

文章编号: 1674-9057(2019)02-0420-07

doi:10.3969/j.issn.1674-9057.2019.02.022

联合水汽因子的 GA-BP 神经网络 PM_{2.5} 质量浓度预测

刘林波, 刘立龙, 黎峻宇, 黄良珂, 周 威, 甘祥前

(桂林理工大学 a. 测绘地理信息学院; b. 广西空间信息与测绘重点实验室, 广西 桂林 541006)

摘 要: 针对北京地区频繁的雾霾灾害天气, 利用影响雾霾的主要因素作为预报因子, 采用遗传算法优化后的 BP(GA-BP)神经网络建立了 PM_{2.5} 质量浓度预测模型, 并对其进行可靠性分析。结果表明: 1) 通过 CMONOC 数据和相关气象温度、气压数据反演得到的 3 个 CMONOC 站点 BJFS、BJGB、BJSH 的大气水汽 (GPS-PWV) 数据与 PM_{2.5} 质量浓度之间的相关系数分别为 0.642 2、0.644 1、0.641 3, 呈中度相关; 2) 在预报因子不同的条件下, GA-BP 神经网络模型均比 BP 神经网络的预测值精度更高; 3) 利用 CO、NO₂、SO₂、O₃、湿度等 5 种主要影响因素作为预报因子, 采用 GA-BP 神经网络模型预测的 PM_{2.5} 数据, 其平均相对误差为 11.0%, 将这 5 种因素分别结合同一时段 3 个站点的 GPS-PWV 数据作为预报因子进行预测, 其平均相对误差分别降低了 2.8%、0.8%、3%; 4) 联合 GPS-PWV 数据与影响雾霾天气的主要因素, 基于 GA-BP 神经网络预测的 PM_{2.5} 值具有相对较好的精度和可靠性。

关键词: GPS-PWV; PM_{2.5}; GA-BP 神经网络; 相关性; 相对误差

中图分类号: P228.49; X513; X16

文献标志码: A

0 引 言

近年来, 我国多个城市(北京、天津、西安等)不断发生大规模的雾霾灾害天气, 现已成为制约我国区域经济发展和影响城市环境的热点问题^[1]。产生雾霾的主要原因是空气污染和一些气象因素的共同作用^[2], 在组成雾霾的众多污染物中, 最具有局域性特征、对人体危害最严重的污染物为大气中细颗粒物 PM_{2.5}, 其可以作为反映雾霾灾害严重程度的一个指标。研究表明, 雾霾灾害的成因与 CO、NO₂、SO₂、O₃ 等大气环境因子和湿度、温度、风速等气象因子密切相关^[3-5]。随着 GPS 气象学的发展, 文献[6]利用 2016 年北京市 GPS-PWV (GPS-Precipitable Water Vapor) 数据与

PM_{2.5} 质量浓度进行比较, 发现两者存在着明显的正相关性。文献[7-8]研究表明, GPS-PWV 是影响雾霾天气的一个重要因素, 可将其作为预测雾霾的一个关键参数。

针对雾霾与诸多大气环境和气象因子密切相关的这一特性, 国内外许多气象学者们都在研究建立科学、准确、实时的雾霾预测方法。由于人工神经网络方法可以解决预测数据的非线性和强突变特征等问题, 因此将其作为预测雾霾严重程度的主要方法^[9-11]。但是, 使用单一的 BP 神经网络模型预测的输出结果误差较大^[12], 且容易突变, 不能满足实际雾霾预测的要求。为解决这些重要问题, 本文采用了 GA-BP 神经网络模型进行预测。遗传算法^[13]是一种全局优化搜索算法, 可对 BP 神经网络

收稿日期: 2017-06-10

基金项目: 国家自然科学基金项目(41664002); 广西中青年教师基础能力提升项目(2017KY0267); 广西空间信息与测绘重点实验室基金项目(桂科能 16-380-25-27); 广西区八桂学者团队项目(002401006017)

作者简介: 刘林波(1993—), 男, 硕士, 研究方向: GNSS 技术及应用, 924134799@qq.com。

通讯作者: 刘立龙, 博士, 教授, hn_liulilong@163.com。

引文格式: 刘林波, 刘立龙, 黎峻宇, 等. 联合水汽因子的 GA-BP 神经网络 PM_{2.5} 质量浓度预测 [J]. 桂林理工大学学报, 2019, 39 (2): 420-426.

络的初始权值和阈值进行优化选择, 能够避开局部极小点, 克服 BP 神经网络的缺点, 两者的结合使得模型具有更高的稳定性和预测精度。

基于以上研究, 本文通过对比分析北京市 2015 年第 275 ~ 280 天 3 个陆态网络 (crustal movement observation network of China, CMONOC) 站点的 GPS - PWV 数据以及影响雾霾的其他大气环境数据、气象数据与 PM_{2.5} 质量浓度数据之间的相关程度, 利用 GPS - PWV 数据联合相关性较强的影响因素作为预报因子, 采用 GA-BP 神经网络模型进行 PM_{2.5} 质量浓度预测, 并将其与 BP 神经网络模型的预测值进行精度对比, 分析引入 GPS - PWV 因子能否提高 PM_{2.5} 数据的预测精度。

1 主要影响因素分析与选取

本实验采用了 CMONOC 基准站数据和 NOAA 气象站的相关气象数据, 通过萨斯塔莫宁 (Saastamoinen, SA) 模型联合反演求取北京市 2015 年第 275 ~ 280 天雾霾严重时期 3 个 CMONOC 基准站点 (BJFS、BJGB、BJSH) 天顶方向的 GPS - PWV 值。

SA 天顶静力学延迟计算模型为

$$\begin{cases} ZHD = (2.2768 \pm 0.0024) \frac{P_s}{f(\theta, H)}, \\ f(\theta, H) = 1 - 0.00266 \cos(2\theta) - 0.00028H. \end{cases} \quad (1)$$

式中: ZHD 为对流层天顶静力学延迟, mm; P_s 为地面气压, hPa; θ 指的是观测站的大地纬度, °; H 为测站的大地高, km。利用下载 CMONOC 基准站数据中对流层天顶总延迟 (ZTD) 减去 ZHD 即得到对流层湿延迟 (ZWD), 则大气水汽总量 (PWV) 可表示为

$$\begin{cases} PWV = K \times ZWD, \\ K = \frac{10^6}{\rho_w R_v (k_2' + k_3/T_m)}. \end{cases} \quad (2)$$

式中: K 代表转换系数; ρ_w 表示水的密度; R_v 表示水汽气体常数; k_2' 、 k_3 为大气折射常数; T_m 表示大气加权平均温度, K。

通过中国空气质量在线监测分析平台 (<http://www.aqistudy.cn/>) 下载得到 2015 年第 275 ~ 280 天北京市的 CO、NO₂、SO₂、O₃、PM_{2.5} 等大气环境数据及湿度、温度、风速等气象数据, 这 8 组数据时间分辨率均为 1 h, 每组共 144 个样本值。各影

响因素与 PM_{2.5} 之间的时间序列关系见图 1。

可知, 在北京市 2015 年第 275 ~ 280 天, 3 个 CMONOC 站的 GPS - PWV 以及 NO₂、CO 含量分别与 PM_{2.5} 质量浓度数据在时间序列的变化下具有很好的对应关系, 各影响因素的上升、下降过程对应了 PM_{2.5} 质量浓度数据的上升下降过程, 这种现象尤其明显地表现在各数据曲线的峰顶和低谷处。而 SO₂、O₃、风速、温度、湿度与 PM_{2.5} 质量浓度数据在时间序列的变化下的对应关系并不是特别明显。为进一步分析各因素与 PM_{2.5} 之间的关系, 本文采用相关性分析方法^[14] 表征其相关程度, BJFS 站 GPS - PWV、BJGB 站 GPS - PWV、BJSH 站 GPS - PWV、NO₂、SO₂、CO、O₃、风速、温度、湿度等影响因素与 PM_{2.5} 质量浓度的相关系数分别是 0.642 2、0.644 1、0.641 3、0.648 7、0.348 5、0.958 4、0.364 1、-0.248 4、0.198 4 和 0.451 1。

结合图 1 可知: 北京市 3 个站点的 GPS - PWV 与 PM_{2.5} 质量浓度在时间序列的变化下具有很好的对应关系, 总体呈中度相关, 因此由 GPS 反演的 GPS - PWV 可以作为本次预测 PM_{2.5} 质量浓度的一个重要参数。NO₂、SO₂、O₃、湿度与 PM_{2.5} 之间的也具有很好的相关性, 属中度相关。CO 与 PM_{2.5} 之间的相关系数更是达到了 0.958 4, 属高度相关。从整体上看, 风速和温度在该时段与 PM_{2.5} 数据的相关性较小, 属低度相关。

由雾霾的成因分析可知, 影响雾霾程度的因素很多, 通过分析各个影响因素与 PM_{2.5} 之间的相关性, 合理地选取影响雾霾程度的主要预报因子, 剔除相关性较弱的预报因子, 从而避免预报的冗余度, 这对于提高 PM_{2.5} 预测的精度至关重要, 进一步说明了本文工作的必要性。为保障 GA-BP 神经网络预测 PM_{2.5} 质量浓度的精度, 本文选取了 CO、NO₂、SO₂、O₃、湿度以及 GPS - PWV 数据作为预报因子, 以提高其预测精度。

2 GA-BP 神经网络模型的建立

BP 神经网络是近年来应用广泛的一种多层前馈神经网络, 其采用信号前向传递, 误差反向传播的方法, 并以均方根误差最小化为目标调整网络权值和阈值。这种学习方法解决了多层前馈神经网络的权重调整问题, 使得最终预测值不断逼

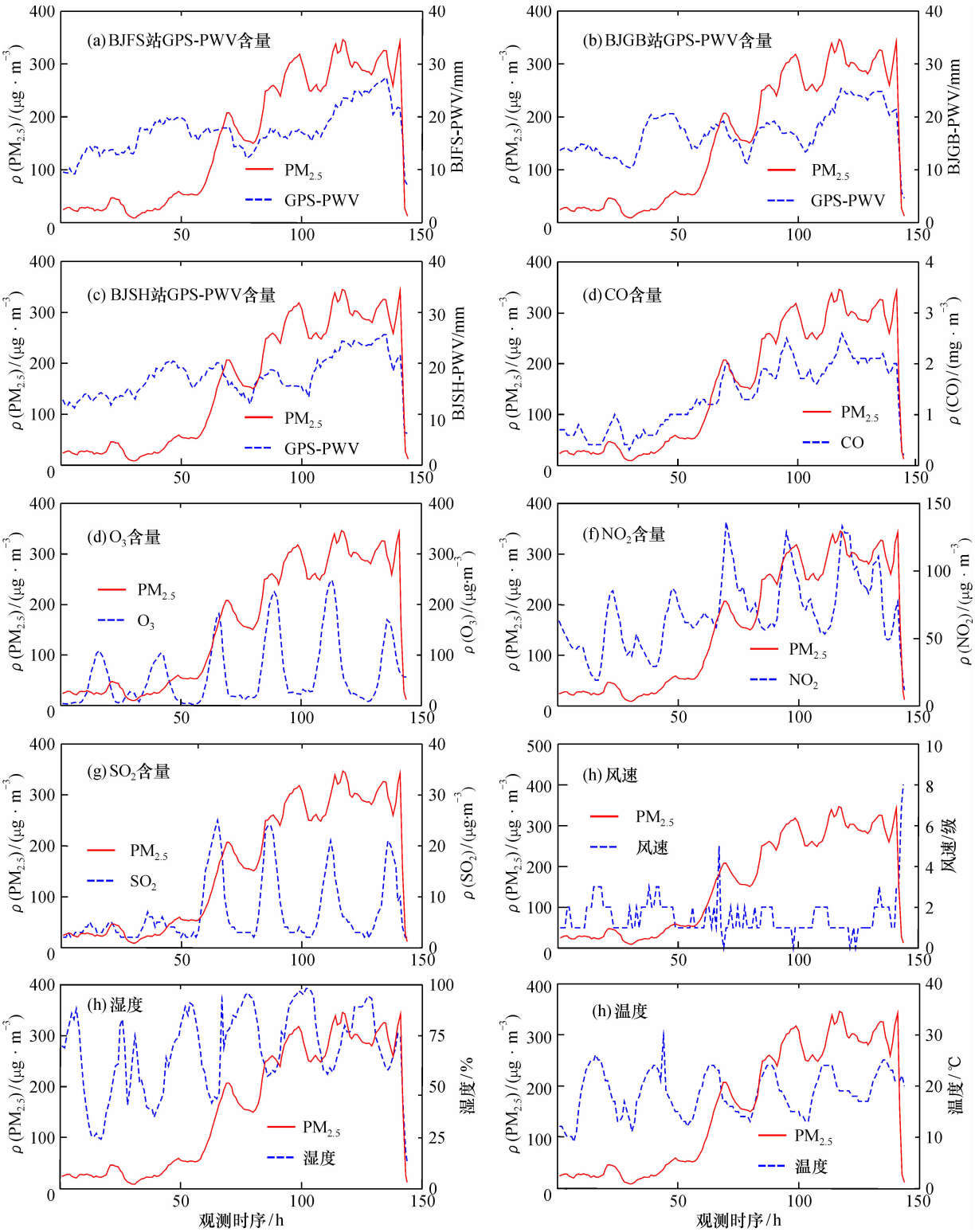


图1 各影响因素与PM_{2.5}质量浓度对比

Fig.1 Comparison of influencing factors and concentration of PM_{2.5}

近真实值。但在实际使用BP神经网络预测过程中发现其存在着隐含层神经元个数的最佳值没有定论、初始权值和阈值是随机产生、容易陷入局部极值等问题^[15]。

遗传算法是一种模拟自然界遗传机制和生物进化论的并行随机搜索最优化方法，具有良好的全局搜索性能，不易陷入局部极小等优点。遗传算法能够优化BP神经网络的初始权值和阈值，且

适用于并行处理，搜索不依赖于梯度信息，可以用来优化 BP 神经网络^[16]。遗传算法优化 BP 神经网络的公式为^[17]

$$\begin{cases} E_1(v,w,\zeta,\eta) = \frac{1}{2} \sum_{i=1}^{M_1} \sum_{t=1}^m [y_i(t) - \hat{y}_i(t)]^2; \\ s,t,w \in R^{m \times p}, v \in R^{p \times n}, \zeta \in R^p, \eta \in R^n. \end{cases} \quad (3)$$

式中： E_1 为网络训练样本总误差，其作为个体适应度倒数； $y_i(t)$ 为真值； $\hat{y}_i(t)$ 为网络的实际输出值； M_1 为训练样本个数； m 为输出层节点个数。

$$\hat{y}_i(t) = f\left\{ \sum_{i=1}^m v_{ui} g\left[\sum_{j=1}^p w_{ij} x_j(t) - \zeta_i \right] \right\} - \eta_i, \quad (4)$$

式中： w_{ij} 为输入层到隐含层节点的连接权值； v_{ui} 为隐含层到输出层节点的连接权值； ζ_i 为过程神经元的输出阈值； η_i 为输出神经元阈值； g 为隐含层神经元的激励函数； f 为输出层神经元的激励函数。其中，隐含层和输出层的激励函数均为 $f(x) = \frac{1}{1 + e^{-x}}$ 。设 $E^2 = \frac{1}{M - M_1} \sum_{k=1}^M \sum_{t=1}^m [y_i(t) - \hat{y}_i(t)]^2$ 为训练样本均方差，为使所设计的网络具有良好的泛化能力，应当使得 E^2 小于一个给定误差 ε ，以满足网络输出的可靠性。

GA-BP 模型预测 PM_{2.5} 质量浓度可分为 BP 预测模型结构确定、遗传算法优化 BP 预测模型和预测 3 部分。BP 神经网络结构可由拟合函数输入输出参数的个数确定，进而确定遗传算法个体的长度。遗传算法可优化 BP 神经网络的权值和阈值，可通过选择、交叉和变异操作找到最优适应度值对应的个体。BP 神经网络预测是将遗传算法得到的最优个体对网络初始权值和阈值赋值，网络经训练后预测输出。

经分析可知：在本文试验中，当不含水汽影响因子时，其输入参数 5 个、输出参数 1 个，所以设置的 GA-BP 神经网络输入层有 5 个节点，隐含层有 5 个节点（单隐层），输出层有 1 个节点，共有 $5 \times 5 + 5 \times 1 = 30$ 个权值， $5 + 1 = 6$ 个阈值，在遗传算法中采用实数法对个体进行编码，其长度为 $30 + 6 = 36$ ；当含水汽影响因子时，其输入参数 6 个，输出参数 1 个，所以设置的 GA-BP 神经网络输入层有 6 个节点，隐含层有 6 个节点，输出层有 1 个节点，共有 $6 \times 6 + 6 \times 1 = 42$ 个权值， $6 + 1 = 7$ 个阈值，遗传算法个体编码长度为 $42 + 7 = 49$ 。本次实验选择北京市 2015 年第 275 ~ 280 天

雾霾严重时期的 144 组数据，前 120 组（275 ~ 279 天）数据作为神经网络训练样本，后 24 组（280 天）作为预测样本。遗传算法的参数设置为：进化次数为 1 000，种群规模为 10，变异概率为 0.2，交叉概率为 0.4。其 GA-BP 神经网络模型通过 MATLAB R2014a 软件进行编程实现。

GA-BP 模型预测 PM_{2.5} 质量浓度网络算法流程图见图 2。

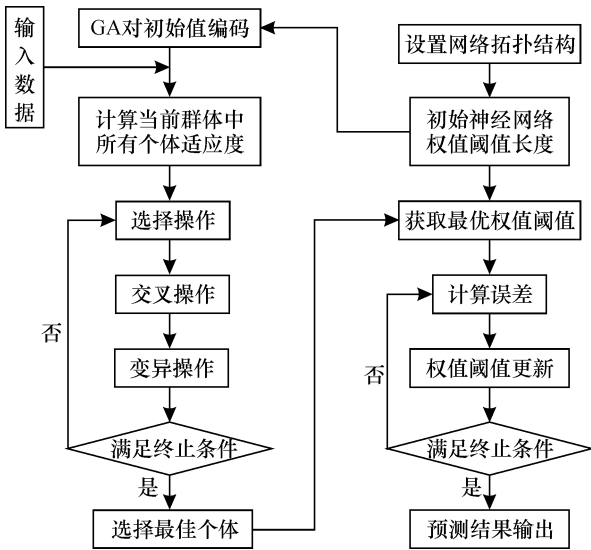


图 2 GA-BP 神经网络预测 PM_{2.5} 质量浓度流程图
Fig. 2 Prediction of PM_{2.5} mass concentration flow chart by GA-BP neural network

3 实验结果分析

本文利用北京市 2015 年第 275 ~ 279 天 120 h 的 PM_{2.5} 数据作为样本，分别通过 GA-BP 神经网络模型和 BP 神经网络模型预测第 280 天 24 h 的 PM_{2.5} 值，其中不含水汽预测值指将 CO、NO₂、SO₂、O₃、湿度作为输入因子的预测结果，含水汽预测值则是分别增加了各 CMONOC 站点（BJFS、BJGB、BJSH）的 GPS - PWV 数据作为输入因子的预测结果，分别如图 3、图 4 所示。

可知，在预报因子包含或不包含 GPS - PWV 的情况下，使用 GA-BP 神经网络模型和 BP 神经网络模型均可以预测出与实际 PM_{2.5} 值大致相近的数据。但是，通过对比两种模型的预测值和真实值之间的曲线图可以明显看出：整体上，GA-BP 神经网络的预测值曲线更加逼近真实值的曲线，且 4 条不同预报因子的预测值曲线靠拢得更加紧密，其在第 280 天的 22:00—24:00 PM_{2.5} 值突变的情况下预测值

更加稳定。为进一步分析这两种模型在不同预报因子下的预测效果,将其平均偏差 ($Bias = (\sum_{i=1}^n |y_i(t) - \hat{y}_i(t)|)/n$) 和均方根误差 ($RMSE = \sqrt{(\sum_{i=1}^n (y_i(t) - \hat{y}_i(t))^2)/n}$) 列于表 1。

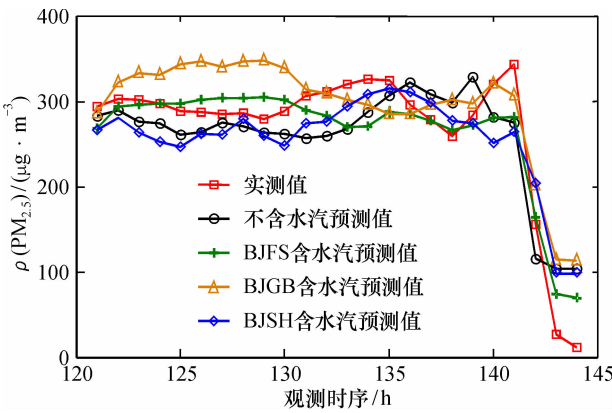


图3 2015 年第 280 天 24 h 实测 $PM_{2.5}$ 含量和 BP 神经网络预测 $PM_{2.5}$ 含量对比

Fig.3 Comparison of measured $PM_{2.5}$ content and BP neural network prediction of $PM_{2.5}$ content on 280th day and 24 hours in 2015

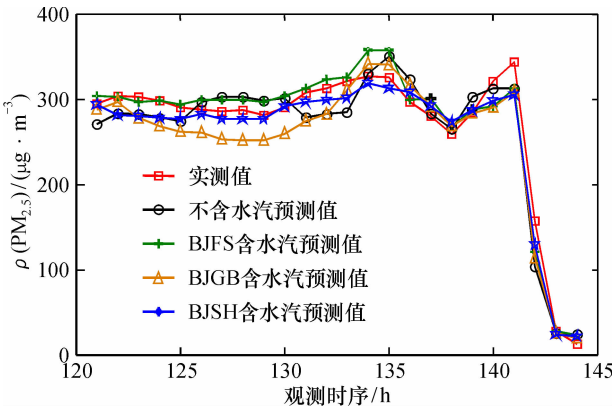


图4 2015 年第 280 天 24 h 实测 $PM_{2.5}$ 含量和 GA-BP 神经网络预测 $PM_{2.5}$ 含量对比

Fig.4 Comparison of measured $PM_{2.5}$ content and GA-BP neural network prediction of $PM_{2.5}$ content on 280th days and 24 hours in 2015

表 1 不同模型各组预报因子预测值的 Bias 和 RMSE

Table 1 Bias and RMSE of the predicted values of different forecasting factors for each model

预测因素	BP		GA-BP	
	Bias	RMSE	Bias	RMSE
不含 GPS - PWV	35.8	23.7	37.3	34.6
BJFS 含 GPS - PWV	41.3	30.1	45.5	40.8
BJGB 含 GPS - PWV	18.8	12.8	20.3	13.2
BJSH 含 GPS - PWV	22.2	16.8	23.4	16.0

综合可知,BP 神经网络模型预测值的 Bias 和 RMSE 均远远大于 GA-BP 神经网络模型预测值。通过对比分析两种模型的预测值可以看出:在预报因子不同的条件下,使用 GA-BP 神经网络模型预测的数据精度更高,数据稳定性更好,可以预测出接近实际的 $PM_{2.5}$ 数据。因此,本文将着重对 GA-BP 神经网络模型的预测值作出详细分析。

在 GA-BP 神经网络中,BJFS 和 BJSH 站点的 GPS - PWV 数据联合其他主要影响因素作为预报因子的预测值,其 Bias 和 RMSE 均小于不含 GPS - PWV 时的预测值,而含 BJGB 站点 GPS - PWV 预测值的 Bias 和 RMSE 分别为 20.3、23.4,大于不含 GPS - PWV 时的 18.8 和 22.2。这种情况是由于各站的高程不同引起的,3 个 CMONOC 站 BJFS、BJGB、BJSH 的海拔高为 87.5、213.0、155.4 m,均高于北京市的平均海拔 55.0 m, BJGB 站与北京市的海拔差值最大,达到 158 m。随着测站海拔与北京市的海拔高度差值增大,当大于 64 m 后, $PM_{2.5}$ 质量浓度会表现出随高度增加而递减的趋势,因此对于 BJGB 站点的预测值影响更加明显,使得预测值精度相对有所降低。

为进一步比较不同预报因子 GA-BP 神经网络预测值的优劣,利用各预测值的相对误差来进行精度分析,使其能更好地反映预测值的可靠程度。其中,相对误差 = $(|真实值 - 预测值| / 真实值) \times 100\%$,相对误差越小,预测精度越高。不同预报因子的预测值相对误差见图 5。

在预报因子不同的 4 种情况下,总体上模型预测的结果与真实监测值前 21 h 预测值相对误差

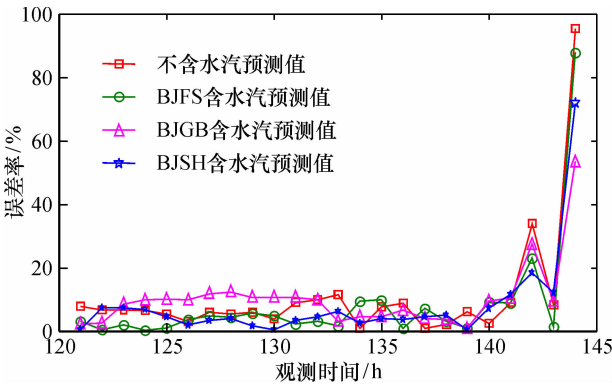


图5 2015 年第 280 天 24 h 各组 GA-BP 神经网络预测 $PM_{2.5}$ 含量相对误差对比

Fig.5 Comparison of the relative errors of GA-BP neural network prediction of $PM_{2.5}$ content in each group on 280th day and 24 hours in 2015

均在 10% 左右。在 22:00—24:00, 4 种情况下的预测值相对误差均产生极大变动, 尤其是在预测因子不含 GPS - PWV 的情况下的预测值相对误差变化最为明显。产生这种情况是由于风速的快速变化, 北京市在 2015 年第 280 天 22: 00 时风速 3 级, 随后在接下来的 1 h 内迅速增加到 7 级, 最终 24:00 时达到 8 级。随着风速的增加, 大气中的 PM_{2.5}微颗粒物被吹散, PM_{2.5}数值由 21: 00 的 343 μg/m³ 急速降到 24: 00 的 12 μg/m³。为更好地描述不同预报因子的预测误差, 将相对误差详细情况列于表 2。

4 种预测因素预测值的相对精度在 21:00 前都比较高, 其相对误差都在 10% 以内, 在 22:00—24:00 内, 其预测精度均有所降低。从 24:00 整体来看, 预报因子中分别包含 3 个 CMONOC 站 GPS - PWV 数据的预测值, 相对误差分别为 8.2%、10.2%、8.1%, 较单纯使用其他影响因子预测的相对误差 11.0% 分别降低了 2.8%、0.8%、3%, 其误差相对较稳定。这表明, 引入 GPS - PWV 数据联合其他影响雾霾的主要大气环境数据和气象数据作为预报因子可明显提高预测 PM_{2.5}质量浓度的精度和可靠性。

4 结 论

本文分析了北京市 2015 年第 275 ~ 280 天各雾

霾影响因素与 PM_{2.5}质量浓度之间的相关性, 为提高模型预测值精度, 选取了 CO、NO₂、SO₂、O₃、湿度数据并将其分别与 BJFS、BJGB、BJSH 3 个 CMONOC 站反演得到的 GPS - PWV 数据联合作为预报因子的 4 种情况, 采用 BP 神经网络模型和 GA-BP 神经网络模型分别预测了 PM_{2.5}质量浓度值。通过分析这 4 种情况下两种模型预测出的 PM_{2.5}值与真实值的误差, 结果表明: 在预报因子不同的条件下, 使用 GA-BP 神经网络模型比 BP 神经网络模型预测的数据精度更高, 更加稳定; 在 GA-BP 神经网络预测 PM_{2.5}质量浓度时, 利用影响雾霾的主要因子联合不同高程站点的 GPS - PWV 数据预测的精度有所差异, 测站高程与当地高程相差越大, 预测值精度相对越低; 但就整体而言, 使用包含 GPS - PWV 数据比单纯使用影响雾霾主要的大气环境因素和气象因素作为预报因子的 GA-BP 神经网络预测值相对误差小, 其预测值的精度和可靠性均有所提高; 因此, 将 GPS - PWV 数据联合影响雾霾主要的大气环境因素和气象因素作为 GA-BP 神经网络模型预测 PM_{2.5}含量的主要预报因子, 有利于提高预测 PM_{2.5}含量的精度和可靠性。本文仅针对于北京市的 PM_{2.5}预测得到验证, 对于其他城市添加水汽因素对于预测 PM_{2.5}可以起到一定的参考作用。

表 2 不同组预报因子的 GA-BP 神经网络预测值的相对误差
Table 2 Relative error of GA-BP neural network predictive value for different group forecasting factors %

预报因子	max	min	时 段				平均值
			21:00 前	22:00	23:00	24:00	
不含 GPS - PWV	95.4	1.0	6.1	33.9	8.2	95.4	11.0
BJFS 含 GPS - PWV	87.8	0.3	4.1	23.1	1.3	87.8	8.2
BJGB 含 GPS - PWV	53.4	0.6	7.3	27.2	9.4	53.4	10.2
BJSH 含 GPS - PWV	71.8	0.3	4.3	18.2	12.1	71.8	8.0

参考文献:

[1] Pan X L, Yan P, Tang J, et al. Observational study of influence of aerosol hygroscopic growth on scattering coefficient over rural area near Beijing mega-city [J]. Atmospheric Chemistry and Physics, 2009, 9: 7519 - 7530.

[2] 戴树桂. 环境化学进展 [M]. 北京: 化学工业出版社, 2005: 1 - 50.

[3] 宋宇, 唐孝炎, 张远航, 等. 北京市大气能见度规律及下降原因 [J]. 环境科学研究, 2003, 16 (2): 10 - 12.

[4] Tao J, Ho K F, Chen L G, et al. Effect of chemical composition of PM_{2.5} on visibility in Guangzhou, China, 2007 spring [J]. Particuology, 2009, 7 (1): 68 - 75.

[5] Sun Y L, Zhuang G S, Tang A H, et al. Chemical characteristics of PM_{2.5} and PM₁₀ in haze-fog episodes in Beijing [J]. Environmental Science & Technology, 2006, 40 (10): 3148 - 3155.

[6] 王勇, 刘严萍, 李江波, 等. 不同季节的北京市 GPS 可降水量与 PM_{2.5}/PM₁₀的比较研究 [J]. 大地测量与地球动力学, 2015, 35 (3): 365 - 367.

[7] 王勇, 刘严萍, 李江波, 等. 河北省 GPS 水汽与 PM_{2.5}质量浓度的比较研究 [J]. 大地测量与地球动力学, 2016, 36 (1): 40 - 42.

[8] 王勇, 刘严萍, 李江波, 等. 水汽和风速对雾霾中 PM_{2.5}/

PM₁₀变化的影响 [J]. 灾害学, 2015, 30 (1): 5 – 7.

[9] Hsieh W W, Tang B Y. Applying neural network models to prediction and data analysis in meteorology and oceanography [J]. Bulletin of the American Meteorological Society, 1998, 79 (9): 1855 – 1870.

[10] Voukantsis D, Karatzas K, Kukkonen J, et al. Intercomparison of air quality data using principal component analysis, and forecasting of PM₁₀ and PM_{2.5} concentrations using artificial neural networks, in Thessaloniki and Helsinki[J]. Science of the Total Environment, 2011, 409(7): 1266 – 1276.

[11] Pasini A, Pelino V, Potesta S. A neural network model for visibility nowcasting from surface observations: results and sensitivity to physical input variables [J]. Journal of Geophysical Research, 2001, 106 (D14): 14951 – 14959.

[12] 艾洪福, 石莹. 基于 BP 人工神经网络的雾霾天气预测

研究 [J]. 计算机仿真, 2015, 32 (1): 402 – 405, 415.

[13] 梁月吉, 任超, 刘立龙, 等. 基于 EMD 和 GA-BP 神经网络的大坝变形预测 [J]. 桂林理工大学学报, 2015, 35 (1): 111 – 116.

[14] 姜高霞, 王文剑. 时序数据曲线排齐的相关性分析方法 [J]. 软件学报, 2014, 25 (9): 2002 – 2017.

[15] 仇丽霞. 基于遗传算法的最优决策值选择及医药学应用研究 [D]. 太原: 山西医科大学, 2007.

[16] 石刚, 井元伟, 徐皑冬, 等. 利用改进的遗传算法解决全局寻优问题 [J]. 仪器仪表学报, 2006, 27 (S3): 2329 – 2332.

[17] 彭基伟, 吕文华, 行鸿彦, 等. 基于改进 GA-BP 神经网络的湿度传感器的温度补偿 [J]. 仪器仪表学报, 2013, 34 (1): 153 – 160.

Prediction of PM_{2.5} mass concentration
based on GA-BP neural network with water vapor

LIU Lin-bo, LIU Li-long, LI Jun-yu, HUANG Liang-ke, ZHOU Wei, GAN Xiang-qian
(a. Guangxi Key Laboratory of Spatial Information and Geomatics; b. College of Geomatics and Geoinformation, Guilin University of Technology, Guilin 541006, China)

Abstract: The concentration of PM_{2.5} can act as an index of the severity of haze. According to the fact that haze occurs frequently in Beijing, a concentration prediction model based on BP (GA-BP) neural network optimized by genetic algorithm is established under the condition of regarding the main influence factors of haze as forecast factors. The reliability of the prediction model was analyzed. The results prove that: (1) The correlation coefficients among the GPS – PWV of BJFS, BJGB and BJSH retrieved by the CMONOC data and relative meteorological data such as temperature, pressure and the concentration of PM_{2.5} are respectively 0.642 2, 0.644 1 and 0.641 3. They are moderately correlated; (2) GA-BP model is more accurate than BP model with different prediction factors; (3) CO, NO₂, SO₂, O₃ and humidity were used as the forecast factors to predict the concentration of PM_{2.5} based on GA-BP. The average relative error is 11.0%. In condition of GPS – PWV added to the forecast factors, the average relative error was deduce in 2.8%, 0.8% and 3%. The five factors were combined with three sites for the same period; (4) With the combination of GPS – PWV data and main influence factors of haze, the PM_{2.5} value based on GA-BP neural network prediction has relatively good accuracy and reliability.

Key words: GPS – PWV; PM_{2.5}; GA-BP neural network; correlation; relative error