R, BLOND N, SCHMIDT H, et al. Multi-model ensemble ozone forecasts over Europe: analysis of uncertainty[C]//BRANDT J, ed. Mesoscale Transport of Air Pollution. OA15. EGS XXXVI General Assembly. Nice. European Geophysical Society, Katlenburg-Lindau, Germany, France. 2001.
- [44] DORAISWAMY P, HOGREFE C, HAO W, et al. Preliminary experiences with the multi-model air quality forecasting system for New York state[C]//Presented at the 8th Annual Community Modeling and Analysis System (CMAS) Conference. Chapel Hill, NC, 2009.

- [45] SAN JOSÉ R, PÉREZ J L, GONZÁLEZ R M. CFD and mesoscale air quality applications in urban environments: Madrid case study[C]//Proceedings of the 4th WSEAS International Conference on Environment, Ecosystems, and Development (EED '06). Venice, Italy, 2006.
- [46] SAN JOSÉ R, PÉREZ J L, MORANT J L, et al. The use of Modern third-generation air quality models (MM5-EMIMO-CMAQ) for real-time operational air quality impact assessment of industrial plants[J]. Water, air, & soil pollution: focus, 2009, 9(1-2): 27-37.
- [47] EBEN K, JURUŞ P, RESLER J, et al. An ensemble Kalman filter for short-term forecasting of tropospheric ozone concentrations[J]. Quarterly journal of the royal meteorological society, 2005, 131(613): 3313-3322.
- [48] GUILLAS S, BAO J H, CHOI Y, et al. Statistical correction and downscaling of chemical transport model ozone forecasts over Atlanta[J]. Atmospheric environment, 2008, 42(6): 1338-1348.
- [49] KANG D W, MATHUR R, RAO S T, et al. Bias adjustment techniques for improving ozone air quality forecasts[J]. Journal of geophysical research: atmospheres, 2008, 113: D23308.
- [50] ROUÏL L, HONORÉ C, VAUTARD R, et al. Prev'air: an operational forecasting and mapping system for air quality in Europe[J]. Bulletin of the American meteorological society, 2009, 90(1): 73-84.
- [51] COBOURN W G, HUBBARD M C. An enhanced ozone forecasting model using air mass trajectory analysis[J]. Atmospheric environment, 1999, 33(28): 4663-4674.
- [52] VIOTTI P, LIUTI G, DI GENOVA P. Atmospheric urban pollution: applications of an artificial neural network (ANN) to the city of Perugia[J]. Ecological modelling, 2002, 148(1): 27-46.
- [53] SOHN S H, OH S C, JO B W, et al. Prediction of ozone formation based on neural network[J]. Journal of environmental engineering, 2000, 126(8): 688-696.
- [54] JORQUERA H, PÉREZ R, CIPRIANO A, et al. Forecasting ozone daily maximum levels at Santiago, Chile[J]. Atmospheric environment, 1998, 32(20): 3415-3424.
- [55] SHAD R, MESGARI M S, ABKAR A, et al. Predicting air pollution using fuzzy genetic linear membership kriging in GIS[J]. Computers, environment and urban systems, 2009, 33(6): 472-481.
- [56] CHENEVEZ J, JENSEN C Ø. Operational ozone forecasts for the region of Copenhagen by the Danish Meteorological Institute[J]. Atmospheric environment, 2001, 35(27): 4567-4580.
- [57] 王芳, 程水源, 李明君, 等. 遗传算法优化神经网络用于大气污染预报[J]. 北京工业大学学报, 2009, 35(9): 1230-1234.
- [58] ZHU H X, LU X H. The prediction of PM_{2.5} value based on ARMA and improved BP neural network model[C]//Proceedings of the 2016 International Conference on Intelligent Networking and Collaborative Systems. Ostrava, Czech Republic: IEEE, 2016: 515-517.
- [59] XING J, ZHENG S X, DING D, et al. Deep learning for prediction of the air quality response to emission changes[J]. Environmental science and technology, 2020, 54(14): 8589-8600.

Evolution and Development of Air Quality Prediction and Simulation Technology

LU Yaling^{1,2,3}, LI Bo^{1,2}, FAN Zhaoyang⁴, WANG Jiantong^{1,2}, ZHANG Hongyu^{1,2,3}, JIANG Hongqiang^{1,2*}

(1. State Environmental Protection Key Laboratory of Environmental Planning and Policy Simulation, Chinese Academy of Environmental Planning, Ministry of Ecology and Environment, Beijing 100012, China; 2. Environmental Research Center of Beijing-Tianjin-Hebei Region, Chinese Academy of Environmental Planning, Ministry of Ecology and Environment, Beijing 100012, China; 3. School of Environmental Science and Engineering, Tianjin University, Tianjin 300072, China; 4. College of Environmental Science and Engineering, Nankai University, Tianjin 300350, China)

Abstract: In recent years, China's environmental problems have become increasingly noticeable, and the problem of air pollution has become particularly prominent. The air quality prediction method can predict the concentration of pollutants in a certain area in advance, and its development is very rapid. This article introduces the principles and cases of some commonly used air quality prediction techniques, summarizes their results and limitations, and proposes prospects for future development. Air quality prediction methods are divided into two categories: numerical simulation and statistical learning. Numerical models can usually be divided into first, second and third generations. Statistical learning can be divided into simple empirical statistical methods and machine learning. The purpose of the two types of methods is to achieve the prediction of air pollutants in a specific time and within a specific area as truthfully and accurately as possible, but the principles and algorithms of the two types of methods are quite different. This research systematically sorts out the evolution progress and development status of air quality prediction technology, looks forward to the development trend of air quality prediction technology, analyzes the importance of establishing and perfecting air pollutant concentration prediction standards and regulations, and gives relevant reference suggestions.

Keywords: air quality; concentration prediction technology; evolution; development