

吴玉霜,黄小燕,陈家正,等.机器学习在广西台风极大风速预报中的应用[J].气象研究与应用,2021,42(4):26–31.

Wu Yushuang, Huang Xiaoyan, Chen Jiazheng, et al. Application of machine learning in forecasting maximum wind speed of typhoon in Guangxi[J]. Journal of Meteorological Research and Application, 2021, 42(4): 26–31.

## 机器学习在广西台风极大风速预报中的应用

吴玉霜<sup>1,2</sup>, 黄小燕<sup>3\*</sup>, 陈家正<sup>4</sup>, 赵华生<sup>3</sup>

(1.广西壮族自治区气象台, 南宁 530022; 2.广西壮族自治区气象灾害防御技术中心, 南宁 530022;

3.广西壮族自治区气象科学研究所, 南宁 530022; 4.广西民族大学, 南宁 530006)

**摘要:**以 1980—2020 年广西台风期间桂林、梧州、龙州、南宁、玉林等 5 个气象观测站的地面日极大风速为研究对象,采用多元线性回归(MR)、支持向量机(SVM)、模糊神经网络(FNN)等三种较为常用的线性和非线性方法分别进行预报建模,对 2011—2020 年共 10a 独立样本的检验。结果表明,在全样本风速预报的平均绝对误差上,FNN 模型对桂林站、梧州站、龙州站、玉林站共 4 个站点预报的平均绝对误差最小,总体预报精度最好,MR 预报模型则对南宁站有较好的预报能力,SVM 模型预报效果总体偏差。对于 6 级以上大风的 TS 评分、命中率、空报率和预报偏差等 4 个评估指标的统计,FNN 模型的预测精度最高且相对稳定,MR 方案次之,SVM 在三种方案中预报效果最差。FNN 方法对广西台风期间地面日极大风速的预报有较好的参考作用。

**关键词:**机器学习;台风;极大风;预报建模

**中图分类号:** P457.8

**文献标识码:** A

**doi:** 10.19849/j.cnki.CN45-1356/P.2021.4.05

**OSID:**

## 引言

近年来,许多专家学者采用机器学习方法对台风大风进行预报建模研究,尝试从具有随机波动性、不确定性和非线性的风速数据间挖掘出变化规律,提高风速预报的精度<sup>[1-5]</sup>。杨晓君等<sup>[6]</sup>利用了天津中尺度天气预报模式产品和 EC 数值模式产品,建立了渤海海风的 BP 神经网络两级海风预报模型,对于灾害大风的预报取得了较好的效果。朱智慧等<sup>[7]</sup>利用 2011—2012 年台风影响期间上海沿海的 5 个浮标站的 2min 平均风速和极大风速,运用回归分析方法得到了台风期间沿海极大风速的客观方程,并将客观方程带入 WRF 模式进行极大风速预报,结果较为理想。钱燕珍等<sup>[8]</sup>将支持向量机方法应用于台风影响下的站点风速预报,表明在适当的样本截取和预报因子选取后,风速预报 48h 内效果较好。

目前对于台风影响广西期间的地面站点极大风速的预报研究相对较少,客观定量预报方法缺乏,主要还是根据预报员的主观分析进行实际业务预报<sup>[9-11]</sup>。陈润珍等<sup>[12]</sup>利用线性回归方法和旋衡风方程,建立了广西沿海台风大风预报模型,预报效果较好。董彦等<sup>[13]</sup>采用遗传-神经网络集合预报方法预报广西台风大风,模型对 $\geq 10\text{m}\cdot\text{s}^{-1}$ 的强风有较好的预报性能。本研究尝试以台风影响广西期间部分地面站点的日极大风速作为预报对象,采用多元线性回归(Multiple Regression, MR)、支持向量机(Support Vector Machine, SVM)、模糊神经网络(Fuzzy Neural Network, FNN)等三种较为常用的线性和非线性机器学习方法分别进行预报建模研究,讨论分析和对比三种不同的机器学习算法对广西台风极大风速的预报效果,为台风影响广西期间的地面极大风速预报工作提供预报参考。

收稿日期: 2021-11-10

基金项目: 国家自然科学基金项目(41765002)、广西自然科学基金重点项目(2017GXNSFDA198030)、广西气象局重点基金项目;桂气科 2021Z02、广西台风与海洋预报服务创新团队项目

作者简介: 吴玉霜(1994—),女,助理工程师,从事专业气象服务。E-mail: 1248893015@qq.com

\* 通讯作者: 黄小燕(1978—),女,博士,正研级高级工程师,从事智能计算预报技术在天气预报中的应用。E-mail: Gx\_huangxy@163.com

## 1 研究方法

### 1.1 多元线性回归

本文将多元线性回归(MR)模型设定如下,设预报量  $Y$ :

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m + \varepsilon \quad (1)$$

式中  $\beta_0, \beta_1, \dots, \beta_m$  均为对应特征向量  $x_1, x_2, \dots, x_m$  的回归系数,  $\varepsilon$  为随机残差。

### 1.2 支持向量机方法

支持向量机(SVM)以结构风险最小化原则为理论基础,首先在低维空间进行计算,通过核函数将低维的原始数据非线性映射到高维的新特征空间,最终在高维特征空间中转换为线性学习构造出最优分离超平面。在实际问题中,大量风速样本数据为非线性集合,SVM 对非线性数据有较好的处理能力<sup>[14-15]</sup>。对 SVM 非线性回归模型,预报量与指标量的一般形式为:

$$f(x) = w^T \varphi(x) + b \quad (2)$$

式中:  $w$  是法向量,  $b$  是位移项,  $\varphi(x)$  是非线性映射函数。

引入拉格朗日因子和核函数  $K(x_i \cdot x_j)$ , 本文主要选取的是多项式核函数作为模型的核函数,可以得到 SVM 非线性回归形式如下:

$$\max_a \sum_{i=1}^m a_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m a_i a_j y_i y_j K(x_i, x_j) \quad (3)$$

$$\text{s.t.} \sum_{i=1}^m a_i y_i = 0, 0 \leq a_i \leq C, i = 1, 2, \dots, m \quad (4)$$

式中:  $a_i$  为最大化目标函数。

### 1.3 模糊神经网络

模糊神经网络(FNN)主要以模糊控制理论为基础,融合了模糊系统的非线性处理能力以及人工神经网络的自学能力等优点<sup>[16]</sup>。这里采用的是一种结论为数值型的模糊神经网络,由输入层、隶属函数层、推理层和反模糊化输出层四个环节组成:

$$\text{输入层: 输入状态变量 } \chi = [x_1, x_2, \dots, x_i] \quad (5)$$

隶属函数层: 通过采用高斯函数可以得到状态变量  $x_1$  的隶属度  $\mu_{ij}$ , 其基本表达式为:

$$\mu_{ij} = \exp \left[ -\frac{(x_i - a_{ij})^2}{\sigma_{ij}^2} \right] \quad (6)$$

其中  $1 \leq i \leq n; 1 \leq j \leq m, \mu_{ij}, a_{ij}, \sigma_{ij}$  与隶属度函数层的各节点相对应, 输出  $\mu_{ij}$  的下标依次为  $\mu_{11}, \mu_{12}, \dots, \mu_{1m}; \mu_{21}, \mu_{22}, \dots, \mu_{2m}; \dots; \mu_{n1}, \mu_{n2}, \dots, \mu_{nm}$ 。

推理层: 包含了模糊规则和模糊推理两个部分, 在模糊逻辑中, 模糊推理是根据建立的模糊规则进

行输出, 本文按照“and”运算和 IF-THEN 的判断形式构建模糊规则形式:

$$\text{if } x_1 \text{ is } \mu_{1k} \text{ and } x_2 \text{ is } \mu_{2k}, \dots, \text{ then } y = \omega_k, k = 1, \dots, m \quad (7)$$

运用相乘的方式计算模糊规则中各节点的输出值, 分别是该节点所有输入的代数乘积  $\Pi$ :

$$\pi_j = \mu_{1j} \times \mu_{2j} \times \cdots \times \mu_{mj} = \prod_{i=1}^n \mu_{ij}, (j = 1, \dots, m) \quad (8)$$

输出层: 使用反模糊化网络进行输出:

$$y = \omega_1 \pi_1 + \omega_2 \pi_2 + \cdots + \omega_m \pi_m \quad (9)$$

其中  $\omega_j (j = 1, 2, \dots, m)$  为连接权。

## 2 台风极大风速预报试验

### 2.1 预报资料

本文主要是为了对比评估 3 种预报方法对台风极大风速的预报效果, 因此选取预报对象为台风进入广西影响范围时 ( $19^\circ\text{N}$  以北,  $112^\circ\text{E}$  以西) 的地面日极大风速值。由于统计预报方法需要较大样本的数据才能进行更好地建模预报, 为此, 挑选了广西站点日极大风速记录中有较长时间序列并且资料较为完善的站点, 从而得到 1980—2020 年位于广西东部的梧州、东南部的玉林、中南部的南宁、北部的桂林和西南部的龙州 5 个气象观测站的地面日极大风速实测值, 资料来自广西气象局气象信息中心。

台风资料选取 1980—2020 年共 41a 影响广西的台风路径、中心风速、最大风速等数据集资料, 来自 CMA 最优台风路径数据集 (<https://tcdata.typhoon.org.cn/>)。物理量风速预报资料选自欧洲中期天气预报中心(ECMWF)ERA5 再分析数据的各层各要素场的数据资料。时间分辨率是 6h (02:00, 08:00, 14:00 和 20:00), 格距:  $0.25 \times 0.25$ ; 纬度:  $9.75^\circ\text{N} \sim 40.5^\circ\text{N}$ ; 经度为  $79.5^\circ\text{E} \sim 120^\circ\text{E}$ , 共计 61364 个格点。为了对预报方法进行客观比较与检验, 统一选取 1980—2010 年数据作为建模样本, 2011—2020 年的样本作为独立测试样本进行预报测试, 其中需对缺失样本进行剔除处理。表 1 为各个气象站用于预报试验的样本数。

表 1 广西台风地面日极大风速预报建模样本(d)

气象站	样本总数	建模样本数	独立样本数
桂林	451	336	115
梧州	387	273	114
龙州	447	335	112
南宁	451	336	115
玉林	448	333	115

## 2.2 预报因子初选

### 2.2.1 物理量预报因子

将  $18.25^{\circ}\text{N}\sim 29.75^{\circ}\text{N}$ ,  $100.25^{\circ}\text{E}\sim 114.75^{\circ}\text{E}$  范围 ERA5 再分析的各层要素场和物理量场格点数据 (共计 15342 个格点, 图 2) 作为待选预报因子, 通过相关分析和显著性检验, 桂林站入选预报因子数量为 70 个, 梧州站入选 81 个, 龙州站入选 59 个, 南宁站入选 81 个, 玉林站入选 76 个。

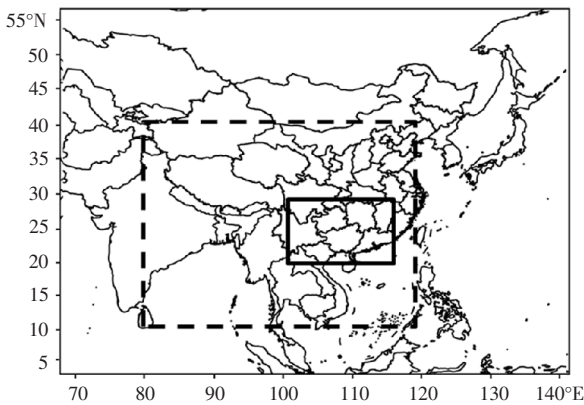


图 1 原物理量场选择范围(短断线方框)以及裁剪后预报因子场区域(实线方框)图

### 2.2.2 台风预报因子

由于研究对象是台风造成的极大风速, 因此台风本身所处的经纬度、气压、中心风速等特征量也是预报地面极大风速的重要因素。这里选取了起报时刻台风所处的经度和纬度、中心最低气压、中心附近最大风速、地面极大风速、过去 24h 变压、前 24h 中心附近最大风速增量等 7 个台风气候持续预报因子作为极大风速预报的待选因子。分析发现, 入选的台风预报因子与风速的相关有地域性的差异, 相关率呈现为玉林、南宁、梧州、龙州到桂林的逐渐递减。究其原因, 可能影响广西的台风大多数是从东南部的玉林市进入的, 相关高 ( $\geq 0.04$ ); 相邻的南宁和梧州影响次之; 北部的桂林和西南部的龙州一般处于台风影响后期, 因而相关性偏小。

## 2.3 构建预报模型

本文以 1980—2010 年的数据作为基础建模样本, 对 2011—2020 年共 10a 台风影响广西期间 5 个气象站的地面日极大风速样本作为独立预报样本进行预报试验, 选择机器学习中的 MR、SVM、FNN 三种方法分别构建不同站点的台风极大风速预报模型。模型构建思路如下:

(1) 建立 MR 台风极大风速预报模型: 经过多

次试验, 当设定多元线性回归中的统计量  $F$  值=3 的入选因子数的结构较为适宜, 得到的预报值也是最优, 各站点分别入选的预报方程因子数量, 桂林站为 6 个, 梧州站入选 5 个, 龙州站 8 个, 南宁站 8 个, 玉林站 8 个。利用确定的预报因子, 分别建立了 5 个站点的台风极大风速预报方程, 由于每个站点的建模样本量不同, 回归方程的系数也不同, 得到方程如下:

$$\text{桂林: } Y_{\text{桂林}} = 8.05 - 0.008 \times X_{17} - 0.01 \times X_{20} - 0.57 \times X_{33} - 0.07 \times X_{36} - 0.0002 \times X_{53} - 0.0001 \times X_{59} \quad (10)$$

$$\text{梧州: } Y_{\text{梧州}} = 7.66 - 0.005 \times X_{33} - 0.799 \times X_{42} - 0.111 \times X_{50} - 0.0003 \times X_{73} - 0.0002 \times X_{79} \quad (11)$$

$$\text{龙州: } Y_{\text{龙州}} = 72.19 - 0.066 \times X_1 - 0.0001 \times X_7 - 0.061 \times X_{30} - 0.291 \times X_{42} - 7.116 \times X_{45} - 0.985 \times X_{49} - 0.569 \times X_{51} - 0.774 \times X_{54} \quad (12)$$

$$\text{南宁: } Y_{\text{南宁}} = 65.7 - 0.058 \times X_1 - 0.0001 \times X_{10} - 0.0001 \times X_{12} - 0.332 \times X_{39} - 0.082 \times X_{42} - 0.123 \times X_{47} - 1.243 \times X_{63} - 1.181 \times X_{66} \quad (13)$$

$$\text{玉林: } Y_{\text{玉林}} = 8.17 + 0.116 \times X_4 - 0.016 \times X_{30} - 0.167 \times X_{37} - 1.912 \times X_{53} - 1.031 \times X_{55} - 1.215 \times X_{58} - 1.888 \times X_{59} + 0.049 \times X_{70} \quad (14)$$

式(10)—(14)中  $Y_{\text{桂林}}$ 、 $Y_{\text{梧州}}$ 、 $Y_{\text{龙州}}$ 、 $Y_{\text{南宁}}$ 、 $Y_{\text{玉林}}$  表示极大风速的预报值,  $X$  表示通过多元线性回归方案设定  $F$  值为 3 时自动筛选得到的预报因子, 下标为预报因子在所有初选预报因子集合排序中所在的序号。

(2) 建立 SVM 台风极大风速预报模型: 为了进行客观对比, 在建立 SVM 台风极大风速预报模型时, 同样以多元线性回归方案筛选出来的预报因子集作为预报模型的因子输入, 实际观测的极大风速值作为训练目标样本。本方案设置拉格朗日乘子上界为 200, 并构造核函数  $K(x, x_i) = [(x \cdot x_i) + 1]^q$ , 其中  $q$  是多项式的阶次, 所得到的是  $q$  阶多项式分类器。最后利用三次多项式求解线性方程得到  $a_i$  和  $b$ , 建立 SVM 非线性回归模型对台风极大风速进行预报。

(3) 建立 FNN 台风极大风速预报模型: 采用模糊神经网络预报模型对广西台风期间地面日极大风速进行预报试验时, 因子输入和上述 2 个模型一致, 这里的模型参数统一设置为: 网络的输入节点为筛选出的各站点的预报因子数量, 网络的输出节点为 1, 设置 3 个推理层节, 网络训练次数设定为 300 次, 学习因子取 0.9, 总体误差定为 0.0001。



3 三种机器学习方法预报结果对比分析

按照 2.3 节构建的预报建模试验思路, 三种机器学习方法对广西 5 个气象站点台风期间地面日极大风速的预报结果见表 2。平均绝对误差(MAE)可以很好地反映预报模型的总体预报精度情况, 因此本文主要采用该评估指标对三种机器学习方法的独立样本预报结果进行分析评估。分析表 2 可知, 三种预报模型的平均绝对误差在  $1.7\sim 3.3\text{m}\cdot\text{s}^{-1}$  之间, 三种方案对桂林站的预报效果最理想, 其次为龙州站, 误差最大均为玉林站。进一步分析可知, MR 预报模型 5 个站点的预报平均绝对误差为  $2.53\text{m}\cdot\text{s}^{-1}$ , 其中预报误差最小为桂林站  $2.25\text{m}\cdot\text{s}^{-1}$ , 误差最大为玉林站的  $2.96\text{m}\cdot\text{s}^{-1}$ 。SVM 预报模型 5 个站点的预报平均绝对误差为  $2.76\text{m}\cdot\text{s}^{-1}$ , 其中预报误差最小为桂林站  $2.28\text{m}\cdot\text{s}^{-1}$ , 误差最大为玉林站的  $3.29\text{m}\cdot\text{s}^{-1}$ 。FNN 预报模型 5 个站点的预报平均绝对误差为  $2.38\text{m}\cdot\text{s}^{-1}$ , 其中预报误差最小为桂林站  $1.74\text{m}\cdot\text{s}^{-1}$ , 误差最大为玉林站的  $2.92\text{m}\cdot\text{s}^{-1}$ 。对比分析可知, FNN 预报模型对桂林站、梧州站、龙州站、玉林站共 4 个站点预报的平均绝对误差最小, 总体预报精度最好; 而 MR 预报模型对南宁站预报的平均绝对误差最小, 该方法对南宁站有较好的预报能力。

其中 FNN 预报模型的平均绝对误差比 MR 预

表 2 三种机器学习预报模型平均绝对误差统计(单位:  $\text{m}\cdot\text{s}^{-1}$ )

F=3	MR	SVM	FNN
桂林	2.25	2.28	1.74
梧州	2.48	2.63	2.42
龙州	2.35	2.60	2.01
南宁	2.59	3.01	2.83
玉林	2.96	3.29	2.92

报模型减少了 1%~29%(除南宁站外)。FNN 预报模型的平均绝对误差比 SVM 预报模型减少了 6%~29%。MR 预报模型的平均绝对误差比 SVM 预报模型减少了 5%~13%(除桂林站外)。

本文把预报绝对误差  $\leq 2\text{m}\cdot\text{s}^{-1}$  的样本视为预报效果较好, 表 3 是三种机器学习预报模型对 5 个站点的独立预报样本预报较好的样本个数情况。分析可知, 对于桂林、龙州和玉林等 3 个站点的预报, FNN 预报模型预报效果好的样本个数最多; MR 和 SVM 预报模型预报对梧州站预报效果好的样本个数一样多; 对南宁站的预报, MR 预报模型表现最优。

总体上, 通过 MAE 检验统计的分析表明, FNN 预报模型的预报精度最高, MR 模型低于 FNN 预报模型但优于 SVM 预报模型。

表 3 三种机器学习预报模型平均绝对误差个数对比

站点 预报模型	桂林			梧州			龙州			南宁			玉林		
	MR	SVM	FNN	MR	SVM	FNN	MR	SVM	FNN	MR	SVM	FNN	MR	SVM	FNN
MAE $\leq 2 (\text{m}\cdot\text{s}^{-1})$	56	55	75	64	64	60	55	55	65	57	52	54	52	53	55
比例(%)	49	48	65	56	56	53	49	49	58	50	45	47	45	46	48

在实际的业务预报中, 更关注的是 6 级( $10.8\text{m}\cdot\text{s}^{-1}$ )以上风速的预报, 为了分析三种机器学习模型对广西台风期间地面 6 级以上日极大风速的预报能力, 这里重点分析了 2011—2020 年 10a 期间各个站点的实际 6 级以上风速的风速预报情况。主要采用 TS 评分、命中率、空报率和预报偏差等 4 种检验方法进行对比检验。

从为三种预报模型 6 级以上风速预报的 TS 评分(图 2a)的对比分析可知, 5 个站点(桂林、梧州、龙州、南宁、玉林)的 MR 预报模型的 TS 评分为: 0.22、0.32、0.09、0.30、0.36; SVM 预报模型的 TS 评分为: 0.24、0.28、0.11、0.26、0.32; FNN 报模型 FNN 的 TS

评分为: 0.29、0.33、0.06、0.21、0.45。对于桂林、梧州和玉林 3 个站点, FNN 预报模型的评分均高于 MR 预报模型和 SVM 预报模型, 分别提高了 22.5%、2.04%、19.7%和 17.2%、13.4%、27.7%。分析发现, 对龙州站的预报中, FNN 预报模型和 MR 预报模型的预报效果非常差, 而 SVM 预报模型相对最理想。在 MR 预报模型的评分略高于 SVM 预报模型, 除了桂林和龙州站, 其余站点提高了 11.6%、12.1%和 10%。

命中率结果显示(图 2b), 总体上 FNN 预报模型的命中率更高。具体分析可知, 三种机器学习模型对桂林站的命中率基本都较高; 对于梧州和玉林站, FNN 预报模型相比 MR 预报模型和 SVM 预报模

型,分别提高了 5.88%、10%和 11.7%、26.6%;三种方案对龙州站的命中率都偏低。

图 2c 为三种预报模型地面日极大风速预报 FAR 评分的对比,从总体上看,FNN 预报模型的空报率低于其它两个机器学习模型。具体分析可知,FNN 预报模型在桂林、龙州、南宁、玉林站的空报率比 MR 预报模型降低了 19.1%、42.8%、10.3%、68.4%;FNN 预报模型在桂林、梧州、龙州、玉林站的空报率比 SVM 预报模型降低了 14.5%、14.4%、25%、9.45%。

对预报偏差进行分析对比(图 2d),FNN 预报模型的预报偏差最小,SVM 报模型次之,MR 预报模型的预报偏差最大。FNN 预报模型比 MR 预报模型分别减少了 35%、100%、31.5%、12.8%(除梧州站外),SVM 预报模型比 MR 预报模型分别减少了 8%、27.2%、38.8%、37.5%(除梧州站外)。

以上分析结果表明,FNN 预报模型的预报性能优于 MR 预报模型和 SVM 预报模型,对广西台风期间地面日极大风速有较好的预报能力。

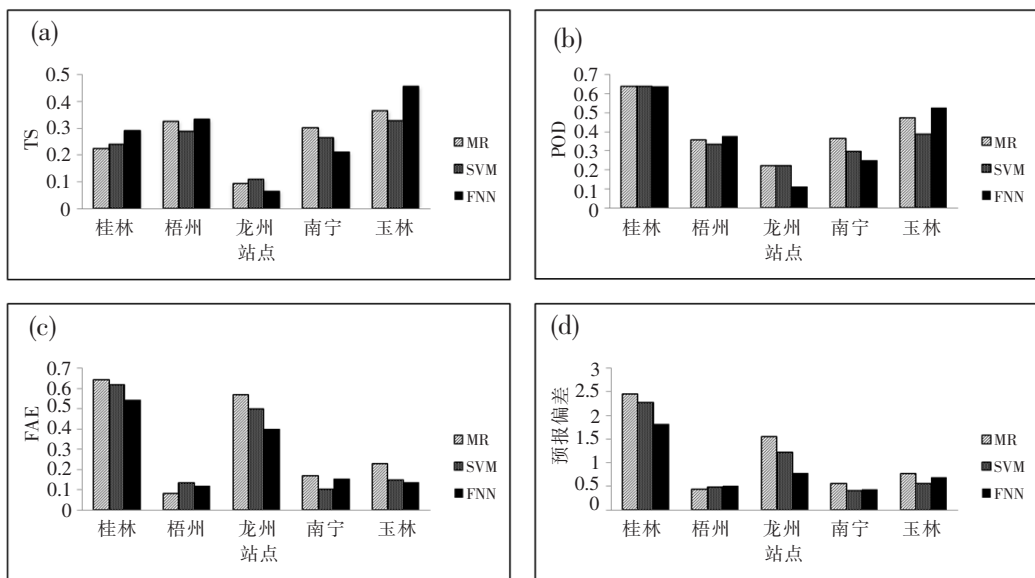


图 2 MR、SVM 和 FNN 预报模型对 2011—2020 年独立样本预报的地面日极大风 ( $>10.8\text{m}\cdot\text{s}^{-1}$ ) 的 TS 评分 (a)、命中率 (b)、空报率 (c) 和预报偏差 (d)

## 4 结论与讨论

采用多元线性回归、支持向量机、模糊神经网络三种不同的机器学习方法进行广西台风期间的 5 个气象观测站桂林、梧州、龙州、南宁、玉林的地面日极大风速进行预报建模试验,全部独立预报样本的绝对误差以及 6 级以上风速的 TS 评分、命中率、空报率和预报偏差的结果都表明,模糊神经网络预报模型的预测精度最高且相对稳定,多元线性回归方案次之,支持向量机在三种方案中预报效果最差。模糊神经网络对极大风速预报具有一定的适用性,可对广西台风期间地面日极大风速的预报有较好的参考作用,并可为后期进行广西的大风灾害预测研究的开展提供理论参考和实证基础。

另外,需要说明的是,本文只选取了广西 5 个具

有区域代表性的气象站点对预报模型的预报能力进行对比说明,后续可进一步将预报范围辐射至整个广西区域的所有站点进行预报试验。在试验中的某些指标上,我们发现模糊神经网络方法并非最优,其他机器学习方法也有很好的表现,可据此深入分析,考虑用将预报性能较好的机器学习算法进行有效集合,应该可以更好地提高风速的预报精度,为广西的大风灾害预报提供参考。

### 参考文献:

- [1] Knofczynski G T. Mundfrom D. Sample Sizes when Using Multiple Linear Regression for Prediction. [J]. Educational & Psychological Measurement, 2008, 68 (3): 431-442.
- [2] 修春波,任晓,李艳晴,等.基于卡尔曼滤波的风速序列短期预测方法[J].电工技术学报,2014,29(2):253-259.

- [3] Nayak M A, Ghosh S. Prediction of Extreme Rainfall Event Using Weather Pattern Recognition and Support Vector Machine Classifier[J]. Theoretical & Applied Climatology, 2013, 114(3-4):583-603.
- [4] Cevik H H, Cunkas M, Polat K. A New Multistage Short-term Wind Power Forecast Model using Decomposition and Artificial Intelligence Methods[J]. Physica A: Statistical Mechanics and its Applications, 2019(534): 122-177.
- [5] 勾海芝,赵征,夏子涵.基于经验模式分解的神经网络组合风速预测研究[J].电力科学与工程,2017,33(10):62-67.
- [6] 杨晓君,张楠,陈宏,等.基于人工神经网络算法的渤海海风预报方法研究[J].干旱气象,2019,37(1):146-152.
- [7] 朱智慧,黄宁立,王延凤.台风影响期间上海沿海极大风速预报方法研究[A].中国气象学会.创新驱动发展提高气象灾害防御能力—S2 灾害天气监测、分析与预报[C].中国气象学会,2013:8.
- [8] 钱燕珍,孙军波,余晖,等.用支持向量机方法做登陆热带气旋站点大风预报[J].气象,2012,38(3):300-306.
- [9] 肖志祥,姚才,赵金彪,等.广西热带气旋预报业务和研究进展[J].气象研究与应用,2020,41(4):20-27.
- [10] 金龙,黄颖,姚才,等.人工智能技术的热带气旋预报综述(之一)——BP神经网络和集成方法的热带气旋预报研究和业务应用[J].气象研究与应用,2020,41(2):1-6.
- [11] 郭彬,卓健,周冬静,等.基于智能计算的广西大风短临预报预警系统的产品检验[J].气象研究与应用,2021,42(1):80-84.
- [12] 陈润珍,孔宁谦.广西沿海热带气旋大风数值预报探讨[J].海洋预报,2004(1):52-55.
- [13] 董彦,林开平,黄小燕.南海热带气旋大风的遗传-神经网络集合预报[J].气象研究与应用,2014,35(1):40-45.
- [14] 李忠,刘景霞.基于遗传算法和最小二乘支持向量机的风电场超短期风速预测[J].电工技术,2021(13):56-59+66.
- [15] 钱斌凯,何彩芬,金炜,等.基于支持向量机的多因子风速预测[J].宁波大学学报(理工版),2018,31(3):14-19.
- [16] 黄小燕,史旭明,刘苏东,等.模糊神经网络方法在热带气旋强度预报中的应用研究[J].高原气象,2009,28(6):1408-1413.

## Application of machine learning in forecasting maximum wind speed of typhoon in Guangxi

Wu Yushuang<sup>1,2</sup>, Huang Xiaoyan<sup>3\*</sup>, Chen Jiazheng<sup>4</sup>, Zhao Huasheng<sup>3</sup>

(1. Guangxi Meteorological Observatory, Nanning Guangxi 530022, China;

2. Guangxi Meteorological Disaster Prevention Technology Center, Nanning Guangxi 530022, China;

3. Guangxi Institute of Meteorological Sciences, Nanning Guangxi 530022, China;

4. Guangxi University of Nationalities, Nanning Guangxi 530006, China)

**Abstract:** Taking the ground daily maximum wind speeds of five meteorological observation stations in Guilin, Wuzhou, Longzhou, Nanning and Yulin during the typhoon in Guangxi from 1980 to 2020 as the research object, three commonly used linear and nonlinear methods such as multiple linear regression (MR), support vector machine (SVM), and fuzzy neural network (FNN) were used for prediction modeling respectively, and the independent samples from 2011 to 2020 were tested. The results show that in terms of the average absolute error of full-sample wind speed prediction, the average absolute error of FNN model for Guilin, Wuzhou, Longzhou and Yulin is the smallest, and the overall prediction accuracy is the best; MR prediction model has better prediction ability for Nanning station, while the prediction effect of SVM model is generally poor. The statistical results of the four evaluation indexes of TS score, hit rate, false alarm rate and prediction deviation for strong winds above level 6 show that the prediction accuracy of FNN model is the highest and relatively stable, followed by MR, and SVM is the worst among the three schemes. FNN method has a good reference for the prediction of daily maximum wind speed on the ground during typhoon in Guangxi.

**Key words:** machine learning; typhoon; maximum wind; forecast modeling