

基于WRF-CMQ-BP神经网络下空气质量的二次建模研究

何家伟¹, 张孙杰², 尚朝辉¹, 李逢源¹

¹上海理工大学, 理学院, 上海

²上海理工大学, 光电信息与计算机工程学院, 上海

收稿日期: 2022年4月22日; 录用日期: 2022年5月20日; 发布日期: 2022年5月27日

摘要

随着工业化城市经济的飞速发展, 大气污染严重危害着生态环境和人体健康。污染防治实践表明, 建立空气质量预报模型, 提前获知可能发生的大气污染过程并采取相应控制措施, 是减少大气污染对人体健康和环境等造成的危害并提高环境空气质量的有效方法之一。本文针对污染物浓度对其进行预报, 基于空气质量预报模式系统(WRF-CMAQ)等一次预报模型模拟结果, 结合更多的实测数据源进行再建模, 以提高预报的准确性, 优化预报模型。在建立WRF-CMAQ-BP模型中, 采用的数据是六种常规污染物的单日浓度, 用模拟预测的值来逼近真实的值, 从而达到较为精准的预报值。同时采用皮尔逊相关系数方法来找到影响空气质量的主要因素。

关键词

二次建模, WRF-CMAQ模型, BP神经网络, 皮尔逊相关系数

A Secondary Modeling Study of Air Quality under WRF-CMQ-BP Neural Network Based

Jiawei He¹, Sunjie Zhang², Chaohui Shang¹, Fengyuan Li¹

¹College of Science, University of Shanghai for Science and Technology, Shanghai

²School of Optical Information and Computer Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Apr. 22nd, 2022; accepted: May 20th, 2022; published: May 27th, 2022

Abstract

With the rapid development of industrialized urban economy, air pollution is seriously endangering ecological environment and human health. The practice of pollution prevention and control shows that establishing air quality forecasting models to know the possible air pollution process in advance and take corresponding control measures is one of the effective methods to reduce the harm caused by air pollution to human health and environment and improve the ambient air quality. In this paper, we forecast pollutant concentrations based on the simulation results of primary forecasting models such as the Weather Research and Forecasting-Community Multiscale Air Quality (WRF-CMAQ), and re-model them with more actual measurement data sources to improve the accuracy of forecasting and optimize the forecasting models. In the WRF-CMAQ-BP model, the data used are the single-day concentrations of six conventional pollutants, and the predicted values are used to approximate the real values, so as to achieve more accurate forecast values. The Pearson correlation coefficient method is also used to find the main factors affecting air quality.

Keywords

Secondary Modeling, WRF-CMAQ Model, BP Neural Network, Pearson Correlation Coefficient

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在城市经济飞速发展的同时, 环境治理问题日益突出, 空气质量问题逐渐成为人们共同关注的重要话题[1]。污染物浓度一年又一年的加剧促使呼吸道疾病的发病率明显呈上升趋势。近年来, 由于污染治理不到位, 多地遭遇较强的雾霾, 雾霾降低了能见度, 使得城市交通系统几度陷入瘫痪, 同时给人体健康带来了不能忽略的负面影响。可见建立合理的空气质量预报模型, 提前获知可能发生的大气污染过程并采取相应控制措施势在必行[2]。

BP (back propagation)神经网络是一种按照误差逆向传播算法训练的多层前馈神经网络, 是目前应用最广泛的神经网络。BP神经网络能学习和贮存大量的输入、输出模式映射关系, 而无需事前解释描述这种映射关系的数学方程。它的学习规划则是使用最速下降, 通过反向传播来不断调整网络的权值和阈值, 使网络的误差平方和最小。BP神经网络模型拓扑结构包括输入层(input layer)、隐藏层(hidden layer)和输出层(output layer)。

BP神经网络具有以下优点: 1) 具有较强的非线性映射能力, 这使得其特别适合于求解内部机制复杂的问题; 2) 具有高度自学习的自适应能力, 具有将学习成果应用于新知识的能力; 3) 具有一定的容错能力; 4) 简单、易操作、计算量小、并行性。

目前常用的预报空气质量的模型为 WRF-CMAQ 模拟体系。WRF 是集数值天气预报、大气模拟、数据同化于一体的模型系统, 主要用于大气环境模拟、天气研究、气象预报等[3] [4] [5]; CMAQ 模型主要用于环境规划、环境保护标准、环境影响评价等有关政策的制定与编制。虽然 WRF-CMAQ 模拟体系起到了不错的效果, 但是由于精确性不是很高, 故需要进行二次建模[6]。

2. 符号说明

表 1, 表 2, 表 3, 表 4 都是进行相关的解释, 为之后的建立模型做铺垫。

Table 1. System resulting data of standard experiment

表 1. 标准试验系统结果数据

符号	意义
$IAQI_p$	污染物 P 的空气质量分指数, 结果进位取整数
C_p	污染物 P 的质量浓度值
BP_{H_i}, BP_{L_0}	与 C_p 相近的污染物浓度限值的高位值与低位值
$IAQI_{H_i}, IAQI_{L_0}$	与 BP_{H_i}, BP_{L_0} 对应的空气质量分指数。
D	距离

Table 2. Meteorological data

表 2. 气象数据

	气象条件名称	单位
一次预报数据	近地 2 米温度	°C
	地表温度	K
	比湿	kg/kg
	湿度	%
	近地 10 米风速	m/s
	近地 10 米风向	°
	雨量	mm
	云量	mm
	边界层高度	m
	大气压	kpa
	感热通量	W/m ²
	潜热通量	W/m ²
	长波辐射	W/m ²
	短波辐射	W/m ²
	地面太阳能辐射	W/m ²

Table 3. Actual measurement data

表 3. 实际测量数据

实际测量数据	温度	°C
	湿度	%
	气压	Mbar
	风速	m/s
	风向	°

Table 4. Pollutant data**表 4.** 污染物数据

污染物名称	单位
SO ₂	μg/m ³
NO ₂	μg/m ³
PM10	μg/m ³
PM2.5	μg/m ³
O ₃	μg/m ³
CO	mg/m ³

Table 5. Air quality sub-index (IAQI) and the corresponding concentration limits of pollutant item**表 5.** 空气质量分指数(IAQI)及对应的污染物项目浓度限值

序号	指数或污染物项目	空气质量分指数及对应污染物浓度限值								单位
0	空气质量分指数(IAQI)	0	50	100	150	200	300	400	500	-
1	一氧化碳(CO) 24 小时平均	0	2	4	14	24	36	48	60	mg/m ³
2	二氧化硫(SO ₂) 24 小时平均	0	50	150	475	800	1600	2100	2620	
3	二氧化氮(NO ₂) 24 小时平均	0	40	80	180	280	565	750	940	
4	臭氧(O ₃)最大 8 小时滑动平均	0	100	160	215	265	800	-	-	
5	粒径小于等于 10 μm 颗粒物(PM10) 24 小时平均	0	50	150	250	350	420	500	600	μg/m ³
6	粒径小于等于 2.5 μm 颗粒物(PM2.5) 24 小时平均	0	35	75	115	150	250	350	500	

注: 1) 臭氧(O₃)最大 8 小时滑动平均浓度值高于 800 的, 不再进行其空气质量分指数计算。2) 其余污染物浓度高于 IAQI = 500 对应限值时, 不再进行其空气质量分指数计算。

表 5 是空气质量分指数及对应的污染物项目浓度限值, 可以看出二氧化硫空气质量分数偏高。

3. 数据分析与处理

使用 SPSS 统计软件, 分别得出各污染物浓度与各气象条件的相关性:

Table 6. Correlation of each pollutant concentration with each measured meteorological condition**表 6.** 各污染物浓度与各实测气象条件的相关性

	相关性				
	温度(°C)	湿度(%)	气压(Mbar)	风速(m/s)	风向(°)
SO ₂ 监测浓度(μg/m ³)	-0.141**	-0.446**	0.293**	-0.098**	-0.025**
NO ₂ 监测浓度(μg/m ³)	-0.359**	-0.044**	0.309**	-0.465**	-0.080**
PM10 监测浓度(μg/m ³)	-0.214**	-0.416**	0.370**	-0.237**	-0.021**
PM2.5 监测浓度(μg/m ³)	-0.287**	-0.303**	0.393**	-0.319**	-0.016*
O ₃ 监测浓度(μg/m ³)	0.329**	-0.557**	-0.053**	0.208**	-0.047**
CO 监测浓度(mg/m ³)	-0.372**	-0.061**	0.331**	-0.328**	0.032**

**在 0.01 级别(双尾), 相关性显著。*在 0.05 级别(双尾), 相关性显著。

由表 6 知, 实测气象条件中 SO₂ 的浓度与湿度呈较强的负相关性, NO₂ 的浓度与风速呈较强的负相关性, PM10 的浓度与湿度呈较强的负相关性, PM2.5 的浓度与气压呈较强的正相关性, O₃ 的浓度与湿

度呈较强的负相关性，CO 的浓度与温度呈较强的正相关性。由此得到湿度有利于 SO₂、PM10、O₃ 的扩散或沉降，风速有利于 NO₂ 的扩散或沉降，气压有利于 PM2.5 的扩散或沉降，而温度有利于 CO 的扩散或沉降。反之，风速和风向相对湿度不利于 SO₂ 的扩散或沉降，同样湿度和风向相对风速不利于 NO₂ 的扩散或沉降，温度和风向相对湿度不利于 PM10 的扩散或沉降，风向相对于气压不利于 PM2.5 的扩散或沉降，风向和气压相对湿度不利于 O₃ 的扩散或沉降，同样风向和气压相对温度不利于 CO 的扩散或沉降。(这里指的是采用皮尔逊(Pearson)相关系数得到的变量间的相关性分析，相关系数的绝对值越大，相关性越强：相关系数越接近于 1 或-1，正或负相关度越强，相关系数越接近于 0，相关度越弱。)

使用 Python 画出各污染物在所给时间范围内的月平均质量浓度与各气象条件的趋势关系图如下：

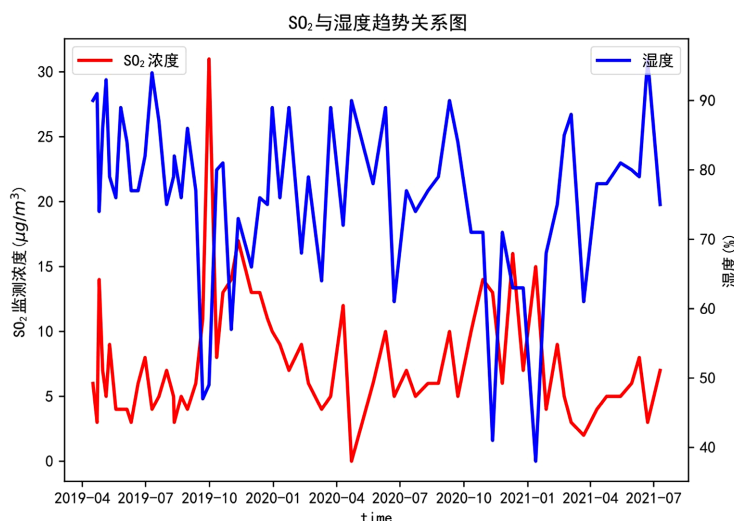


Figure 1. Trend of SO₂ vs. humidity
图 1. SO₂ 与湿度的趋势关系图

图 1 为 SO₂ 与湿度的趋势关系图，观察得知 SO₂ 和湿度整体呈明显的负相关趋势，即随着湿度的增大，SO₂ 的浓度减小；反之随着湿度的减小，SO₂ 的浓度增大。

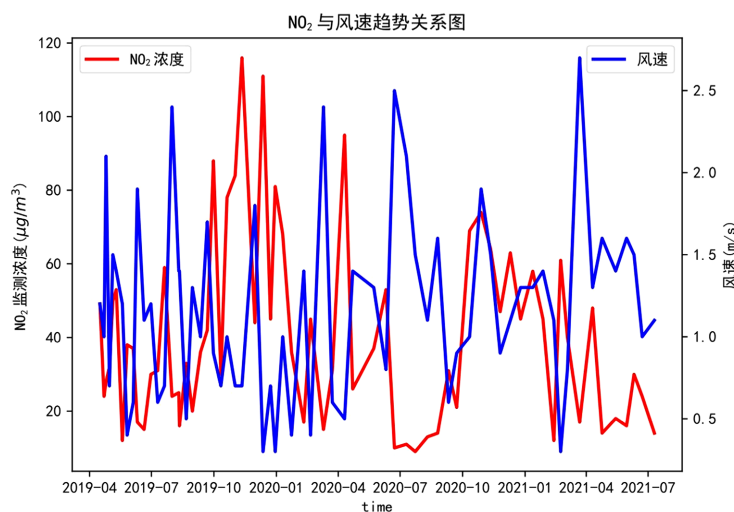


Figure 2. Trend of NO₂ versus wind speed
图 2. NO₂ 与风速的趋势关系图

图 2 为 NO_2 与风速的趋势关系图，观察得知 NO_2 和风速整体呈明显的负相关趋势，即随着风速的增强， NO_2 的浓度缩小；反之随着风速的减弱， NO_2 的浓度增大。

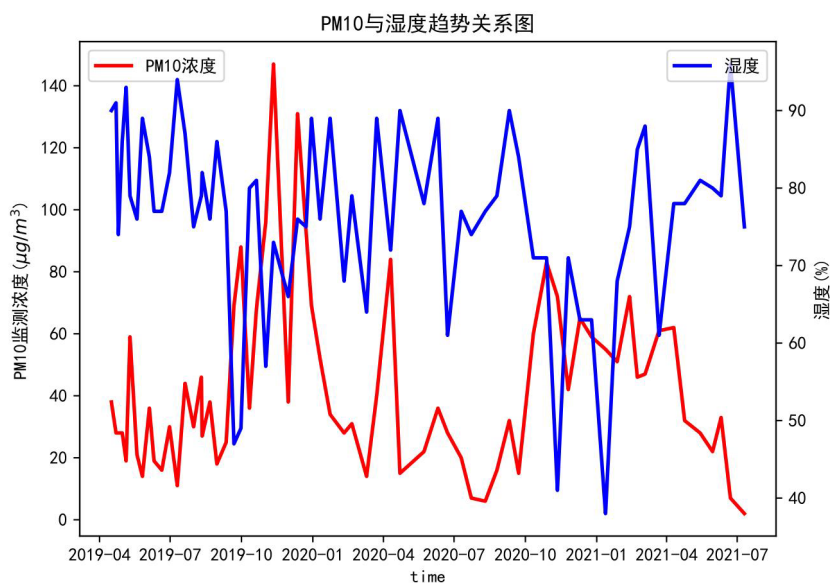


Figure 3. Trend graph of PM10 vs. humidity

图 3. PM10 与湿度的趋势关系图

图 3 为 PM10 与湿度的关系图，观察知 PM10 和湿度整体呈明显的负相关趋势，即随着湿度的增大，PM10 的浓度减小；反之随着湿度的减小，PM10 的浓度增大。

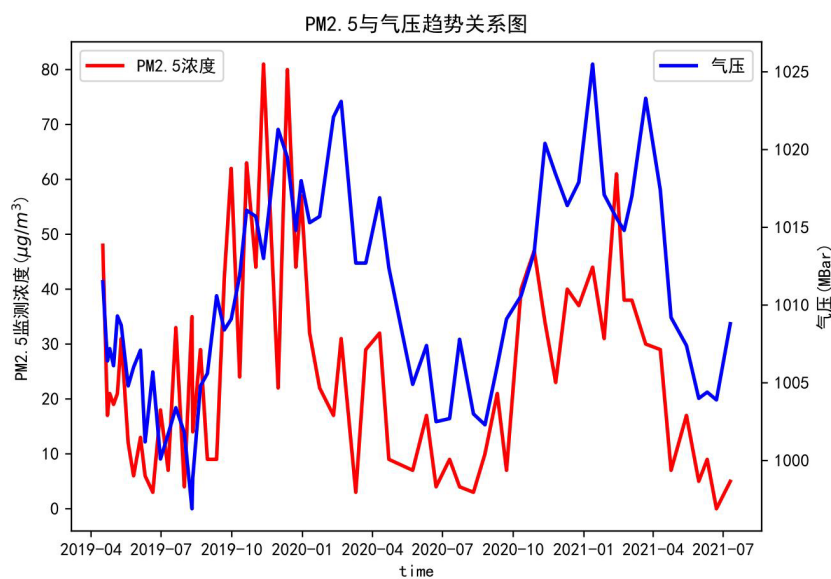


Figure 4. Trend of PM2.5 vs. barometric pressure

图 4. PM2.5 与气压的趋势关系图

图 4 为 PM2.5 与气压的趋势关系图，观察知 PM2.5 和气压整体呈明显的正相关趋势，即随着气压的增大，PM2.5 的浓度增大；反之随着气压的减小，PM2.5 的浓度减小。

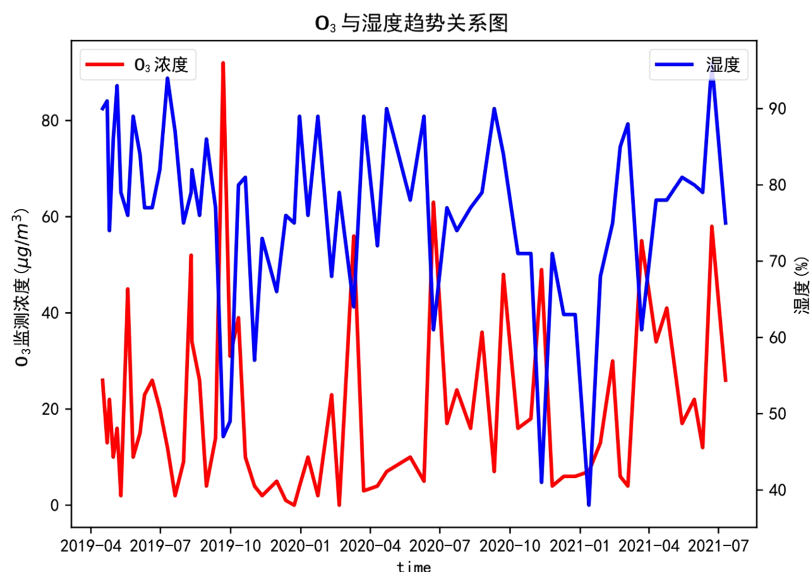


Figure 5. Trend diagram of ozone versus humidity

图 5. O_3 与湿度的趋势关系图

图 5 为 O_3 与湿度的趋势关系图，观察知 O_3 和湿度整体呈明显的负相关趋势，即随湿度的增大， O_3 的浓度减小；反之随着湿度的减小， O_3 的浓度增大。

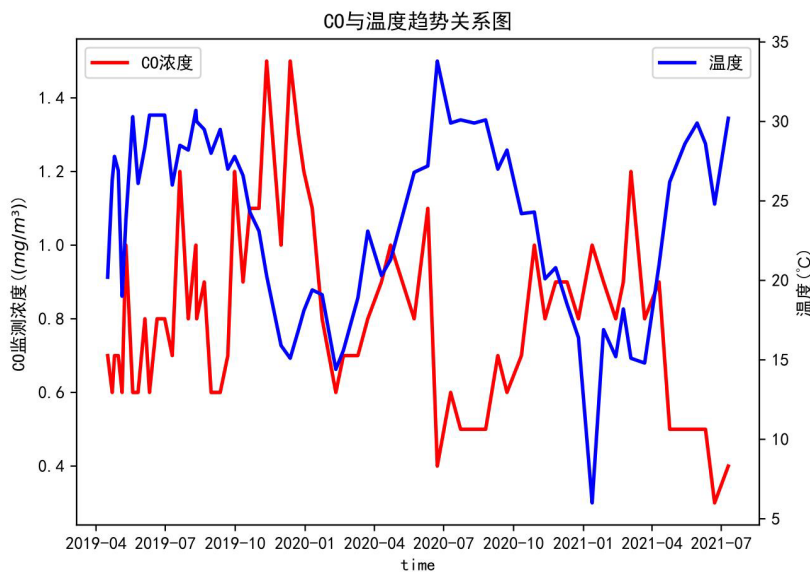


Figure 6. CO versus temperature trend graph

图 6. CO 与温度的趋势关系图

图 6 为 CO 与温度的趋势关系图，观察知 CO 和温度整体呈明显的负相关趋势，即随着温度的升高，CO 的浓度减小；反之随着温度的降低，CO 的浓度增大。

综上所述，对于不同污染物浓度的影响程度，气象条件的分类不同。总的来说比湿、大气压、长波辐射、近地 2 米温度、湿度有利于 O_3 外各污染物的扩散或沉降；而短波辐射、地面太阳能辐射、潜热通量、感热通量有利于 O_3 的扩散或沉降；相比而言湿度更有利于 SO_2 、 PM_{10} 、 O_3 的扩散或沉降，风速更

有利于 NO_2 的扩散或沉降，气压更有利于 $\text{PM}_{2.5}$ 的扩散或沉降，而温度更有利于 CO 的扩散或沉降。

4. 二次模型建立

首先，构建数据集。该问题中我们所建立的模型是 BP 神经网络二次预测模型。该模型的输入数据采用一次预报的气象条件，输出数据则是监测点逐小时污染物浓度与气象一次预报数据污染物浓度与监测点逐小时污染物浓度与气象实测数据污染物浓度的差值，所建模型的功能是构建出输入数据和输出数据的关系，从而找出它们之间存在的规律。BP 神经网络模型见图 7：

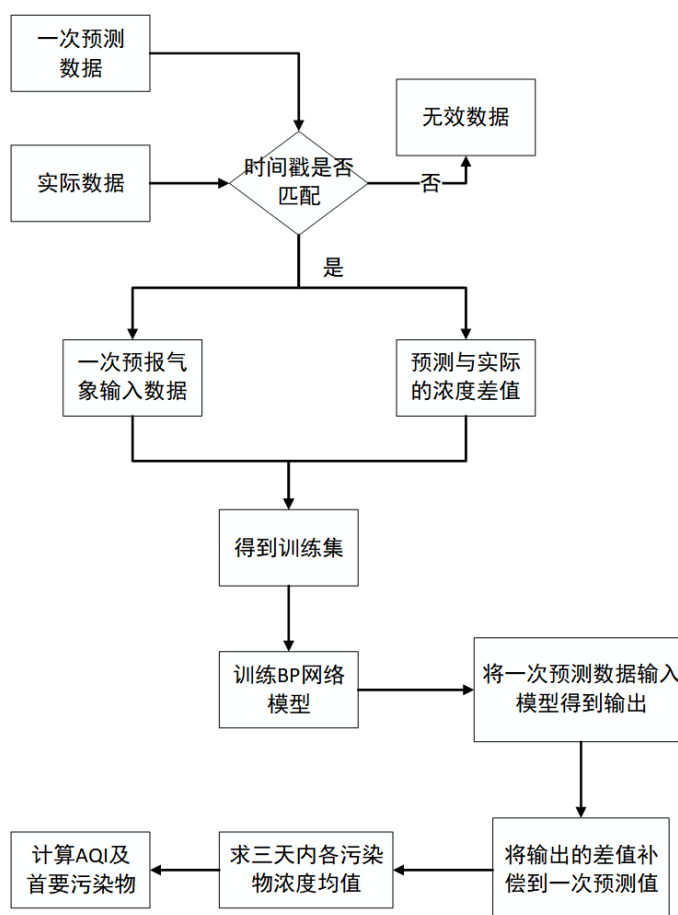


Figure 7. BP neural network model construction process based on forecast base data

图 7. 基于预报基础数据的 BP 神经网络模型构建过程

训练集数据构建步骤如下：

1) 数据时间戳对齐。对于监测点 A、B、C 的一次预报数据，选择预测时间在上文所述范围的所有数据进行读取，同时读取监测点 A、B、C 监测时间范围与预测时间范围相同的污染物测量数据。由于在监测点的一次预报数据中，每个运行日期对应三天的逐小时预报数据，且预报数据和实际监测数据各有缺失，因此，我们首先将预报数据中的预测时间和实测数据中的监测时间进行对齐。

2) 生成数据集。在完成时间戳对齐后，我们将选定时间范围内的监测点 A、B、C 的一次预报气象数据构建为输入数据，每个时间戳的数据看作是一个样本，将每个气象数据对应的污染物一次预报数据和实际监测数据做差，以此作为模型训练集的输出数据。

3) 导入模型。使用 python 的 numpy 和 pandas 库等完成对数据集的构建后, 将输入数据和输出数据分别生成两个.mat 文件, 导入我们的 BP 神经网络二次预测模型。

接着, 构建二次预报模型。综合选题以及网络上的信息我们得知, WRF-CMAQ 一次预报模型的结果并不理想, 与实测数据的偏差较大, 由于实际气象条件对空气质量影响很大(例如湿度降低有利于臭氧的生成), 且污染物浓度实测数据的变化情况对空气质量预报具有一定参考价值, 因此需要参考空气质量监测点获得的气象与污染物数据进行二次建模, 以优化预报模型。

我们将监测点 A、B、C 一次预报气象数据作为模型输入, 将对应的一次预报污染物数据和实测污染物数据的差值作为输出数据, 同时利用一次预报的气象数据和实测数据信息, 以构建他们之间的关系模型来进行预测。考虑到较大的输入数据量, 为了简模型复杂度, 我们使用 MATLAB 构建 BP 神经网络模型对数据进行建模以获得较好的二次预报数据。

我们将上文中所述构建的数据集导入神经网络模型, 模型的相关参数及运行结果如表 7:

Table 7. Parameter table

表 7. 参数表

隐藏层层数	隐藏层节点数	训练目标值	学习率	动量
1	39	0.001	0.0001	0.9

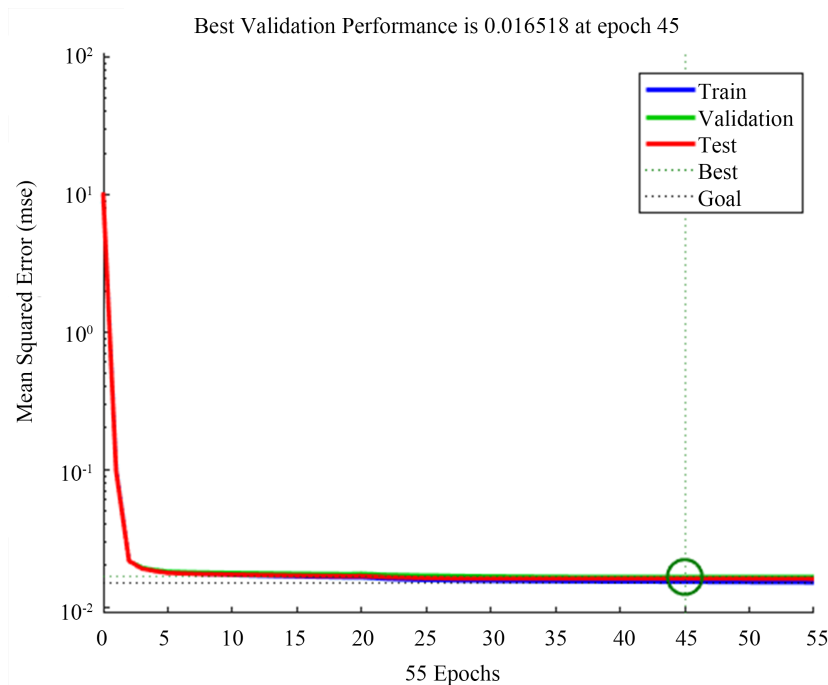


Figure 8. Network mse

图 8. 网络性能函数 mse

图 8 是网络性能函数, 在模型训练完成后, 我们将监测点 A、B、C 运行时间在 2021 年 7 月 13 日所获得的 2021 年 7 月 13 日至 2021 年 7 月 15 日的一次预报的气象数据作为输入, 获得的模型输出即是要补偿到对应一次预报污染物的数据, 将逐小时的一次预报污染物数据相加, 得到优化后的逐小时二次预报数据, 然后分别将三天的逐小时二次预报数据取平均, 得到监测点 A、B、C 在 2021 年 7 月 13 日至 7 月 15 日 6 种常规污染物的单日浓度值, 然后计算相应的 AQI 和首要污染物。

将所需值代入公式求得监测点 A、B、C 相应的 AQI 和首要污染物结果如表 8、表 9、表 10:

Table 8. AQI and primary pollutants measured at monitoring site A since July 13~July 15, 2021

表 8. 监测点 A 自 2021 年 7 月 13 日~7 月 15 日实测的 AQI 和首要污染物

预报日期	地点	二次模型日值预测						AQI	首要污染物
		SO ₂ (μg/m ³)	NO ₂ (μg/m ³)	PM10 (μg/m ³)	PM2.5 (μg/m ³)	O ₃ 最大 八小时滑动 平均 (μg/m ³)	CO (mg/m ³)		
2021/7/13	监测点 A	7	11	20	5	72	0.4	36	臭氧
2021/7/14	监测点 A	8	21	21	7	70	0.5	35	臭氧
2021/7/15	监测点 A	6	13	13	5	75	0.4	38	臭氧

Table 9. AQI and primary pollutants measured at monitoring site B since July 13~July 15, 2021

表 9. 监测点 B 自 2021 年 7 月 13 日~7 月 15 日实测的 AQI 和首要污染物

预报日期	地点	二次模型日值预测						AQI	首要污染物
		SO ₂ (μg/m ³)	NO ₂ (μg/m ³)	PM10 (μg/m ³)	PM2.5 (μg/m ³)	O ₃ 最大 八小时滑动 平均 (μg/m ³)	CO (mg/m ³)		
2021/7/13	监测点 B	5	9	20	4	66	0.3	33	臭氧
2021/7/14	监测点 B	6	9	19	7	31	0.4	19	PM10
2021/7/15	监测点 B	6	6	16	5	81	0.3	41	臭氧

Table 10. AQI and primary pollutants measured at monitoring site C since July 13~July 15, 2021

表 10. 监测点 C 自 2021 年 7 月 13 日~7 月 15 日实测的 AQI 和首要污染物

预报日期	地点	二次模型日值预测						AQI	首要污染物
		SO ₂ (μg/m ³)	NO ₂ (μg/m ³)	PM10 (μg/m ³)	PM2.5 (μg/m ³)	O ₃ 最大 八小时滑动 平均 (μg/m ³)	CO (mg/m ³)		
2021/7/13	监测点 C	8	17	19	11	201	0.5	138	臭氧
2021/7/14	监测点 C	6	15	18	16	197	0.4	134	臭氧
2021/7/15	监测点 C	8	18	50	20	189	0.4	127	臭氧

由图 9 和图 10 的 test 拟合效果可以看出, 经过误差加权后的预测数据其拟合优度系数明显高于初始数据, 故误差加权后的预测数据效果要好于初始预测数据, 由此说明协同预报模型能够提升针对监测点 A、B、C 的污染物浓度预报准确度。

5. 结语

本文提出了基于 WRE-CMAQ-BP 神经网络下的二次建模, 先是用 SPSS 中的皮尔逊相关系数的方法得到各污染物浓度与各气象条件的相关性, 然后根据 BP 神经网络知识进行构建二次预测模型[7] [8]。本文采用预测监测点 A、B、C 在 2021 年 7 月 13 日至 7 月 15 日六种常规污染物单日浓度的相关数据, 发

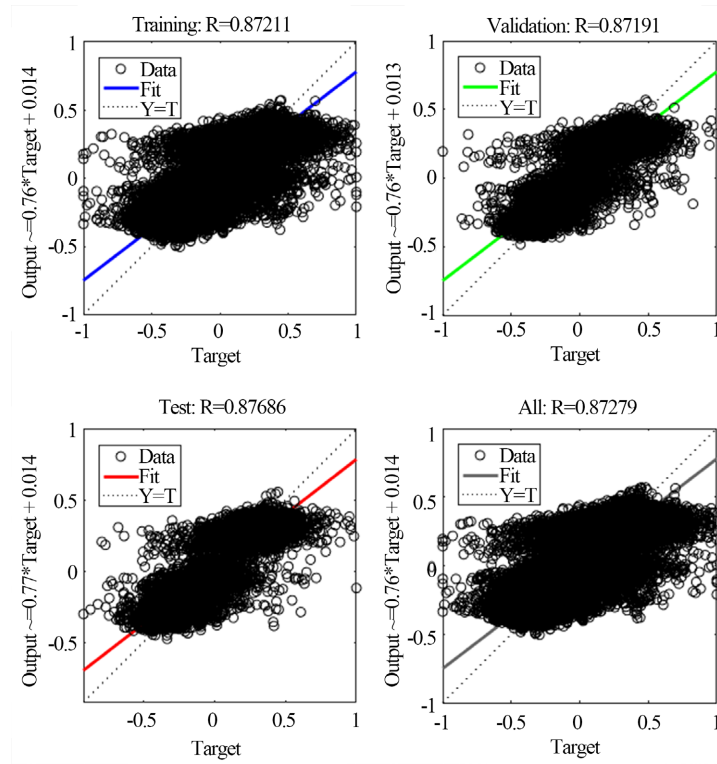


Figure 9. Initial fitting effect mse
图 9. mse 初始拟合效果图

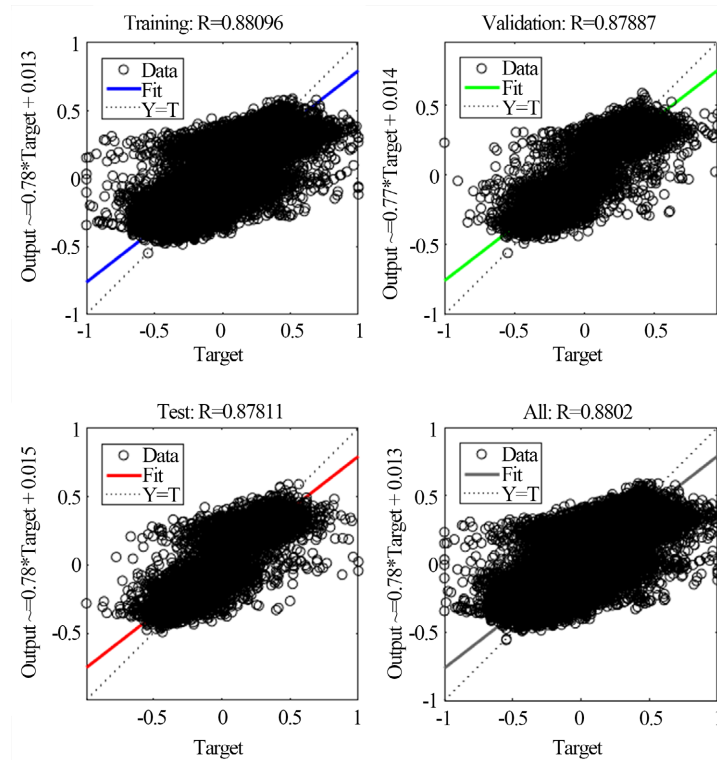


Figure 10. Error-weighted fitting effect graph
图 10. 误差加权拟合效果图

现二次建模模型比一次预测模型效果更优[9] [10]。本文的研究成果可以推广到其他未来的研究课题，例如：空气质量的变化预测趋势。

参考文献

- [1] 赵秀玲. 中国城市典型空气污染物时空分布特征与影响因素研究[D]: [博士学位论文]. 合肥: 中国科学技术大学, 2021.
- [2] 金宏, 刘焯焜. PM_{2.5}污染与气象条件的关系[J]. 长春工业大学学报, 2019, 40(4): 404-408.
- [3] 杜英魁, 张乙芳, 原忠虎, 关屏, 彭跃. 数据预处理对 LSTM 网络大气污染预测精度分析[J]. 计算机与数字工程, 2021, 49(7): 1400-1404+1425.
- [4] 刘青, 李典, 王立为, 徐亚琪, 吴宇童. 空气污染物浓度变化特征与气象因子的相关性建模研究[J]. 环境科学与管理, 2021, 46(4): 136-140.
- [5] Liu, H.X., Li, Q., Yu, D.B. and Gu, Y. (2019) Air Quality Index and Air Pollutant Concentration Prediction Based on Machine Learning Algorithms. *Applied Science*, **9**, 4069. <https://doi.org/10.3390/app9194069>
- [6] Perez, P. and Reyes, J. (2006) An Integrated Network Model for PM₁₀ Forecasting. *Atmospheric Environment*, **40**, 2845-2851. <https://doi.org/10.1016/j.atmosenv.2006.01.010>
- [7] 金仁浩, 曾国静, 王莎. 基于神经网络模型的空气质量预测研究[J]. 黑龙江科学, 2021, 12(12): 15-19.
- [8] LeCun, Y., Touresky, D., Hinton, G., *et al.* (1988) A Theoretical Framework for Back-Propagation. *Proceedings of the 1988 Connectionist Models Summer School*, **1**, 21-28.
- [9] Hecht-Nielsen, R. (1992) Theory of the Backpropagation Neural Network. *Neural Networks for Perception*, Academic Press, Cambridge, 65-93. <https://doi.org/10.1016/B978-0-12-741252-8.50010-8>
- [10] Chen, T. and Guestrin, C. (2016) Xgboost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, 13 August 2016, 785-794. <https://doi.org/10.1145/2939672.2939785>