

基于NARX 神经网络的 PM_{2.5/10} 浓度值预测模型 ——以咸阳市两寺渡监测站为例

张丹宁,张 猛,张 博 西安交通大学 人居环境与建筑工程学院,西安 710049

摘 要: PM_{2.5}和 PM₁₀(记为 PM_{2.5/10})对空气质量和人类健康有着严重威胁,日益引起国内外的关注, 并成为大气污染控制工程中最重要的部分。基于陕西省咸阳市两寺渡监测站的污染物(PM_{2.5}、PM₁₀、 NO₂、NO、NO_x、CO)和相关气象参数的监测数据,建立起基于非线性有源自回归神经网络的预测模 型,并分别针对不同预测时间段确定最优网络结构,从而实现了对未来6小时、12小时以及24小时 PM_{2.5/10}浓度的有效预测。实验结果表明:(1)NARX 神经网络模型可对未来24小时内的 PM_{2.5/10} 污染 物浓度进行较为准确的预测;(2)对于 PM_{2.5/10}未来6小时的预测能力优于对12小时、24小时的预测; (3)预测值偏高或偏低的结果与前后时间段内的气象因素及其他污染物浓度变化情况也具有相关性。 关键词:PM_{2.5}; PM₁₀; 空气质量; NARX; 递归神经网络; 大气污染预测

The model to predict PM_{2.5/10} concentrations based on NARX neural network—taking Liangsidu monitoring station in Xianyang as an example

ZHANG Danning, ZHANG Meng, ZHANG Bo

School of Human Settlements and Civil Engineering, Xi'an Jiaotong University, Xi'an 710049, China

Abstract: *Background, aim, and scope* As known, both $PM_{2.5}$ and PM_{10} (denoted as $PM_{2.5/10}$) are typical fine particles matters, which have serious threats to air quality and human health. To better understand the temporal variations of air quality, this research is dedicated to the prediction of $PM_{2.5/10}$ concentrations, and we regard Xianyang, Shaanxi, China as study area, taking Liangsidu Monitoring Station as example. *Materials and methods* In this research, the RNN (recurrent neural network) model based on the NARX (nonlinear autoregressive with external input) method has been proposed by using the hourly monitoring data of pollutants (incl. $PM_{2.5}$, PM_{10} , NO_2 , NO, NO_x and CO) and meteorology (incl. wind direction, wind speed, temperature, humidity, pressure, etc.). *Results* Six optimal structures of the neurons in hidden layer have been determined to predict the $PM_{2.5}$ and PM_{10} concentrations in the following 6 h, 12 h and 24 h, respectively. The conducted experiments shown that (1) for the $PM_{2.5}$ prediction of 6 h, the performance becomes the best when the neuron number of hidden layers has been

收稿日期: 2019-04-02; 录用日期: 2019-07-11; 网络出版: 2019-07-16

Received Date: 2019-04-02; Accepted Date: 2019-07-11; Online first: 2019-07-16

基金项目:国家自然科学基金项目(41871315)

Foundation Item: National Natural Science Foundation of China (41871315)

通信作者: 张 猛, E-mail: sql_zhang@hotmail.com

Corresponding Author: ZHANG Meng, E-mail: sql_zhang@hotmail.com

Citation: Zhang D N, Zhang M, Zhang B. 2020. The model to predict PM_{2.5/10} concentrations based on NARX neural network—taking Liangsidu monitoring station in Xianyang as an example [J]. Journal of Earth Environment, 11(2): 161–168.

引用格式: 张丹宁,张 猛,张 博. 2020. 基于 NARX 神经网络的 PM_{2.5/10} 浓度值预测模型—— 以咸阳市两寺渡监测站为例 [J]. *地球环境 学报*, 11(2): 161-168.

settled as 8, while for the $PM_{2.5}$ prediction of 12 h and 24 h, the neuron number of hidden layers should be turned to 12 and 7 for the best prediction accuracy; and (2) for the prediction PM_{10} prediction of 6 h, 12 h and 24 h, the optimal settlements of the neurons number in hidden layers are 12, 10 and 13 respectively. *Discussion* In general, the proposed model has revealed satisfactory performance for the predictions of $PM_{2.5/10}$ concentrations and the prediction accuracy for the next 6 h is slightly better than that for 12 h and 24 h. Some uncertain predictions, however, still exist especially when unusual meteorological situation occurs. In addition, the data used for the neural-network training are not quite enough. *Conclusions* It has been demonstrated that the established RNN model based on the NARX network can be implemented to effectively predict the concentration of $PM_{2.5/10}$: the *R* values for the $PM_{2.5}$ predictions of the following 6 h, 12 h and 24 h reaches 0.929281, 0.906767 and 0.889691, respectively; the corresponding RMSE values are 0.0008, 0.0010 and 0.0012; for the prediction of PM_{10} , the *R* reaches 0.929867, 0.921972 and 0.917757, and the corresponding RMSE are 0.0013, 0.0014 and 0.0017 for the following 6 h, 12 h and 24 h respectively. *Recommendations and perspectives* In order to further improve the prediction performance of the PM_{2.5/10} concentrations, the effect of unusual methodological conditions should be considered by the proposed RNN model; moreover, the sensitivity analysis of the different input parameters need to be further investigated.

Key words: PM_{2.5}; PM₁₀; air quality; NARX; recurrent neural network; air pollution prediction

近几年,由于雾霾天数的显著增多和呼吸系 统疾病持续高发,空气质量日益引起关注,更多 的研究指向空气质量的预测问题,希望通过建立 科学合理的空气质量预测模型,对污染物浓度进 行有效的预测预报,从而降低大气污染带来的危 害(徐敬等,2007;杨新兴等,2012;陈广银等, 2017)。

大气污染物的预测工作历经几十年发展,预 测方法和预测技术都得到了全面的提升。但是, 如何改善空气污染物的预测精度,一直是空气质 量预测领域必须考虑和研究的重点(郑国威等, 2018)。常见的时间序列预测模型有: AR 模型 (自回归模型)、MA模型(滑动平均模型) 和 ARMA 模型(自回归滑动平均模型或混合模 型)。针对空气中细颗粒物的预测模型主要包括: 空气质量模型 CMAQ、高斯轨迹烟云扩散模型、 GEOS-CHEM 模型、神经网络模型等。由于环境 空气中污染物浓度的预测是一个相对复杂的非线 性问题,因此在各种预测方法中,人工神经网络 (artificial neural network, ANN) 以其刻画非线 性现象强的特点,对于空气中污染物浓度的高精 度预测具有独有的优越性(薛文博等, 2013; 彭斯俊, 2014)。张怡文等(2015)通过与多 元回归模型的对比,建立了考虑多种污染物的 BP 神经网络模型,对 PM25 的等级进行预测,结果相 对较为准确,但有待提高。相比于 BP 神经网络,

递归神经网络通过将长度不定的输入分割为多个 等长度的小块, 再将其依次输入到网络中, 实现 了神经网络对变长输入的处理(Li et al, 2016), 它能充分考虑预测目标同时间序列的关系及各神 经元间的联系,因而在包含时间及时间序列的预 测方面有独特的优越性。Kim et al (2010)通过对 比 RNN(递归神经网络)、MLR(多元线性回归 模型)及NN(神经网络)三种方法的预测结果, 证明了 RNN 的预测结果较其他两种具有更小的 RMSE 值,结果更精确,同时,过多的参考变量 会对预测结果产生不利影响。由于 NARX (非线 性有源自回归)网络模型输出层不断将包含历史 的输出数据反馈到输入层,从而使其具有记忆能 力,相比于传统回归神经网络能包含更长时间的 网络历史状态和实时状态信息。袁红春等(2018) 建立了 PCA-NARX 网络模型,结果证明其具有 更好的泛化能力,预测精度得到了提高。Cai et al (2010)应用 NARX 网络对暴时 SYM-H 指数进 行预测,并取得了不错的效果。

当前,一些基于神经网络建立的 PM_{2.5} 和 PM₁₀ (记为 PM_{2.5/10})预测模型虽然达到了较高的预测 精度,但是大都没有全面考虑到各种污染物浓度 和气象因素共同作用的影响。事实上,细颗粒物 PM_{2.5/10}的来源既包括污染源直接排放,也包括氮 氧化物,碳氧化物,挥发性有机物等的转化,同 时,其稀释及迁移转化过程与气象因素和时间序 列密不可分,因而建立模型时应对多方面都加以 考虑。本研究以陕西省咸阳市为研究区,根据两寺 渡监测站长期监测的污染物数据(PM_{2.5}、PM₁₀、 NO₂、NO、NO_x、CO)及气象数据(包括风向、 风速、气温、湿度、气压),建立了基于非线性 有源自回归网络(NARX)的递归模型,并通过改 变神经网络隐藏层神经元节点个数确定不同时间 尺度范围下的最优网络模型结构,从而实现了对 未来6小时、12小时以及24小时的PM_{2.5/10}浓度 值的有效预测,旨在为大众出行提供指导,为环 保决策提供意见。

1 研究方法及技术路线

技术路线如图 1 所示。首先,进行相关性分析 确定建模数据。接着,填补缺失数据,并对所有数 据做 0—1 归—化处理,并进行纵向数据分割。再 将完成预处理的数据分为训练数据集 *A* 和验证数 据集 *B*,在 MATLAB 中的时间序列工具箱中选择 NARX 网络模型,分别对延时 6 h,12 h 以及 24 h 的 PM₂₅和 PM₁₀污染浓度值进行训练和学习;根 据属性简约结果,输入矩阵为 11 维,输出为 1 维。 其中,在学习部分,*A* 数据集中 70% 的数据用作 训练,15% 的数据用以验证,剩余的 15% 用做测 试。研究选定列文伯格马夸尔特法,并根据经验 公式 (1)确定神经元个数大致范围,再通过多次改 变隐藏层神经元节点数进行训练,每种结构进行 10 次实验,并采用多项指标对结果进行评价,最 终确定了不同情况下的最优网络结构。

 $h=\sqrt{m+n}+a$

(1)

式中: h 为隐含层神经元数, m 是输入神经元数, n 是输出神经元数, a 为 1—10 的任意常数,本研 究中, m 与 n 分别为 11 和 1。经计算,隐含层神 经元数大致为 5 至 14 较为合适。

2 实验数据及预处理

2.1 数据选取及相关性分析

PM_{2.5} 及 PM₁₀ 都是成分复杂的污染物,其浓 度受多种其他物质的影响,并和多种其他气象因 素相关。因而,本研究通过与 PM_{2.5/10} 显著相关的 污染物建立预测模型:在 SPSS 中输入一系列可能 影响 PM_{2.5} 及 PM₁₀ 变化的污染物数据及气象数据, 并进行相关性分析;根据皮尔森相关系数,最终 确定与之呈显著相关的污染物种类数据和气象数 据(PM₁₀、PM_{2.5}、NO、NO_x、NO₂、CO、风速、 风向、湿度、气温、气压)(表1)。



Fig.1 Technical route

咸阳市两寺渡监测站 2017 年 1 月 1 日—2017 年 5 月 6 日的以上 11 种数据,共 3013 组,被分 为 *A* 和 *B* 两段。其中,*A* 段数据(共 2399 组)用 来训练模型,*B* 段数据(共 614 组)用来验证模型。 此外,还需对外部输入变量中的异常值进行处理, 并对极个别缺失的数据,取其前两日及后两日共 四日的数据进行平均。

2.2 归一化及数据分割

由于污染物数据和气象数据的极值点均不可忽略,因而不进行数据平滑化处理,只做简单的数据 归一化。归一化是为了消除不同量纲及数量级对网 络训练的影响,以保证后期数据便于处理,同时使 得程序运行时的收敛加快;并在保留数据变化规律 的完整性的前提下,让数据均落在0—1。归一化的 具体方法为:首先,分别找出每一项指标的最大值

DOI: 10.7515/JEE192013

地球环境学报

第11卷

(2)

和最	小值并将二者做差;	ſſI,	f 后依据以下公式进行计	
算,	最终得到大于0小于	1	的数据集:	

表 1 SPSS 相关性分析											
Tab.1 SPSS correlation analysis											
	NO	NO	NO	DM		PM _{2.5}	风速	风向	气温	湿度	气压
	NO ₂	NO	NO _x	PM_{10}	0		Spe	Dir	Temp	Hum	Pre
NO_2	1										
NO	0.46**	1									
NO_x	0.74**	0.94**	1								
PM_{10}	0.46**	0.41**	0.49**	1							
CO	0.57**	0.55**	0.64**	0.72**	1						
PM _{2.5}	0.52**	0.44**	0.54**	0.85**	0.90**	1					
」速 Dir	-0.47**	-0.39**	-0.48**	0.19**	-0.31**	-0.29**	1				
向 Spe	0.07**	0.15**	0.14**	0.04*	0.01	0.02	-0.45**	1			
温 Temp	-0.09**	-0.26**	-0.24**	-0.21**	-0.39**	-0.39**	0.06**	0.03	1		
度 Hum	0.06**	0.21**	0.18**	0.05*	0.31**	0.28**	-0.13**	0.10**	-0.32**	1	
压 Pre	-0.01	0.14**	0.10**	0.07*	0.13**	0.13**	-0.11**	0.06**	-0.74**	-0.02	1
表示在 0.01 尾上显著相关,* 表示在 0.05 尾上显著相关 Dir Spe Temp Hum 和 Pre 分别为 direction speed											

** 表示在 0.01 层上显者相关; * 表示在 0.05 层上显者相关。Dir、Spe、Iemp、Hum 相 Pre 分别为 direction、speed、 temperature、pressure、 humidity 的简写。

** means a significant correlation on the 0.01 level; * indicates a significant correlation on the 0.05 level; Dir, Spe, Temp, Hum and Pre are short for direction, speed, temperature, humidity and pressure respectively.

数据分割是指把逻辑上是统一整体的数据分 割成较小的、可以独立管理的物理单元进行存储, 以便于重构、重组和恢复。为了使用方便,选取 垂直分割方法(vertical splitting),使具有不同特 性的数据更加均匀地分布于不同的样本集合中, 从而让后续数据输入与计算更具有代表性和普遍 性,同时还可使模型的整体稳定性得到提高。具 体做法为:将 $m \times n$ 的矩阵划分为1×m的行向量, 其中,每个元素均为 $n \times 1$ 的列向量。如公式(3)— (7)所示, A_1 、 B_1 、 A_2 、 B_2 分别为训练输入矩阵、 训练输出矩阵、验证输入矩阵、验证输出矩阵。 其中:m=2399,n=11,t=614。p、q、r、s、t、u、v、w、x、y、z分别代表 PM₂₅、PM₁₀、NO₂、 NO、NO_x、CO、风向、风速、气温、湿度、气压, b为 PM₁₀、PM₂₅的输出值。

$$A_1 = |a_1 \ a_2 \ a_3 \cdots a_{m-1} \ a_m| \tag{3}$$

$$a_{i} = \left| p_{i} \ q_{i} \ r_{i} \ s_{i} \ t_{i} \ u_{i} \ v_{i} \ w_{i} \ x_{i} \ y_{i} \ z_{i} \right|^{T}$$
(4)

$$B_1 = |b_1 \ b_2 \ b_i \cdots b_{m-1} \ b_m| \tag{5}$$

$$A_2 = |a_1 \ a_2 \ a_3 \cdots a_{t-1} \ a_t| \tag{6}$$

$$B_2 = |b_1 \ b_2 \ b_3 \cdots b_{t-1} \ b_t| \tag{7}$$

3 模型及算法

3.1 非线性有源自回归神经网络

根据网络结构中是否存在反馈回路,神经网 络可分为动态神经网络和静态神经网络(陈哲

DOI: 10.7515/JEE192013

等,2014;吴春霖等,2018)。常用于解决时间 序列问题的动态神经网络有三种:非线性有源自 回归网络 (nonlinear autoregressive with external input, NARX)、非线性自回归网络(nonlinear autoregressive, NAR)和非线性输入输出网络 (nonlinear input-output),均可用以解决非线性 时间序列问题,其结构分别如图 2a、b、c 所示。 其中,非线性有源自回归算法是一种结构清晰的 动态神经网络,是以一切相关因素,包括自身历 史值进行预测,因而常用以解决非线性的时间序 列问题(Roghanchi and Kocsis, 2019); 非线性 自回归网络是仅仅基于过去一段时间 PM25/10 本身 浓度值的预测模型; 而非线性输入输出网络是根 据过去一段时间的其他因素(除自身以外的其他 因素)来预测自身浓度,因而对相关性及对应关 系要求较高。

 $(n=1, 2, 3\cdots)$

NARX 的基本结构包括:输入层,隐藏层,输出层及输出到输入的延时。它能将神经网络的输出延时保存之后,通过外部反馈,引入到输入向量中,因而,具有记忆力,可以代替 BP 等神经 网络进行更长时间尺度的学习和预测,且效果良好(杨丽等,2018)。本研究所用的 NARX 网络的具体模型结构设计如图 3。其中,X₀ 在本研究中包括 10 个量,即除被预测量以外的其余污染物和气象因素; Y₀ 包括 1 个量,即被预测量,X₀ 和

区风气湿气

 $Y_{(0)}$ 共同组成了 A 矩阵(公式(3)—(7)), $Y_{(0)}$ 是 B 矩阵(公式(3)—(7)); d 为延时时间(本研究 d 值分别为 6 h、12 h 及 24 h); x 为隐藏层神经元 个数; w 是隐藏层到隐藏层之间的权重; b (bias) 代表偏差。与其他两种网络相比,由于增加了对 各因素的考虑以及对前期输出值的反复学习,效 率更高。 PM_{25} 及 PM_{10} 均为成分复杂的混合物,其 形成、迁移转化过程和浓度变化情况是诸多因素 共同作用的结果。由于本研究基于多种污染物浓 度数据及气象因素监测数据,因而将模型确定为 NARX 网络最为合适。



图 2 三种不同网络模型结构





图 3 NARX 结构设计 Fig.3 Structure design of NARX

3.2 列文伯格 - 马夸尔特法

列文伯格-马夸尔特法(Levenberg-Marquardt, LM),是使用最广泛的非线性最小二乘算法,属 于基于标准数值最优化技术的快速算法,用于解 决非线性问题。相比于贝叶斯正规化法(Bayesian regularization)和量化共轭梯度法(scaled conjugate gradient),列文伯格-马夸尔特法算法能将梯度 下降法与高斯-牛顿法相结合,因而其既有高斯-牛顿法的局部收敛性,又具有梯度下降法的全局 特性,寻优速度十分快。因此,本研究选取列文 伯格-马夸尔特法作为模型的训练算法,用于提高 迭代收敛的速度,从而高效确定不同情况下的最 优结构模型。

3.3 模型性能评价及建模过程

为了评价所建立的 NARX 神经网络模型的预 测性能,对 PM_{2.5}及 PM₁₀ 在未来 6 h、12 h、24 h 的预测值和实际观测值进行比对分析。具体的评 价指标包括均方根误差 (root mean square error, RMSE) (公式 (8))、相关系数 (correlation coefficient, R) (公式 (9)):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (V_a - V_b)^2}$$
(8)
$$R = \frac{\frac{1}{n} \sum_{i=1}^{n} (V_a - V_a') (V_b - V_b')}{\sqrt{\sum_{i=1}^{n} (V_a - V_a')^2} \sqrt{\sum_{i=1}^{n} (V_b - V_b')^2}}$$
(9)

公式(8)和公式(9)中: n为B 段测试样本部分的 样本个数,即 n=614; V_a表示目标输出值,即预 测值; V'_a表示目标输出平均值; V_b表示实际输出 值; V'_b表示预测平均值。

4 结果与分析

经过多次实验对比可知:针对不同污染物,不 同时段,模型最优结构有所不同。在训练过程中, 实验效果均十分理想。测试过程中,当隐藏层神 经元个数分别为12、10和13时,对于PM₁₀浓度 未来6h、12h及24h的预测能力达到最优;对 于PM₂₅:在未来6h的预测中,当隐藏层神经元 个数为8时准确度达到最佳;预测未来12h浓度 时,隐含层神经元最适个数为12;对于未来24h 的浓度值预测,神经元个数为7可达到最好效 果。对于两种污染物,随着未来小时数的增加, 预测能力逐渐下降。其性能测试结果如表2、表3 所示。

图 4 — 6 分 别 为 6 h、12 h、24 h 的 $PM_{2.5}$ 和 PM_{10} 浓度预测情况。对于 6 h PM_{10} 的预测情况, 预测值与实际值的趋势变化情况基本相似, 吻合 程度较高,存在个别时刻,预测值略大于实际值。 其中,在 2017 年 4 月 12 日 22 时对次日 4 时的预测值偏高较为明显,经分析,此段时间内的气象特 征及此污染物变化情况为:空气湿度较高且基本 保持稳定,风速在期间某一时刻突然变大, PM_{10} 自身也在期间出现极值,而后下降;4月 17 日 3 时—4月 18 日 3 时、4 月 18 日 15 时的预测值整 体偏高于实际值,原因可能是这段时间内,风速有 较为明显的变化且气温较低;2017年4月29日6时,其向后6h内,由于温度较低且持续下降,因而预测值也偏高;2017年5月4日11时—15时

的风速很大,温度较小且温度持续下降,因而11 时的6h预测情况偏高。综上,当预测时间段内风 速大,温度较低的时候,预测结果会偏高。

表 2 PM ₂₅ /PM ₁₀ 模型训练性能测试							
Tab.2 Training performances of the $PM_{2.5}/PM_{10}$ model							
	预测时间	神经元数		目标值个数	均方误差	相关系数	
	/h	Neurons		Target	MSE	R	
			Training	1679	0.000246010	0.983558	
	6	12	Validation	360	0.000256164	0.983986	
			Testing	360	0.000280521	0.982099	
			Training	1679	0.000168048	0.988706	
PM ₁₀	12	10	Validation	360	0.000414316	0.972748	
			Testing	360	0.000468368	0.971421	
		13	Training	1679	0.000107805	0.992960	
	24		Validation	360	0.000451070	0.971335	
			Testing	360	0.000556232	0.962324	
			Training	1679	0.000216135	0.986932	
	6	8	Validation	360	0.000608079	0.959266	
			Testing	360	0.000285577	0.966998	
	12	12	Training	1679	0.000784900	0.989577	
PM _{2.5}			Validation	360	0.000151811	0.976476	
			Testing	360	0.000120015	0.984521	
	24	7	Training	1679	0.000857873	0.988816	
			Validation	360	0.000129170	0.980423	
			Testing	360	0.000183106	0.974858	

表 3 PM _{2.5} /PM ₁₀ 预测性能测试								
Tab.3 Prediction performances of the $PM_{2.5}/PM_{10}$ model								
	预测时间	神经元个数	均方根误差	相关系数				
	/h	Neurons	RMSE	R				
	6	12	0.0013	0.929867				
PM_{10}	12	10	0.0014	0.921972				
	24	13	0.0017	0.917757				
	6	8	0.0008	0.929281				
PM _{2.5}	12	12	0.0010	0.906767				
	24	7	0.0012	0.889691				



图 4 未来 6 小时 $PM_{10}(a)$ 和 $PM_{25}(b)$ 的预测值与实测值比对图 Fig.4 Comparison between the prediction and monitoring values of $PM_{10}(a)$ and $PM_{25}(b)$ in 6 hours



图 5 未来 12 小时 $PM_{10}(a)$ 和 $PM_{25}(b)$ 的预测值与实测值比对图 Fig.5 Comparison between the prediction and monitoring values of $PM_{10}(a)$ and $PM_{25}(b)$ in 12 hours



图 6 未来 24 小时 $PM_{10}(a)$ 和 $PM_{25}(b)$ 的预测值与实测值比对图 Fig.6 Comparison between the prediction and monitoring values of $PM_{10}(a)$ and $PM_{25}(b)$ in 24 hours

2017年4月12日13时预测6h后PM₁₀浓度时, 测值偏低。期间,风速先下降再升高,PM₁₀出现 过极小值,且温度在逐渐下降;在4月17日15时 进行预测时,由于在往后6h内,风速先增大再减 小,湿度不变,温度相对较低,因而预测值偏低; 4月28日11时—16时,PM₁₀本身和NO出现异 常极值,风速先变大后变小,预测结果的值较低; 5月5日14时的6h预测值明显低于实际值,观察 其前后的5h风速值可知,变化明显,同时,NO浓 度出现较大波动,温差较大。综上,当风速渐变或 突变,NO浓度出现波动,或预测点后6h内出现 PM₁₀极点的时候,不利于预测,会使预测值偏低。

 $PM_{2.5}$ 的 6 h 预测值与实际值变化趋势一致, 但是整体值偏高。存在个别极值点,低于实际值。 但是在整体上可以较为准确地预测 $PM_{2.5}$ 浓度情况及极值点的存在。而对于 12 h 及 24 h 的 $PM_{2.5}$ 及 PM_{10} 预测来说, 12 h 预测浓度的准确性均低于 6 h, 但高于 24 h, 对 PM_{10} 的预测准确率普遍高于 $PM_{2.50}$

5 结论

本研究所建立的 PM₁₀/PM_{2.5} 模型运行成本 低且具有较强的实用性和时效性,可较为准确 地预测未来 6 h、12 h 及 24 h 的 PM₁₀ 及 PM_{2.5} 浓 度。PM₁₀ 预测模型的 *R* 值分别达到了 0.929867、 0.921972、0.917757, RMSE 值 分 别 为 0.0013、 0.0014、0.0017; PM_{2.5} 预测模型的 *R* 值分别为 0.929281、0.906767、0.889691, RMSE 值分别为 0.0008、0.0010、0.0012。由此可见,该预测模型 可靠性强,可在很大程度上为人们提供预警服务, 为环保决策提供指导依据。

由预测值与实际值的对比情况可知:对未来 24 h 以内的 PM₁₀ 的预测值与实际观测值趋势基本 一致,尤其是 6 h 预测值,十分准确。因而,此 模型对未来 24 h 内的 PM₁₀ 浓度可以实现有效预 测。此外,在个别峰值的预测方面,只能预测到 峰值存在,但峰值预测值整体略小于实测值。尤 其是突发性的极大值,偏差略大,后期可针对峰 值做进一步的分析处理。而对于未来 24 h 的 PM₂₅

168

地球环境学报

预测,预测值与实际值趋势相似,可以实现预警的作用,但预测值整体略微高于实际值,同时与 PM₁₀的预测曲线相似,对PM₂₅的预测,也存在 着峰值预测缺陷的问题。同时,不论是PM₁₀,还 是PM₂₅,此模型的预测精度均随预测小时数的增 加而有所降低,后期可通过扩大数据样本量、加 入SO₂等其他污染物的监测值、在预处理部分进 行 PCA 降维处理以及建立修正模型等方法加以弥 补。此外,针对 NO 等污染物监测值在个别时段内 波动极大的情况,可尝试通过统计学或机器学习 的方法找寻其中的规律,并依照规律对数据进行 相应的处理,以得到更为准确的预测效果。

参考文献

- 陈广银,蔡灏兢,姜 欣. 2017. 马尔可夫修正的 BP 神经网络在 PM₂₅ 预测中的应用 [J]. *能源环境保护*, 31(5): 8–11.
 [Chen G Y, Cai H J, Jiang X. 2017. Application of corrected BP neural network in PM₂₅ prediction based on Markov [J]. *Energy Environmental Protection*, 31(5): 8–11.]
- 陈 哲,刘建坤,李冰洁. 2014. 基于 BP 神经网络的空气 质量评价及预测 [J]. 现代经济信息,(24): 387-388. [Chen Z, Liu J K, Li B J. 2014. Air quality assessment and prediction based on BP neural network [J]. *Modern Economic Information*, (24): 387-388.]
- 彭斯俊, 沈加超, 朱 雪. 2014. 基于 ARIMA 模型的 PM_{2.5} 预测 [J]. *安全与环境工程*, 21(6): 125–128. [Peng S J, Shen J C, Zhu X. 2014. Forecast of PM_{2.5} based on the ARIMA model [J]. *Safety and Environmental Engineering*, 21(6): 125–128.]
- 吴春霖,李 琦,侯俊雄,等. 2018. 卷积神经网络的 PM_{2.5} 预报模型 [J]. *测绘科学*, 43(8): 68-75. [Wu C L, Li Q, Hou J X, et al. 2018. PM_{2.5} concentration prediction using convolutional neural networks [J]. *Science of Surveying and Mapping*, 43(8): 68-75.]
- 徐 敬,丁国安,颜 鹏,等. 2007. 北京地区 PM₂₅的成分 特征及来源分析 [J]. 应用气象学报, 18(5): 645-654. [Xu J, Ding G A, Yan P, et al. 2007. Componential characteristics and sources identification of PM₂₅ in Beijing [J]. Journal of Applied Meteorological Science, 18(5): 645-654.]
- 薛文博,王金南,杨金田,等.2013. 国内外空气质量模型 研究进展[J]. 环境与可持续发展,38(3):14-20. [Xue W B, Wang J N, Yang J T, et al. 2013. Domestic and foreign research progress of air quality model [J]. *Environment and*

Sustainable Development, 38(3): 14–20.]

- 杨 丽,吴雨茜,王俊丽,等. 2018. 循环神经网络研究综述[J]. *计算机应用*, 38(S2): 1-6, 26. [Yang L, Wu Y X, Wang J L, et al. 2018. Research on recurrent neural network [J]. *Journal* of Computer Applications, 38(S2): 1-6, 26.]
- 杨新兴, 冯丽华, 尉 鹏. 2012. 大气颗粒物 PM_{2.5} 及其危 害 [J]. *前沿科学*, 6: 22-31. [Yang X X, Feng L H, Wei P. 2012. Air particulate matter PM_{2.5} in Beijing and its harm [J]. *Frontier Science*, 6: 22-31.]
- 袁红春,赵彦涛,刘金生.2018. 基于 PCA-NARX 神经网 络的氨氮预测 [J]. *大连海洋大学学报*, 33(6): 808-813. [Yuan H C, Zhao Y T, Liu J S. 2018. Ammonia nitrogen level forecasting based on PCA-NARX neural network [J]. *Journal of Dalian Ocean University*, 33(6): 808-813.]
- 张怡文,胡静宜,王 冉.2015. 基于神经网络的 PM_{2.5} 预 测模型 [J]. 江苏师范大学学报(自然科学版), 33(1): 63-65. [Zhang Y W, Hu J Y, Wang R. 2015. PM_{2.5} forecasting model based on neural network [J]. Journal of Jiangsu Normal University(Natural Science Edition), 33(1): 63-65.]
- 郑国威, 王腾军, 杨友森, 等. 2018. 基于遗传小波神经网络的 PM_{2.5} 浓度预测模型 [J]. *测绘与空间地理信息*, 41(9): 248-250, 256. [Zheng G W, Wang T J, Yang Y S, et al. 2018. PM_{2.5} concentration prediction model based on genetic wavelet neural network [J]. *Geomatics & Spatial Information Technology*, 41(9): 248-250, 256.]
- Cai L, Ma S Y, Zhou Y L. 2010. Prediction of SYM-H index during large storms by NARX neural network from IMF and solar wind data [J]. *Annales Geophysicae*, 28(2): 381–393.
- Kim M H, Kim Y S, Lim J J, et al. 2010. Data-driven prediction model of indoor air quality in an underground space [J]. *Korean Journal of Chemical Engineering*, 27(6): 1675–1680.
- Li N. 2016. Improved Delay-dependent Stability Criterion for Neutral-type Recurrent Neural Networks with Timevarying Delay [C]// Proceedings of the 35th China Control Conference. China Automation Society Control Theory Professional Committee, China Society of Systems Engineering. Ningbo.
- Roghanchi P, Kocsis K C. 2019. Quantifying the thermal damping effect in underground vertical shafts using the nonlinear autoregressive with external input (NARX) algorithm [J]. *International Journal of Mining Science and Technology*, 29(2): 255–262.