

# 支持向量回归的机器学习方法在海浪预测中的应用

王燕, 钟建, 张志远

(中国人民解放军91001部队, 北京100161)

**摘要:** 基于支持向量回归(SVR)方法,建立了渤海海域近岸海浪有效波高短期预测模型,并设计了多组风浪信息组合输入方案,开展了有效波高预测敏感性试验。研究发现:综合考虑当前风浪信息作为模型的输入,对3 h和6 h有效波高预测具有较高的预报技巧,但随着预测时效的延长其预测准确性迅速降低;若此时引入未来预测风速信息作为模型输入,则可极大提高对12 h和24 h有效波高的预测能力;此外,若输入信息与预测对象之间不存在显著相关,多个信息的输入对有效波高预测效果提高无显著作用。建立的机器学习模型对小样本数据集具有良好的适应能力,能够有效解决海浪预报中的非线性问题,可为近岸海浪有效波高短期预测提供合理的技术参考。

**关键词:** 有效波高;风速;支持向量回归;粒子群优化

**中图分类号:** P731.22 **文献标识码:** A **文章编号:** 1003-0239(2020)03-0029-06

## 1 引言

海浪周期、波向、有效波高等海况特征是海洋工程建设、海上运输、环境保护和军事行动等活动必须考虑的重要安全因素<sup>[1]</sup>。其中,海浪有效波高最为重要,其预测的时效和准确性依然是目前亟需解决的关键问题。为提高海浪预测的准确性,国内外学者开展了不同的预测方法研究,主要有经验模型、数值模拟<sup>[2]</sup>和机器学习法等。基于经验模型,通过先验模型假设进行海浪预测,例如自回归移动平均方法(Auto-Regression and Moving Average Model, ARMA),但预测能力有限。基于数值模型,例如海浪模型(WAVE Model, WAM)<sup>[3]</sup>、近岸海浪数值模型(Sim-ulated WAVes Nearshore, SWAN)<sup>[4]</sup>、第三代海浪模型(WAVEWATCH III)<sup>[5]</sup>和海浪数值模式(MARine Science and NUMerical Modeling, WASNUM)<sup>[6]</sup>等,通过求解海浪谱控制方程来模拟和预测海浪要素,该方法对于广阔海域海浪具有显著的预测能力,但对于复杂地形的近岸海域,其预测能力欠佳。

由于机器学习方法不受高性能计算和大样本数据集输入的限制,近年来,人工神经网络<sup>[2,7-12]</sup>、模糊推理系统<sup>[12,13-14]</sup>和遗传规划<sup>[15-17]</sup>等机器学习方法被广泛应用于海浪要素预测,有效弥补了其他两种海浪预测方法的不足。支持向量机(Support Vector Machine, SVM)<sup>[18]</sup>作为一种新颖的机器学习方法,已广泛用于海上活动和沿海工程大气海洋环境要素的预测。Mohandes等<sup>[19]</sup>利用SVM方法进行风速预测。Asefa等<sup>[20]</sup>利用SVM方法实现了对复杂海况的海流预测。Salcedo-Sanz等<sup>[21]</sup>利用SVM方法开展了风电场风力涡轮机中的风速预测。Raghavendra等<sup>[22]</sup>详细阐述了SVM方法在水文领域的应用情况,并指出其对大气海洋环境的预测能力已得到社会广泛认可。2009年, Mahjoobi等<sup>[1]</sup>利用支持向量回归(Support Vector Regression, SVR)方法建立了海浪有效波高预测模型,将风速作为输入信息,并指出SVR方法对有效波高的预测技巧优于人工神经网络模型、多层感知机模型和径向基函数模型,且所需计算时间相对较少。

收稿日期: 2020-02-04; 修回日期: 2020-04-11。

基金项目: 国家重点研发计划(2017YFC1404005)。

作者简介: 王燕(1983-),女,工程师,硕士,主要从事数值模拟和海洋预报技术研究。E-mail: hhd\_gz@sina.com

通讯作者: 钟建(1982-),男,工程师,硕士,主要从事海洋环境军事保障研究。E-mail: zhongjian\_1982@126.com

海浪通常可分为风浪和涌浪。风浪是受局地风影响的海浪,当海浪传播到远离其产生区域时,或当它们的传播相速度超过局地风速时,则为涌浪<sup>[23]</sup>。由于涌浪的存在,如果在SVR预测模型中仅考虑风速作为输入信息,则无法达到最佳预测效果。针对这个问题,本文考虑历史海浪状态的影响,设计了多组风浪信息组合方案,利用SVR方法建立了渤海海域有效波高预测模型,并开展不同风浪信息组合对有效波高预测的敏感性试验,分析对比不同输入信息对有效波高预测技巧的影响。

## 2 预测区域和实验数据

渤海属于我国内陆海,海岸线崎岖,岛屿众多,地形、地貌复杂,对该海域近岸海浪的准确预测存在较大难度。随着我国经济的快速发展和军事实力的增强,对渤海海域海浪的预测需求日益迫切。本文选取渤海海域为预测区域(见图1),研究采用的观测数据为近岸浮标数据(121°40′48″E、38°9′31″N),观测要素为海浪有效波高和10 m风速,时间分辨率为1 h,时间区间为2012年11月15日—2013年2月15日。本文将观测数据集分为两组,第一组(810个样本)用于预测模型训练,第二组(480个样本)用于预测结果验证。

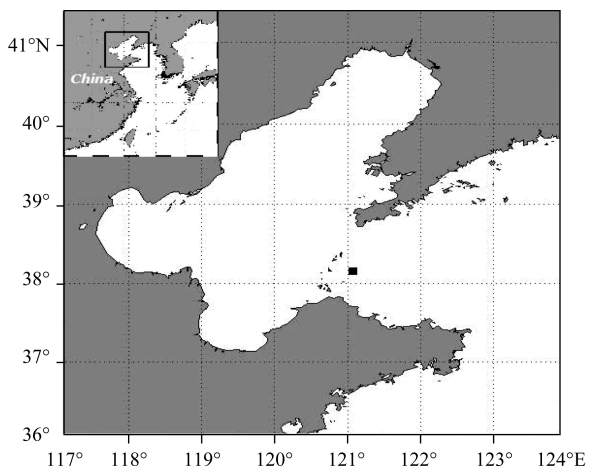


图1 预测区域和浮标位置(见黑色小框)

渤海湾属于半封闭海,太平洋涌浪很难影响到渤海湾海域,因此该海域主要为风浪。Pierson等指出风速 $U$ 和有效波高遵循如下经验关系<sup>[24]</sup>,

$$SWH \approx 0.22 \times \frac{U^2}{g} \quad (1)$$

式中: $g$ 是重力加速度, $U$ 是10 m风速。基于该风浪经验关系公式,得到风浪状态和涌浪状态之间拟合曲线(见图2),曲线下(上)方的值表示风浪(涌浪)为主的海浪状态。统计发现涌浪为主的海浪状态所占比例约为30%,表明该海域涌浪存在且不能忽视。

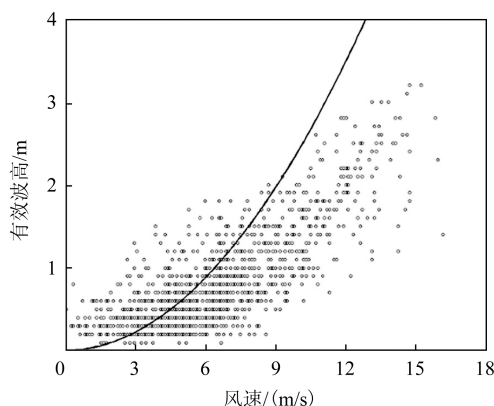


图2 海面风速和有效波高的散点分布图  
(风浪关系曲线根据式(1)叠加得到)

## 3 建模方法和参数优化

SVM是在统计学习理论上发展起来的基于结构风险最小化原则的机器学习理论<sup>[18]</sup>,其具有较高的泛化推广能力。通过将低维的样本空间映射到高维空间来实现低维非线性问题向高维线性问题的转化,能够较好地处理小样本、高维度、非线性等问题。研究发现,SVM方法在数据分类和数据回归方面具有显著优势,且其预测技巧优于人工神经网络模型和模糊推理模型<sup>[20]</sup>。

SVR方法是通过SVM方法进行回归分析的一种机器学习方法。为简化SVR模型复杂度并提升预测技巧,SVR模型通过构造核函数解决上述问题。核函数的选择能够很好地解决维数问题,并克服机器学习方法中存在的收敛速度慢、局部极值、过学习与欠学习等问题。常用的核函数有:线性函数、S函数、多项式函数和径向基函数等。本文通过敏感性数值试验评估后,选择径向基核函数,与其它核函数相比,径向基核函数所需输入参数少,数值计算限制少。具体表示为:

$$K(x, x_i) = \exp(-\gamma \|x - x_i\|^2) \quad (2)$$

式中: $x \in R^n$ ,  $x_i$ 是输入要素向量, $\gamma$ 是径向基核

函数参数。

径向基核函数的参数 $\gamma$ 和SVM模型中的惩罚参数 $C$ 的选择会极大的影响模型的预测结果,应根据不同的应用领域来取值。为了优化参数 $\gamma$ 和 $C$ ,本文使用粒子群优化(particle swarm optimization, PSO)算法选择最佳匹配参数。PSO方法最初由Kennedy等提出<sup>[25]</sup>,与遗传算法相比,PSO没有复杂的继承、变异、选择和交叉的操作,实施更为简单,通过交叉验证进一步确定PSO算法在本文可用。

#### 4 模型建立和性能评估指标

本文基于SVR机器学习方法,建立了渤海海域有效波高短期预测模型,并利用浮标观测数据进行3 h、6 h、12 h和24 h有效波高预测试验和结果验证。前人研究成果表明,与风向、风作用时间和风区长度等输入信息相比,风速对海浪预测技巧的提升更为重要<sup>[2]</sup>。此外,由于历史海浪状态能够改变海面粗糙度分布,因此本文在考虑风速影响的情况下,同时考虑海浪状态的影响,在预测模型中引入历史海浪状态作为输入信息。为分析不同的风浪输入信息组合对有效波高预测能力的影响,本文开展了多组有效波高预测的敏感性试验,共设计了32组数值试验(见表1),分别以M1—M32命名,表中符号 $H$ 和 $U$ 分别表示有效波高和10 m风速,下标表示时间,例如 $U_0$ 表示前当风速, $U_{-3}$ 表示预测前3 h风速,

$U_3$ 表示未来3 h风速, $H_0$ 表示当前有效波高, $H_3$ 表示未来3 h有效波高。

为定量评估SVR模型的预测能力,本文选择均方根误差(Root Mean Square Error, RMSE),相关系数( $R$ )和一致性指数(Index of Agreement, IA)作为评估指标,其定义如下:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2} \quad (3)$$

$$R = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (4)$$

$$IA = 1 - \frac{\sum_{i=1}^n (y_i - x_i)^2}{\sum_{i=1}^n (|y_i - \bar{y}| + |x_i - \bar{x}|)^2} \quad (5)$$

式中: $x_i(y_i)$ 是观测(预测)值, $n$ 表示样本总数。 $\bar{x}(\bar{y})$ 表示观测(预测)值的平均值。

#### 5 输入信息敏感性试验和结果分析

##### 5.1 仅输入10 m风速信息的预测试验

基于SVR模型,将当前风速 $U_0$ 作为输入信息,设计了5组数值试验(M1、M2、M3、M4、M5)来评估当前风速 $U_0$ 对有效波高的预测技巧。预测结果误差统计表明(见表2),在只考虑当前风速 $U_0$ 情况下,仅对未来6 h以内的有效波高具有较高的预测能力,尤其对未来3 h的有效波高预测技巧达到最高,其RMSE、 $R$ 、IA分别为0.277、0.913和0.935,但对6 h以上的有效波高没有显著的预测能力。

表1 不同输入信息组合的数值试验

输入参数	输出参数				
	$H_0$	$H_3$	$H_6$	$H_{12}$	$H_{24}$
$U_0$	M1	M2	M3	M4	M5
$H_0$	—	M6	M7	M8	M9
$H_0+U_{-6}$	—	M10	M11	—	—
$H_0+U_{-3}$	—	M12	M13	M14	—
$H_0+U_0$	—	M15	M16	M17	M18
$H_0+U_3$	—	M19	M20	M21	M22
$H_0+U_6$	—	—	M23	M24	—
$H_0+U_9$	—	—	—	M25	—
$H_0+U_{12}$	—	—	—	M26	—
$H_0+U_{18}$	—	—	—	—	M27
$H_0+U_{21}$	—	—	—	—	M28
$H_0+U_{24}$	—	—	—	—	M29
$H_i+U_i(i=-3,-2,-1,0)$	—	M30	M31	M32	—

表2 仅输入10 m风速信息的预测试验的有效波高预测误差

预报时间/h	RMSE/m	$R$	IA
0	0.356	0.825	0.888
3	0.277	0.913	0.935
6	0.369	0.820	0.872
12	0.569	0.446	0.607
24	0.640	0.100	0.291



## 5.2 仅输入当前海浪信息的预测试验

由于历史海浪状态能够影响海面粗糙度,本文设计了4组(M6、M7、M8、M9)数值试验,将当前有效波高 $H_0$ 作为SVR模型的输入信息,评估其对有效波高的预测技巧。从不同预测时刻的有效波高误差统计结果可见(见表3),未来3 h有效波高预测的 $RMSE$ 、 $R$ 、 $IA$ 分别为0.291、0.886、0.937,其具备一定预测技巧,但随着预测时效的延长,误差越来越大,12 h后完全失去预测能力。此外,比较表2和表3的误差结果,作为输入信息,当前风速 $U_0$ 对有效波高预测的贡献要高于当前有效波高 $H_0$ ,但当前有效波高 $H_0$ 的作用不可忽视,因此基于SVR模型的有效波高预测应当考虑当前有效波高 $H_0$ 的影响。

表3 仅输入当前海浪信息的预测试验的有效波高预测误差

预报时间/h	$RMSE/m$	$R$	$IA$
3	0.291	0.886	0.937
6	0.469	0.680	0.809
12	0.623	0.287	0.505
24	0.691	-0.076	0.271

## 5.3 不同风浪信息组合的预测试验

基于上述试验结果,本文将当前有效波高 $H_0$ 与不同时刻的风速 $U_i$ 作为SVR模型的输入信息,设计了23组数值试验(M10—M32),深入分析不同的风浪输入信息组合对有效波高预测技巧的影响。表4

列出了利用当前有效波高 $H_0$ 和不同时刻的风速 $U_i$ 作为输入信息( $H_0$ 和 $U_{-3}$ 、 $H_0$ 和 $U_0$ 、 $H_0$ 和 $U_3$ )的试验结果,可以看出将 $H_0$ 和 $U_3$ 作为输入信息对有效波高的预测技巧最高,其未来3 h有效波高预测的 $RMSE$ 、 $R$ 、 $IA$ 分别为0.179、0.963、0.977,对未来12 h有效波高仍然具备一定的预测能力,但预报误差变化特征表明,随着预测时效的延长预测误差也随之增加。此外,综合表2、表3和表4的误差统计结果可见,使用风浪组合作为输入信息的预测结果更接近实测有效波高。

为进一步提高对12 h和24 h有效波高的预测能力,本文考虑将预测时刻的风速作为SVR模型的输入信息,进而分别针对12 h和24 h有效波高预测目标独立训练其最优预测参数。根据输入的预测风速时效不同,将数值试验分为当前有效波高 $H_0$ 和预测时刻前6 h风速、当前有效波高 $H_0$ 和预测时刻前3 h风速、当前有效波高 $H_0$ 和预测时刻风速共3类数值试验。误差统计结果表明(见表5)。对12 h和24 h的有效波高预测技巧最高的是M25和M28,其中12 h有效波高预测的 $RMSE$ 、 $R$ 、 $IA$ 分别为0.295、0.892、0.929,也就是说采用当前有效波高 $H_0$ 和预测时刻前3 h风速作为输入信息,其12 h和24 h有效波高预报技巧达到最高。如表5所示,该方法对12 h和24 h有效波高预测误差基本相当,也就是说该方法对24 h有效波高预测结果可用。

为进一步评估多组输入信息对有效波高预测效果的影响,本文在上述数值试验的基础上,将8个

表4 利用当前有效波高 $H_0$ 和不同时刻的风速 $U_i$ 作为输入信息的预测试验的有效波高预测误差

预报时间/h	输入( $H_0+U_{-3}$ )			输入( $H_0+U_0$ )			输入( $H_0+U_3$ )		
	$RMSE/m$	$R$	$IA$	$RMSE/m$	$R$	$IA$	$RMSE/m$	$R$	$IA$
3	0.292	0.887	0.938	0.219	0.940	0.965	0.179	0.963	0.977
6	0.472	0.680	0.811	0.376	0.804	0.880	0.246	0.925	0.954
12	0.646	0.280	0.522	0.583	0.409	0.754	0.509	0.595	0.735
24	—	—	—	0.671	0.097	0.375	0.660	0.081	0.336

表5 输入当前有效波高 $H_0$ 和不同预测时刻的风速 $U_i$ 作为输入信息的预测试验的有效波高预测误差

预报时间/h	输入信息(当前有效波高 $H_0$ + 预报时刻前6 h风速 $U$ )			输入信息(当前有效波高 $H_0$ + 预报时刻前3 h风速 $U$ )			输入信息(当前有效波高 $H_0$ + 预报时刻风速)		
	$RMSE/m$	$R$	$IA$	$RMSE/m$	$R$	$IA$	$RMSE/m$	$R$	$IA$
12(M24/M25/M26)	0.402	0.770	0.855	0.295	0.892	0.929	0.339	0.849	0.902
24(M27/M28/M29)	0.390	0.808	0.863	0.287	0.910	0.933	0.382	0.800	0.866

变量  $H_i$  和  $U_i$  ( $i=-3, -2, -1$  和  $0$ ) 均作为 SVR 模型的输入信息, 设计了 3 组试验 (M30、M31、M32)。实验结果表明 (见表 6), 在认为所有输入信息都具有同等重要性的前提下, 同时引入多个输入信息的并不能有效提升预测效果, 反而增加了 SVR 模型的计算时间。

表 6 使用 8 个变量作为输入信息的误差

预报时间/h	RMSE/m	R	IA
3	0.217	0.939	0.966
6	0.378	0.801	0.879
12	0.592	0.399	0.588

## 6 结论

本文利用 PSO 算法, 建立了基于 SVR 的有效波高短期预测模型, 并开展了不同风浪输入信息组合对有效波高预测影响的敏感性试验。通过优化径向基核函数的参数  $\gamma$  和模型中的惩罚参数  $C$ , 以及采用最优化风浪组合方案作为模型输入信息, 开展了对渤海海域近岸有效波高 3 h、6 h、12 h 和 24 h 的预测可靠性研究。试验结果表明: 风速是影响有效波高预测最重要的因素, 利用当前风速信息能够准确实现对 3 h 和 6 h 有效波高的预测, 但随着预测时效增加, 12 h 和 24 h 预测准确性迅速降低; 当前有效波高仅对 3 h 有效波高预测具备一定预测技巧, 其影响作用不可忽视; 同时引入当前有效波高和不同时刻的风速作为输入信息可有效改善预测能力, 当预测时效延长至 12 h 后, 将预测风速作为模型的输入信息, 可有效提高 12 h 和 24 h 的有效波高预测准确性; 此外, 若模型输入信息与预测对象不存在显著的相关, 多个输入信息不能有效提高模型的预测能力。综上所述, SVR 模型对于小样本数据具有较高的学习能力和适应能力, 为复杂地形近岸海域有效波高短期预测提供了一种有效的机器学习方法, 可在海洋环境业务保障中推广应用。

## 参考文献:

- [1] Mahjoobi J, Mosabbe E A. Prediction of significant wave height using regressive support vector machines[J]. Ocean Engineering, 2009, 36(5): 339-347.
- [2] Mahjoobi J, Etemad-Shahidi A, Kazeminezhad M H. Hindcasting of wave parameters using different soft computing methods[J]. Applied Ocean Research, 2008, 30(1): 28-36.
- [3] The Wamdi Group. The WAM model—A third generation ocean wave prediction model[J]. Journal of Physical Oceanography, 1988, 18(12): 1775-1810.
- [4] Booij N, Ris R C, Holthuijsen L H. A third-generation wave model for coastal regions: 1. Model description and validation[J]. Journal of Geophysical Research, 1999, 104(C4): 7649-7666.
- [5] Tolman H L. User manual and system documentation of WAVEWATCH-III version 1.18[R]. Camp: NCEP, NOAA, 1999: 110.
- [6] 王关锁, 乔方利, 杨永增. 基于 MPI 的 LAGFD-WAM 海浪数值模式并行算法研究[J]. 海洋科学进展, 2007, 25(4): 401-407.
- [7] Deo M C, Jha A, Chaphekar A S, et al. Neural networks for wave forecasting[J]. Ocean Engineering, 2001, 28(7): 889-898.
- [8] Makarynskyy O. Improving wave predictions with artificial neural networks[J]. Ocean Engineering, 2004, 31(5-6): 709-724.
- [9] Makarynskyy O, Pires-Silva A A, Makarynska D, et al. Artificial neural networks in wave predictions at the west coast of Portugal [J]. Computers & Geosciences, 2005, 31(4): 415-424.
- [10] Zamani A, Solomatine D, Azimian A, et al. Learning from data for wind - wave forecasting[J]. Ocean Engineering, 2008, 35(10): 953-962.
- [11] Deka P C, Prahlada R. Discrete wavelet neural network approach in significant wave height forecasting for multistep lead time[J]. Ocean Engineering, 2012, 43: 32-42.
- [12] Castro A, Carballo R, Iglesias G, et al. Performance of artificial neural networks in nearshore wave power prediction[J]. Applied Soft Computing, 2014, 23: 194-201.
- [13] Özger M, Şen Z. Prediction of wave parameters by using fuzzy logic approach[J]. Ocean Engineering, 2007, 34(3-4): 460-469.
- [14] Akpınar A, Özger M, Kömürçü M İ. Prediction of wave parameters by using fuzzy inference system and the parametric models along the south coasts of the Black Sea[J]. Journal of Marine Science and Technology, 2014, 19: 1-14.
- [15] Gaur S, Deo M C. Real-time wave forecasting using genetic programming[J]. Ocean Engineering, 2008, 35(11-12): 1166-1172.
- [16] Kisi O, Shiri J. Precipitation forecasting using wavelet-genetic programming and wavelet-neuro-fuzzy conjunction models[J]. Water Resources Management, 2011, 25(13): 3135-3152.
- [17] Nitsure S P, Londhe S N, Khare K C. Wave forecasts using wind information and genetic programming[J]. Ocean Engineering, 2012, 54: 61-69.
- [18] Vapnik V N. The nature of statistical learning theory[M]. New York: Springer, 1995: 314.
- [19] Mohandes M A, Halawani T O, Rehman S, et al. Support vector machines for wind speed prediction[J]. Renewable Energy, 2004, 29(6): 939-947.
- [20] Asefa T, Kemblowski M, McKee M, et al. Multi-time scale stream flow predictions: the support vector machines approach[J].

- Journal of hydrology, 2006, 318(1-4): 7-16.
- [21] Salcedo-Sanz S, Ortiz-García E G, Pérez-Bellido Á M, et al. Short term wind speed prediction based on evolutionary support vector regression algorithms[J]. Expert Systems with Applications, 2011, 38(4): 4052-4057.
- [22] Raghavendra S, Deka P C. Support vector machine applications in the field of hydrology: a review[J]. Applied Soft Computing, 2014, 19: 372-386.
- [23] Semedo A, Vettor R, Breivik Ø, et al. The wind sea and swell waves climate in the Nordic seas[J]. Ocean Dynamics, 2015, 65 (2): 223-240.
- [24] Pierson Jr W J, Moskowitz L. A proposed spectral form for fully developed wind seas based on the similarity theory of S. A. Kitaigorodskii[J]. Journal of Geophysical Research, 1964, 69(24): 5181-5190.
- [25] Kennedy J, Eberhart R. Particle swarm optimization[C]// Proceedings of 1995 International Conference on Neural Networks. Perth, WA, Australia: IEEE, 1995: 1942-1948.

## Application of support vector regression in significant wave height forecasting

WANG Yan, ZHONG Jian, ZHANG Zhi-yuan

(Unit No.91001 of PLA, Beijing 100161 China)

**Abstract:** A short-term SWH forecasting model for the Bohai Sea is established using the support vector regression (SVR) in this paper, and sensitive experiment is conducted with multiple input combination of wind and wave data. It is found that the forecasting model reveals high forecasting skill for 3-hour and 6-hour prediction forced by the current wind and wave information, while the forecasting skill decreases dramatically along with the extension of the forecast leading time. Nevertheless, the forecasting skill for 12-hour and 24-hour prediction increases significantly under forcing of the future wind speed. Moreover, inputs of multiple information won't increase the forecasting accuracy if there is no significant correlation between the inputs and forecasting variables. In summary, the established SWH forecasting model reveals good adaptability for data of small samples and could effectively solve the nonlinear issues in wave forecasting, which provides technical reference for short-term forecasting of nearshore waves.

**Key words:** significant wave height; wind speed; support vector regression; particle swarm optimization