

# 基于PCA-BP特征工程的近海单点海温预报技术及应用

何恩业<sup>1</sup>, 李琼<sup>2\*</sup>, 张聿柏<sup>2</sup>, 匡晓迪<sup>1</sup>, 王源<sup>2</sup>, 朱现晔<sup>2</sup>

(1. 国家海洋环境预报中心 自然资源部海洋灾害预报技术重点实验室, 北京 100081; 2. 山东省海洋预报减灾中心, 山东 青岛 266104)

**摘 要:** 本文将主成分分析方法(Principal Components Analysis, PCA)和误差后传(Back Propagation, BP)神经网络相结合,建立了一种PCA-BP特征工程的近海单点海温智能预报模型,并对山东荣成近岸海域气象数值预报产品和在线海温监测仪连续观测数据开展了释用技术研究和应用。2021年业务化运行结果显示:该预报模型具有占用内存小、运行速度快、预报误差低的优点,相比近岸基础单元数值预报和经验预报的24 h均方根误差降幅达1.0℃和0.8℃,均方根相对误差降幅达12%~14%,未来48 h和72 h的预报误差也降幅明显,预报计算时间小于10 s,并将预报时效进一步向前扩展了3 d,达到144 h。

**关键词:** 海温预报;主成分分析;神经网络;特征工程;释用技术

**中图分类号:** P731.31 **文献标识码:** A **文章编号:** 1003-0239(2023)03-0035-10

## 0 引言

海表面温度(Sea Surface Temperature, SST, 简称海温)是地球海气能量平衡系统中的一个重要参数<sup>[1]</sup>,是海水在运动过程中热量和动力输送以及表面海气热量交换的体现<sup>[2]</sup>。太阳辐射、海气交换、冷暖洋流、降水和入海径流等是海温的主要影响因素<sup>[3]</sup>。海温对研究海气相互作用、水热量平衡、汛期状况、海洋大气环流以及预测海洋天气、气候、环境极端事件、风场、渔情、海洋锋和降水等有重要作用<sup>[1,4-7]</sup>,同时对于发展水产养殖、保护海洋环境、建设海洋水声工程、开发海洋生物矿产和滨海休闲旅游资源、科学规划岸上和离岸港口活动、开展海洋军事和海上交通运输等均有实际应用价值<sup>[8-14]</sup>,因此亟需准确的海温短期和中长期预报。

国外从20世纪40年代起便开始了与海温预报相关的分析预测<sup>[15]</sup>,国内从20世纪60年代初开展相

应研究<sup>[8]</sup>。随着观测技术的提升,海温预报方法快速发展,主流的海温预报方法可以分为经验预报、统计预报和数值预报3种。经验预报主要使用外推法和归纳法。前者是利用海温自身变动的持续性、保守性和周期性,根据当前的海温观测数据和未来天气系统的变化外推获得短期的海温预测值;后者则是预报员对大量的数据样本进行分类归纳,根据目前天气形势和分类型天气的相似性进行预报<sup>[16]</sup>。经验预报方法曾经长期作为我国沿海城市单点业务预报的主要手段<sup>[17]</sup>,但缺点也很明显,它比较依赖于预报员的个人能力,预报结果差别较大,而且效率较低,难以开展大范围精细化多点预报。统计预报是甄选出影响海温变化的相关因子,应用数理统计方法预报水温变化。从20世纪60年代后期开始,统计预报逐渐占据重要地位,70年代时MOS(Model Output Statistic)方法开始应用于预报业务中<sup>[8]</sup>,直到80年代我国海温预报仍以统计预报为

收稿日期: 2022-10-31。

**基金项目:** 国家自然科学基金面上项目(41976018);海岸灾害及防护教育部重点实验室(河海大学)开放基金(202202);南方海洋科学与工程广东省实验室(珠海)资助项目(SML2020SP008);自然资源部海洋环境信息保障技术重点实验室开放基金课题;自然资源部渤海生态预警与保护修复重点实验室开放基金课题(2022111)。

**作者简介:** 何恩业(1981-),男,高级工程师,硕士,主要从事海洋生态环境、生态灾害和水文气象预报研究。E-mail: heenye@163.com

**\*通信作者:** 李琼(1990-),女,助理研究员,硕士,主要从事海气相互作用和海洋预报研究。E-mail: li\_qiong@163.com

主<sup>[18]</sup>。按照是否考虑外界因子的影响,统计预报可分为单要素时间序列预报和多要素相关预报两类<sup>[19]</sup>;按照使用的统计方法,统计预报可分为回归分析<sup>[20]</sup>、多层向量自回归<sup>[21]</sup>、典型相关分析<sup>[22-23]</sup>、马尔科夫模型<sup>[24]</sup>、经验正交分析<sup>[25]</sup>、复经验正交函数<sup>[26]</sup>、方差分析<sup>[27]</sup>、非线性模糊系统推理<sup>[28]</sup>、一维相空间延拓<sup>[4]</sup>和人工智能方法<sup>[23]</sup>。近年来随着计算能力的提升,多层感知神经网络<sup>[29]</sup>、支持向量机<sup>[5]</sup>、模糊神经网络<sup>[30]</sup>、递归神经网络<sup>[6]</sup>和误差后传(Back Propagation, BP)神经网络<sup>[31-33]</sup>等人工智能方法逐渐占据重要地位。在数值预报方面,海洋动力模式从20世纪80年代开始被应用于海温预报,我国于1989年开始发布海温数值预报<sup>[14]</sup>。得益于近些年高性能计算、模式改进和同化技术的快速发展,海温数值预报在海洋模式分辨率和复杂性等多方面都有显著改善,并取代经验统计预报成为当前主要使用的方法。数值预报方法的优势在于物理意义清晰、预报范围广、精度高,但由于海水运动具有随机性,目前的研究水平在数值模型中仍无法给出完全精确的动力学描述<sup>[31]</sup>,尤其是在不确定的非线性因素多、环境状况复杂的近岸海域。神经网络方法能够在保持精度的前提下,逼近任意非线性函数,建立模型来模拟复杂的动力系统<sup>[34]</sup>。因此,将基于海洋动力学的数值预报和基于神经网络的统计预报相结合发展海温释用技术<sup>[7]</sup>并应用于近岸海域,以此来提高海温预报的准确性是迫切需要的。海温释用技术是对数值预报产品的再次利用,通过建立各种模型和方法对数值预报产品进行修正,提取其有用的预报信息,提高众多不同质量数值预报产品的应用能力<sup>[35]</sup>。研究表明,人工智能预报释用技术可对数值产品进行信息深度挖掘,已经成为数值预报的强力支撑,应用前景广阔<sup>[36-37]</sup>。但将人工神经网络方法单独使用于海温预报释用技术,容易因为变量输入过多产生“维度灾难”,且学习效率低。本文将利用主成分分析(Principal Component Analysis, PCA)和BP神经网络相结合的方法解决这一问题。

## 1 数据来源和方法

### 1.1 数据来源

海温观测资料来源于山东荣成近岸海域水温

在线监测仪器观测的2019年1月1日—2021年9月30日逐时SST数据。海温监测仪器为直读式温盐仪,安装于威海荣成北部海域的浮标上。该浮标于2018年8月布放,同步接入山东省海洋观测数据平台。浮标系统由标体、传感器、通讯系统、供电系统和锚链系统等部分组成。标体上加载的传感器间隔10 min传输水文气象观测数据,其中气象传感器(置于海面以上1.5 m处)可获取气温、气压、湿度、风向和风速等数据;海流传感器(置于海面以下2 m处)可获取表层流速和表层流向数据;波浪传感器(置于海面附近)可获取有效波高、波向和有效波周期等数据;温盐传感器(置于海面以下1 m处)可获取水温、盐度和电导率数据。气象预报数据为区域WRF (Weather Research and Forecasting)模式结果,分辨率为10 km,预报时效为7 d,逐3 h一次,该模式已在国家海洋环境预报中心(National Marine Environmental Forecasting Center, NMEFC)业务化运行。利用双线性插值将WRF气象要素数据插值到海温预报单点上,形成日变化气象要素数据集。选取2019年1月1日—2021年9月30日每日0—168 h的气象要素预报结果,包括海面2 m气温( $T$ )、比湿( $Q_2$ )、纬向10 m风速( $U$ )、经向10 m风速( $V$ )和海平面气压(Sea Level Pressure, SLP)共5个要素的日平均值进行分析。本文的数据质量控制采用 $3\sigma$ 准则对异常值进行处理,同时若某天的海温与前一日和后一日相差 $4\text{ }^{\circ}\text{C}$ 以上,则作剔除处理,并利用五点等权滑动平均滤波法进行缺失值填充<sup>[14,17,38]</sup>。

本文以2019年和2020年的SST和气象数值预报数据作为建模仿真调训集,利用调训好的模型对荣成近岸海域海温开展了2021年1—9月业务化试预报。采用迭代算法扩展预报时效,即下一时刻 $SST_{(t)}$ 的预报因子采用前一时刻的预报值 $SST_{(t-1)}$ 和气象因子的数值预报值,预报时效为144 h。

### 1.2 研究方法

#### 1.2.1 主成分分析法

PCA方法是对协方差矩阵进行特征分析,通过降维将多个指标转化为少数几个互不相关、含有元数据大部分信息的综合指标,即主成分。设原始数据矩阵 $X = \{x_{ij}\}_{n \times m}$ ,  $x_{ij}$ 表示样本 $i$ 的第 $j$ 个因子变量,计算过程为:

①数据标准化,即对所选因子变量做标准化无量纲处理。公式如下:

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}, \sigma_j^2 = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2, y_{ij} = (x_{ij} - \bar{x}_j) / \sigma_j \quad (1)$$

式中: $x_{ij}$ 为原始变量值; $\bar{x}_j$ 为第 $j$ 个因子变量的平均值; $\sigma_j$ 为标准偏差; $y_{ij}$ 为标准化值。由公式可得到标准化变量矩阵 $Y = \{y_{ij}\}_{n \times m}$ 。

②求解协方差矩阵 $R$ 。公式如下:

$$r_{ab} = (\sum (Y_a - \bar{Y}_a)(Y_b - \bar{Y}_b)) / \sqrt{\sum (Y_a - \bar{Y}_a)^2 \sum (Y_b - \bar{Y}_b)^2} \quad (2)$$

式中: $r_{ab}$ 为第 $a$ 个因子变量与第 $b$ 个因子变量的相关系数,值越大表明指标间相关程度越大,有必要对数据进行PCA处理。由公式可得矩阵 $R = \{r_{ab}\}_{m \times m}$ 。

③KMO(Kaiser-Meyer-Olkin)检验和Bartlett检验。对变量进行KMO和Bartlett检验是能否进行PCA处理的前提,一般当 $KMO > 0.6$ 且 $P < 0.05$ 才可进行。

④根据协方差矩阵 $R$ 求解主成分特征值、特征向量。解特征方程 $|\lambda E - R| = 0$ 求出特征值 $\lambda_k (k = 1, 2, \dots, m)$ ,因为 $R$ 是正定矩阵,所以其 $\lambda_k$ 都为正数,设 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq 0$ ,将其按从大到小排列, $\lambda_k$ 是各主成分的方差,其大小反映了各主成分的影响力。根据 $\lambda_k$ 求对应的特征向量 $e_k = (e_{k1}, e_{k2}, \dots, e_{km})$ ,采用加权求和方式便可得到第 $k$ 个主成分:

$$F_k = e_{k1}Y_1 + e_{k2}Y_2 + \dots + e_{km}Y_m (k = 1, 2, \dots, m) \quad (3)$$

⑤获取主成分个数。主成分 $F_k$ 贡献率为 $\lambda_k / \sum \lambda$  ( $k = 1, 2, \dots, m$ ),累计贡献率为 $(\sum_{k=1}^p \lambda_k) / \sum \lambda$  ( $p = 1, 2, \dots, m$ )。根据选取主成分的原则, $F_k$ 的特征值大于1,贡献率大于5%,累计贡献率要求达80%。设符合条件的 $F_k$ 特征值分别为 $\lambda_1, \lambda_2, \dots, \lambda_p (p \leq m)$ ,其中 $p$ 为主成分的个数。

⑥建立因子载荷矩阵。因子载荷量是 $F_k$ 与 $X_j$ 的相关系数 $r(F_k, X_j)$ ,揭示了主成分与因子间的相关密切程度,用以解释主成分的意义。

### 1.2.2 PCA-BP特征工程模型

本文将PCA和BP网络模型相结合,对元数据进行PCA降维处理,将获取的主成分变量作为BP模型的输入因子进行仿真调训和预报。PCA-BP特征工程设计见图1。该设计用于解决神经网络易出现的“维度灾难”问题,能够有效提高学习效率。利用NMEFC-WRF气象预报数据和水温在线监测仪器实时观测数据,预报未来0~144 h的单点海温变化。BP模型构建参考何恩业等<sup>[37]</sup>在叶绿素预测中使用的神经网络方法,神经元采用Sigmoid型激活函数 $f(x) = 1/(1 + e^{-x})$ ,损失函数(即网络总误差)定义为:

$$E_\Sigma = \frac{1}{2} \sum_{p=1}^n (t_i^p - y_i^p)^2 \quad (4)$$

式中: $t_i^p$ 为第 $p$ 个样本的期望输出,即海温实际观测值; $y_i^p$ 为网络模型输出值。将主成分变量输入BP网

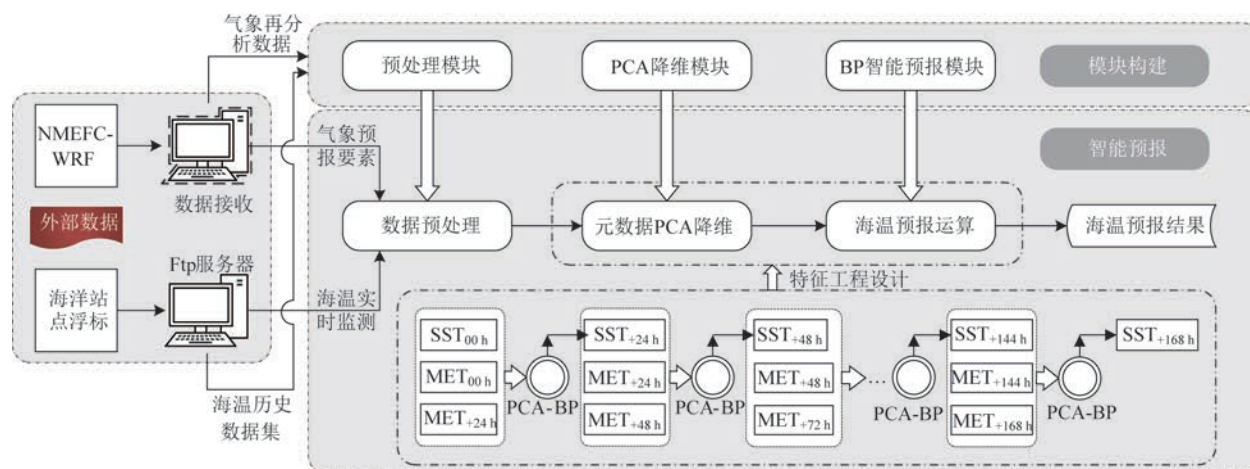


图1 近海单点海温PCA-BP特征工程预报模型构建流程

Fig.1 Construction of offshore SST prediction model based on PCA-BP feature engineering



络进行仿真训练,当 $E_{\Sigma}$ 小于预设值 $\varepsilon$ 或学习次数大于预设值 $q_{\max}$ 时,网络仿真训练结束。具体仿真算法流程见图2。

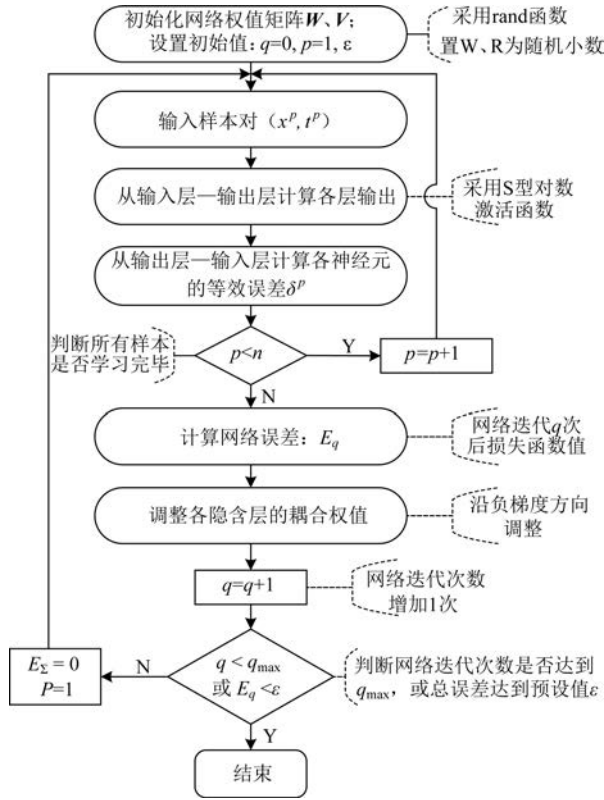


图2 BP仿真训练算法流程

Fig.2 Flowchart of BP learning algorithm

对2019年和2020年仿真调训样本数据采用交叉训练方案,每次预留100个样本作为测试集,其余用作拟合集循环试验,选取测试误差最小的拟合训练集和测试集组合,其对应的网络节点间的耦合权值及网络超参数即为最优网络,建立各要素与海温的非线性人工神经网络模型。

### 1.2.3 检验评估方法

采用均方根误差( $E_{\text{RMS}}$ )、均方根相对误差( $E_{\text{RMSP}}$ )、平均绝对误差( $E_{\text{MA}}$ )和平均相对误差( $E_{\text{MAP}}$ )对海温预测进行检验评估。计算公式分别为:

$$E_{\text{RMS}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (5)$$

$$E_{\text{RMSP}} = \sqrt{\frac{1}{n} \sum_{i=1}^n \left( \frac{\hat{y}_i - y_i}{y_i} \right)^2} \times 100\% \quad (6)$$

$$E_{\text{MA}} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (7)$$

$$E_{\text{MAP}} = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \times 100\% \quad (8)$$

式中: $\hat{y}_i$ 为预报值; $y_i$ 为实况值; $n$ 为检验样本数。

## 2 结果与分析

### 2.1 主成分分析结果

统计了气温( $T$ )、比湿( $Q2$ )、纬向风( $U$ )、经向风( $V$ )、海平面气压( $SLP$ )、SST这6个环境要素的16个时间序列变量: $T_{(00\text{h})}$ 、 $\Delta T_{(-24\sim 00\text{h})}$ 、 $\Delta T_{(-48\sim -24\text{h})}$ 、 $Q2_{(00\text{h})}$ 、 $\Delta Q2_{(-24\sim 00\text{h})}$ 、 $\Delta Q2_{(-48\sim -24\text{h})}$ 、 $U_{(00\text{h})}$ 、 $U_{(-24\text{h})}$ 、 $U_{(-48\text{h})}$ 、 $V_{(00\text{h})}$ 、 $V_{(-24\text{h})}$ 、 $V_{(-48\text{h})}$ 、 $SLP_{(00\text{h})}$ 、 $\Delta SLP_{(-72\sim 00\text{h})}$ 、 $SST_{(00\text{h})}$ 、 $SST_{(-24\text{h})}$ 。Person相关性分析结果显示(见表1), $SST_{(00\text{h})}$ 与 $T_{(00\text{h})}$ 、 $Q2_{(00\text{h})}$ 、 $SST_{(-24\text{h})}$ 分别达到了0.926、0.839、0.999的强相关水平,且均通过了 $P=0.01$ 水平的显著性检验,说明这3个因子变化过程对海温的影响较大,尤其是 $SST_{(00\text{h})}$ 与 $SST_{(-24\text{h})}$ 的相关系数接近1,体现了海温变化的保守性和连续性<sup>[15]</sup>,也正是传统海温预报多参考前1日海温实况数据的原因。此外, $SST_{(00\text{h})}$ 与 $T_{(00\text{h})}$ 、 $Q2_{(00\text{h})}$ 的强相关则因为在海温数值预报模式中,使用了以Galerkin法为基础,从三维海洋动力热力学过程出发构造的具有海洋自由面坐标下的距平方程组<sup>[39]</sup>。海温的距平预报方程包含海面净热通量距平 $Q_s$ ,它是由太阳辐射净通量、净长波辐射通量、海面感热通量和潜热通量组成的,海面气温是海面感热通量计算的重要一环,比湿是潜热通量的重要组成部分,因此海温受海气热量交换、海面蒸发和水汽输送的影响<sup>[2]</sup>,当前时刻的海温与气温和比湿有较强相关性。另外, $SST_{(00\text{h})}$ 与 $SLP_{(00\text{h})}$ 也达到了0.565的中等相关水平,而风的变化与海温也呈现出一定的相关性,且均通过了显著性检验,这说明气压升降过程和风场变化会对海温产生显著影响。对荣成近岸海域而言,在气压出现大幅升降和风场发生快速变化时往往有明显天气系统过境,一方面气流垂直运动和冷暖平流输送会造成局地气温的变化,从而影响海温,另一方面由于近海海流流向改变和局地出现降水,水体交换条件发生变化,从而对海温产生一定的影响。

表1 SST<sub>(00 h)</sub>与15个变量的相关性Tab.1 Correlation between SST<sub>(00 h)</sub> and environment

variables			
变量	相关系数	变量	相关系数
$T_{(00\text{ h})}$	0.926**	$U_{(-48\text{ h})}$	-0.281**
$\Delta T_{(-24\sim 00\text{ h})}$	-0.041	$V_{(00\text{ h})}$	0.174**
$\Delta T_{(-48\sim -24\text{ h})}$	-0.034	$V_{(-24\text{ h})}$	0.185**
$Q2_{(00\text{ h})}$	0.839**	$V_{(-48\text{ h})}$	0.194**
$\Delta Q2_{(-24\sim 00\text{ h})}$	-0.038	SLP <sub>(00 h)</sub>	-0.565**
$\Delta Q2_{(-48\sim -24\text{ h})}$	-0.033	$\Delta\text{SLP}_{(-72\sim 00\text{ h})}$	0.053
$U_{(00\text{ h})}$	-0.275**	SST <sub>(-24 h)</sub>	0.999**
$U_{(-24\text{ h})}$	-0.280**		

注:\*\*表示在0.01水平(双侧)上显著相关。

表2 公因子方差

Tab.2 The variance of variables

变量	初始值	提取后
$T_{(00\text{ h})}$	1	0.974
$Q2_{(00\text{ h})}$	1	0.934
$U_{(00\text{ h})}$	1	0.714
$U_{(-24\text{ h})}$	1	0.785
$U_{(-48\text{ h})}$	1	0.592
$V_{(00\text{ h})}$	1	0.812
$V_{(-24\text{ h})}$	1	0.745
$V_{(-48\text{ h})}$	1	0.864
SLP <sub>(00 h)</sub>	1	0.781
SST <sub>(-24 h)</sub>	1	0.890

对影响荣成近岸海域水温变化过程较大的 $T_{(00\text{ h})}$ 、 $Q2_{(00\text{ h})}$ 、 $U_{(00\text{ h})}$ 、 $U_{(-24\text{ h})}$ 、 $U_{(-48\text{ h})}$ 、 $V_{(00\text{ h})}$ 、 $V_{(-24\text{ h})}$ 、 $V_{(-48\text{ h})}$ 、SLP<sub>(00 h)</sub>、SST<sub>(-24 h)</sub>这10个气象水文因子变量进行主成分分析。经标准化后各变量KMO=0.736>0.6, Bartlett球度检验显著性 $P=0.00<0.05$ , 检验结果表明变量间相关性密切, 可以进行PCA分析。由表2可知,  $T_{(00\text{ h})}$ 、 $Q2_{(00\text{ h})}$ 、 $V_{(-48\text{ h})}$ 、 $V_{(00\text{ h})}$ 、SST<sub>(-24 h)</sub>的方差贡献率超过了0.8, 说明其对公因子贡献率较大; 所有因子的贡献率都超过了0.4, 说明10个因子均可被提取。由表3可知, 成分1、成分2、成分3、成分4的特征值分别为3.272、1.930、1.496、1.394, 贡献率分别为32.721%、19.302%、14.961%、13.938%, 这4个成分的累计贡献率为80.921%, 符合提取主成分的条件, 故可选择前4个成分作为主成分。

表4为根据特征值计算的各因子变量在主成分上的载荷及相应的特征向量。结果表明, 4个主成分包含了所有因子的信息, 且各个主成分主要包含的信息( $r>0.6$ )并不重合, 可以认为荣成海域海温的变化主要受到4类性质不同的主成分影响。F1与载荷较大的 $T_{(00\text{ h})}$ ( $r=0.93$ )、 $Q2_{(00\text{ h})}$ ( $r=0.88$ )、SST<sub>(-24 h)</sub>( $r=0.93$ )呈正相关, 与SLP<sub>(00 h)</sub>( $r=-0.71$ )呈负相关, 该成分主要反映了大气状态, 即气温升高、比湿增大、气压降低有利于海温的升高, 过去时刻海温较高则下一时刻海温也会相应较高; F2与载荷较大的 $U_{(00\text{ h})}$ ( $r=0.80$ )、 $U_{(-24\text{ h})}$ ( $r=0.86$ )、 $U_{(-48\text{ h})}$ ( $r=0.63$ )呈正相关, 主要反映了纬向风分量; F3和F4与载荷较大的 $V_{(00\text{ h})}$ ( $r=0.87$ )、 $V_{(-24\text{ h})}$ ( $r=0.63$ )、 $V_{(-48\text{ h})}$ ( $r=0.91$ )呈正比, 主要反映了经向风信息。

表3 主成分特征值及贡献率

Tab.3 Eigenvalues and cumulative variance contribution rate in principal component analysis

主成分	初始			旋转		
	特征值	贡献率/%	累计/%	特征值	贡献率/%	累计/%
1	4.443	44.427	44.427	3.272	32.721	32.721
2	1.651	16.511	60.938	1.930	19.302	52.022
3	1.149	11.486	72.424	1.496	14.961	66.983
4	0.850	8.496	80.921	1.394	13.938	80.921
5	0.709	7.087	88.008	—	—	—
			.....			
10	0.020	0.201	100.000	—	—	—

表 4 主成分的因子载荷矩阵和系数(特征向量)矩阵

Tab.4 Various factor loadings and eigenvectors in principal component analysis

环境因子	因子载荷矩阵				系数(特征向量)矩阵			
	F1	F2	F3	F4	e1	e2	e3	e4
$T_{(00\text{ h})}$	0.934	-0.199	0.147	0.202	0.543	0.163	0.062	0.114
$Q2_{(00\text{ h})}$	0.884	-0.268	0.202	0.199	-0.005	0.69	-0.079	-0.197
$U_{(00\text{ h})}$	-0.092	0.800	-0.045	-0.254	0.007	0.033	0.711	-0.161
$U_{(-24\text{ h})}$	-0.169	0.862	-0.096	0.065	0.554	0.124	0.063	0.138
$U_{(-48\text{ h})}$	-0.307	0.626	-0.042	0.323	-0.360	0.170	-0.038	0.062
$V_{(00\text{ h})}$	0.245	0.024	0.036	0.866	0.138	0.611	0.234	0.170
$V_{(-24\text{ h})}$	0.166	-0.056	0.627	0.567	0.168	0.053	0.621	0.298
$V_{(-48\text{ h})}$	0.155	-0.096	0.911	-0.035	0.199	-0.197	0.096	0.522
$SLP_{(00\text{ h})}$	-0.712	0.102	-0.443	-0.261	-0.426	-0.185	-0.161	-0.328
$SST_{(-24\text{ h})}$	0.932	-0.145	-0.015	-0.017	0.005	0.046	-0.052	0.633

注：这里的环境因子变量值均为为标准化后的值

## 2.2 荣成海域海温预报释用结果

### 2.2.1 模型仿真调训结果

仿真调训结果显示,网络最终仿真拟合的  $E_{\text{RMS}}$  和  $E_{\text{MA}}$  分别为 0.15 °C 和 0.11 °C, 100 个测试检验样本的  $E_{\text{RMS}}$  和  $E_{\text{MA}}$  分别为 0.17 °C 和 0.12 °C。利用训练好的网络模型对仿真调训样本集合进行 24~144 h 预报时效扩展预测误差检验,结果见表 5。72 h 内海温预测的  $E_{\text{RMS}}$  和  $E_{\text{MA}}$  均小于 0.3 °C, 72~144 h 预测的  $E_{\text{RMS}}$  和  $E_{\text{MA}}$  均小于 0.5 °C,说明模型能够较好地拟合并预测单点海温的变化情况。

表 5 2019 年和 2020 年荣成海域 0~144 h 海温预测误差统计(单位: °C)

Tab.5 SST prediction Errors in Rongcheng sea area from 2019 to 2020(unit: °C)

误差类别	时长					
	24 h	48 h	72 h	96 h	120 h	144 h
$E_{\text{RMS}}$	0.21	0.22	0.28	0.33	0.39	0.45
$E_{\text{MA}}$	0.14	0.20	0.26	0.31	0.36	0.41

### 2.2.2 业务化预报检验结果

PCA-BP、数值预报(NU)和经验预报(EXP)在冬春季的海温预报误差均显著小于夏秋季节(见图 3),但 PCA-BP 模型在夏秋季节预报中表现出较

大优势。李青青等<sup>[3]</sup>认为海面热通量的变化与温度变化有较好的同步性并呈正相关关系,中纬度沿海地区夏秋季气温日较差大于冬春季,SST 日变化幅度较大。我国属于季风气候区,夏季多雨带位于黄河流域及以北地区,夏季环流和降水对 SST 的影响呈显著相关<sup>[40]</sup>,而本文未考虑上述影响因子也是造成夏秋季预报偏差明显增大的原因之一。结果显示,海温预报存在一定的滞后性(见图 3a),个别日期预报误差较大(见图 3b),表明海温变化的保守性和连续性因素在模型预报中的权重占比较大,尤其在海温转向变化时,预报误差会陡然增大。因此后续还需考虑海流、降水等因子的影响,以平衡并抑制过去海温因子权重过高的状况,进一步提高预测准确率。

图 4 为该模型 2021 年 1—9 月在荣成近岸海域海温预报业务化试运行期间预报误差检验结果。相比数值预报和经验预报,PCA-BP 模型 24 h 预报的  $E_{\text{RMS}}$  为 0.31 °C,下降约 1.0 °C,  $E_{\text{RMSP}}$  为 2.7%,下降幅度达 12%~14%,  $E_{\text{MA}}$  为 0.19 °C,下降约 0.8 °C,  $E_{\text{MAP}}$  为 1.85%,下降幅度达 8%~9%;72 h 预报的  $E_{\text{RMS}}$  为 0.60 °C,下降约 0.8 °C,  $E_{\text{RMSP}}$  为 4.9%,下降幅度达 12%~14%,  $E_{\text{MA}}$  为 0.39 °C,下降约 0.6 °C,  $E_{\text{MAP}}$  为 3.62%,下降幅度达 7%~8%;144 h 预报的  $E_{\text{RMS}}$  为 0.78 °C,  $E_{\text{RMSP}}$  为 7.15%,  $E_{\text{MA}}$  为 0.57 °C,  $E_{\text{MAP}}$  为 5.43%。相比当前数值预报和经验预报只能实现未来 3~4 d

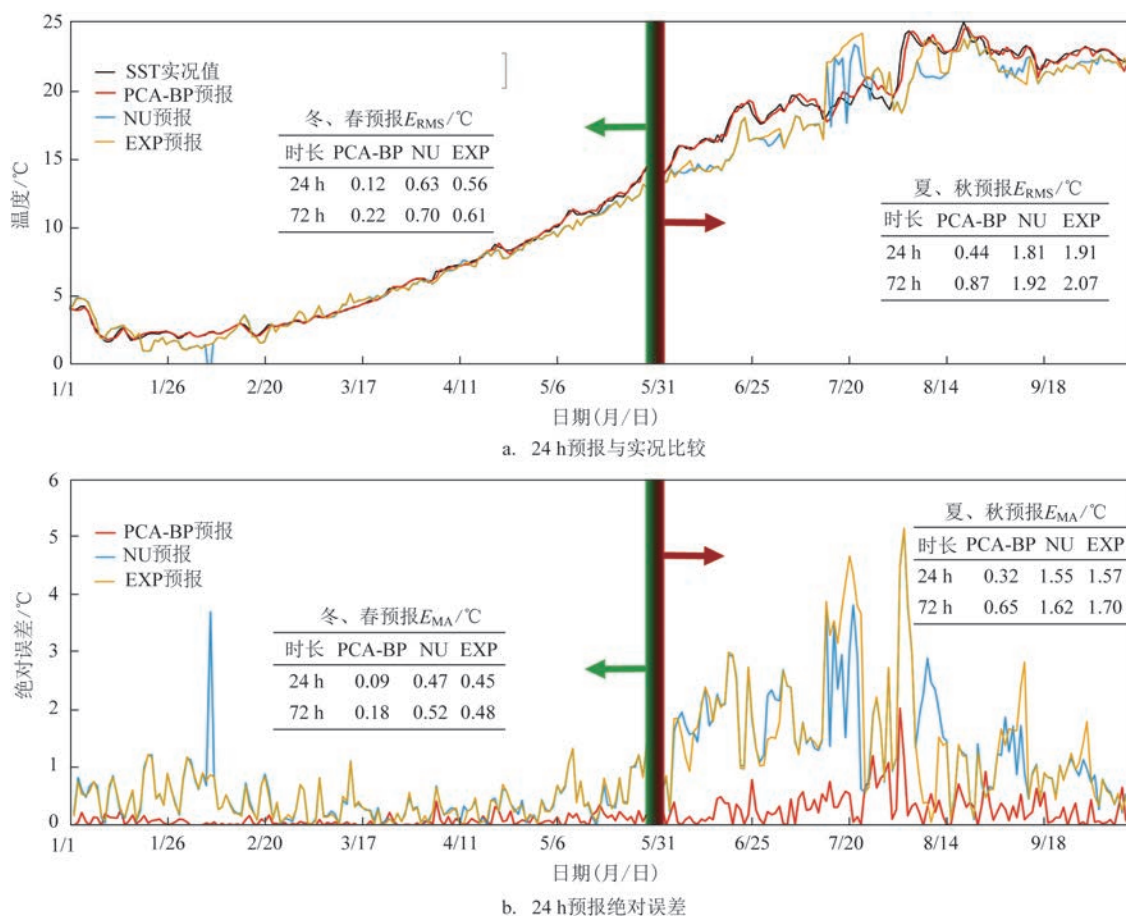


图3 智能预报(PCA-BP)、数值预报(NU)、经验预报(EXP)和实况值的海温比较

Fig.3 Comparison of intelligent forecast(PCA-BP), numerical forecast(NU), empirical forecast (EXP) and observation of SST

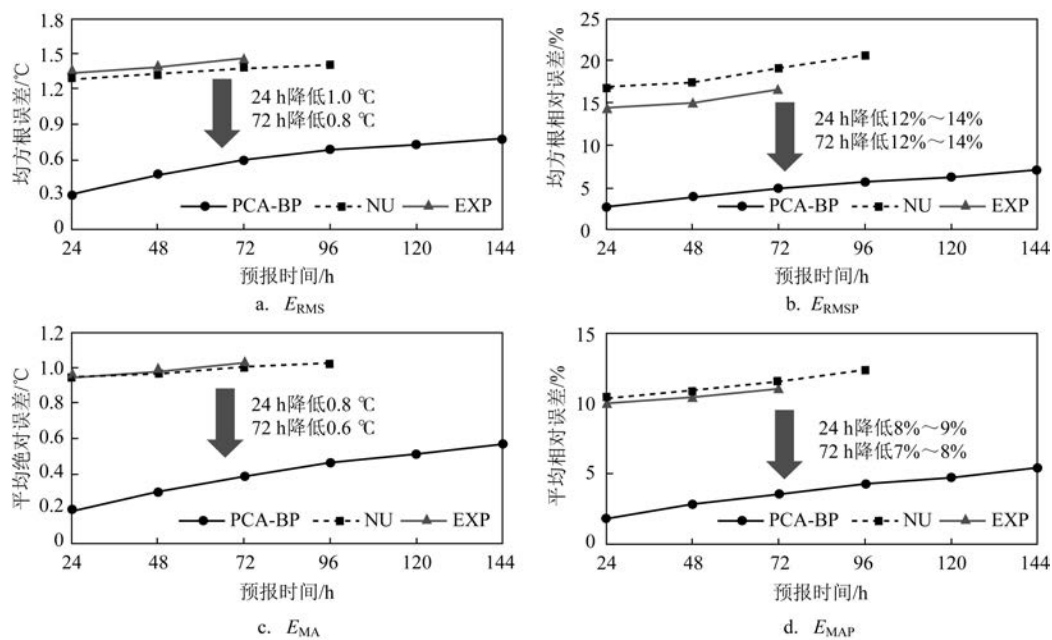


图4 2021年1—9月荣成海域PCA-BP海温预报模型业务化应用检验结果

Fig.4 Results of SST prediction based on PCA-BP model in Rongcheng sea area from January to September, 2021



的预报能力,PCA-BP模型将海温预报进一步向前扩展了3 d,且预报误差下降明显,释用结果表现良好。

### 3 结论

本文建立了一种基于PCA-BP特征工程的近岸海域单点水温预报模型,运用PCA方法对影响海温变化的相关性环境因子进行降维处理,可有效避免由于变量输入过多导致的网络“维度灾难”问题,并可提高仿真训练效率。对NMEFC-WRF气象预报数据和水温在线监测仪器的实时观测数据进行释用产品技术开发,可预报未来0~144 h的单点海温变化。PCA-BP模型具有预报精准度高和预报速度快的优点。业务化预报检验显示,该模型预报时效较之前的海温经验预报和数值预报进一步向前扩展了3 d,并具有占用内存小、运行速度快、预报误差低的优点。运算时间小于10 s,24 h预报的 $E_{\text{RMS}}$ 和 $E_{\text{MA}}$ 分别为0.31 °C和0.19 °C,72 h的为0.60 °C和0.39 °C,144 h的为0.78 °C和0.57 °C,相比近岸基础单元数值预报24 h和72 h预报的 $E_{\text{RMS}}$ 分别下降1.0 °C和0.8 °C, $E_{\text{MA}}$ 分别下降0.8 °C和0.6 °C。同时,该水温预报模型仅需每日的短期气象数值预报产品和过去24 h的在线海温观测数据,泛化效果良好,0~144 h海温预报的 $E_{\text{RMS}}$ 低于0.8 °C, $E_{\text{MA}}$ 低于0.6 °C。这说明本文构建的近岸海域水温智能预报模型可解决数理统计方法对长时间样本数据的需求问题和预报效果不稳定问题,并克服数值预报对近岸海域基础数据的高依赖性问题。但在实际应用中 also 发现,在海温变化转向时,预报误差会陡然增大,后续还需考虑海流、降水等因子的影响,以平衡抑制过去海温因子权重过高的状况,进一步提高预测准确率。总之,PCA-BP模型不需要完整的输入信息仍具有较高的预报精准度,可在沿海海洋预报机构日常业务中投入应用。

### 参考文献:

- [1] ZHANG Q, WANG H, DONG J Y, et al. Prediction of sea surface temperature using long short-term memory[J]. IEEE Geoscience and Remote Sensing Letters, 2017, 14(10): 1745-1749.
- [2] 国家海洋局海洋环境预报中心温盐组. 海温预报传真图的内容及其应用[J]. 海洋预报服务, 1984, 1(1): 85-89.
- Temperature and Salt Group of the Marine Environment Prediction Center of the National Oceanic Administration. Content and application of sea surface temperature forecast facsimile chart[J]. Marine Forecast Service, 1984, 1(1): 85-89.
- [3] 李青青, 何佩东, 蒋风芝, 等. 连云港、吕泗海域的海表温度变化及影响机制[J]. 科学技术与工程, 2013, 13(11): 2930-2937.
- LI Q Q, HE P D, JIANG F Z, et al. The Variation of sea surface temperature with its influencing mechanism of Lianyungang and Lüsi waters[J]. Science Technology and Engineering, 2013, 13(11): 2930-2937.
- [4] 邓冰, 张翔. 南海海温可预报性的诊断研究[J]. 海洋预报, 1997, 14(3): 32-37.
- DENG B, ZHANG X. Diagnosis study on the predictability of SST over South China Sea[J]. Marine Forecasts, 1997, 14(3): 32-37.
- [5] LINS I D, ARAUJO M, MOURA M D C, et al. Prediction of sea surface temperature in the tropical Atlantic by support vector machines[J]. Computational Statistics & Data Analysis, 2013, 61: 187-198.
- [6] ZHANG Y, ZHU J S, LI Z X, et al. Sea surface temperature predictions using a multi-ocean analysis ensemble scheme[J]. Climate Dynamics, 2017, 49(3): 1049-1059.
- [7] 张培军, 周水华, 梁昌霞. 基于卫星遥感海温数据的南海 SST 预报误差订正[J]. 热带海洋学报, 2020, 39(6): 57-65.
- ZHANG P J, ZHOU S H, LIANG C X. Study on the correction of SST prediction in South China Sea using remotely sensed SST[J]. Journal of Tropical Oceanography, 2020, 39(6): 57-65.
- [8] 张建华. 海温预报知识讲座 第一讲 海水温度预报概况[J]. 海洋预报, 2003, 20(4): 81-85.
- ZHANG J H. Lecture on sea temperature forecast: Lecture 1: overview of sea temperature forecast[J]. Marine Forecasts, 2003, 20(4): 81-85.
- [9] TONANI M, BALMASEDA M, BERTINO L, et al. Status and future of global and regional ocean prediction systems[J]. Journal of Operational Oceanography, 2015, 8(S2): S201-S220.
- [10] PATIL K, DEO M C, RAVICHANDRAN M. Prediction of sea surface temperature by combining numerical and neural techniques[J]. Journal of Atmospheric and Oceanic Technology, 2016, 33(8): 1715-1726.
- [11] 刘娜, 王辉, 凌铁军, 等. 全球业务化海洋预报进展与展望[J]. 地球科学进展, 2018, 33(2): 131-140.
- LIU N, WANG H, LING T J, et al. Review and prospect of global operational ocean forecasting[J]. Advances in Earth Science, 2018, 33(2): 131-140.
- [12] 郝日棚, 赵玉新, 何忠杰, 等. 基于 EOF-NAR 神经网络混合模型的海温预报方法研究[C]//中国海洋学会 2019 海洋学术(国际)双年会论文集. 三亚: 海洋出版社, 2019: 31-45.
- HAO R X, ZHAO Y X, HE Z J, et al. Research on SST forecasting method based on EOF-NAR neural network hybrid model[C]//2019 Oceanographic (International) Biennial Meeting



- of China Oceanographic Society. Sanya: Ocean Publishing House, 2019: 31-45.
- [13] 王金萍, 张翔宇. 宁德近海海温短期预报模型的初步研究[J]. 信息技术与信息化, 2019(2): 186-187.
- WANG J P, ZHANG X Y. Preliminary study on short-term prediction model of offshore sea temperature in Ningde[J]. Information Technology and Informatization, 2019(2): 186-187.
- [14] 王兆毅, 李云, 王旭. 中国近岸海域基础预报单元海温预报指导产品研制[J]. 海洋预报, 2020, 37(4): 59-65.
- WANG Z Y, LI Y, WANG X. Development of forecast guidance product for sea temperature of basic forecast units in the Chinese coastal waters[J]. Marine Forecasts, 2020, 37(4): 59-65.
- [15] 张志刚, 蔡东明, 骆永军. 周期外推方法在海水水温预报中的应用[J]. 海洋技术, 2007, 26(4): 69-71.
- ZHANG Z G, CAI D M, LUO Y J. Application of distinct periodicity extrapolation to sea water temperature forecast[J]. Journal of Ocean Technology, 2007, 26(4): 69-71.
- [16] 陶祖钰, 赵翠光, 陈敏. 谈谈统计预报的必要性[J]. 气象科技进展, 2016, 6(1): 6-13.
- TAO Z Y, ZHAO C G, CHEN M. The necessity of statistical forecasts[J]. Advances in Meteorological Science and Technology, 2016, 6(1): 6-13.
- [17] 李燕, 张建华, 刘钦政, 等. 单站海温短期预报自动化[J]. 海洋预报, 2007, 24(4): 33-41.
- LI Y, ZHANG J H, LIU Q Z, et al. The automation of Single Sea Station's surface sea temperature short term forecasting[J]. Marine Forecasts, 2007, 24(4): 33-41.
- [18] 杜碧兰, 宋学家, 张建华. 表层海水温度场的正交综合因子场分解预报方法——东海及其外缘海域月平均表层水温预报[J]. 海洋学报, 1982, 4(2): 149-156.
- DU B L, SONG X J, ZHANG J H. An empirical orthogonal multi-factorial method designed to predict sea surface temperature - on the prediction of mean monthly sea surface temperature for the East China Sea and the adjacent waters[J]. Acta Oceanologica Sinica, 1982, 4(2): 149-156.
- [19] 张建华. 海温预报知识讲座 第二讲 数理统计方法在海温预报中的应用[J]. 海洋预报, 2004, 21(1): 85-90.
- ZHANG J H. Lecture on SST forecast: lecture 2 application of mathematical statistics in SST forecast[J]. Marine Forecasts, 2004, 21(1): 85-90.
- [20] 侯瑞科. 平稳时间序列分析在海温预报中的应用[J]. 海洋预报, 1996, 13(1): 41-45.
- HOU R K. Application of stationary time series analysis in SST prediction[J]. Marine Forecasts, 1996, 13(1): 41-45.
- [21] LEE D E, CHAPMAN D, HENDERSON N, et al. Multilevel vector autoregressive prediction of sea surface temperature in the North Tropical Atlantic Ocean and the Caribbean Sea[J]. Climate Dynamics, 2016, 47(1-2): 95-106.
- [22] 周林, 杨成荫, 王汉杰, 等. 基于CCA-BP-BPNN释用模型的太平洋 SST 预报[J]. 解放军理工大学学报(自然科学版), 2009, 10(4): 391-396.
- ZHOU L, YANG C Y, WANG H J, et al. Interpretation scheme of SST prediction in the tropical Pacific Ocean based on CCA-BP-BPNN[J]. Journal of PLA University of Science and Technology (Natural Science Edition), 2009, 10(4): 391-396.
- [23] 方玥炜, 唐佑民, 李俊德, 等. 几种统计模型对热带印度洋海温异常的预报[J]. 海洋学研究, 2018, 36(1): 1-15.
- FANG Y W, TANG Y M, LI J D, et al. Several statistical models to predict tropical Indian Ocean sea surface temperature anomaly [J]. Journal of Marine Sciences, 2018, 36(1): 1-15.
- [24] XUE Y, LEETMAA A. Forecasts of tropical Pacific SST and sea level using a Markov model[J]. Geophysical Research Letters, 2000, 27(17): 2701-2704.
- [25] NEETU, SHARMA R, BASU S, et al. Data-adaptive prediction of sea-surface temperature in the Arabian sea[J]. IEEE Geoscience and Remote Sensing Letters, 2011, 8(1): 9-13.
- [26] 李繁华, 尹逊福, 曾宪模, 等. 复经验正交函数在水温预报中的应用[J]. 黄渤海海洋, 1996, 14(3): 2-7.
- LI F H, YIN X F, ZENG X M, et al. An application of complex empirical orthogonal function in the prediction of sea temperature [J]. Journal of Oceanography of Huanghai & Bohai Seas, 1996, 14(3): 2-7.
- [27] 侯瑞科. 方差分析及其在海洋水温预报中的应用[J]. 海洋预报, 1994, 11(4): 68-73.
- HOU R K. Analysis of variance and its application in ocean water temperature forecast[J]. Marine Forecasts, 1994, 11(4): 68-73.
- [28] 张韧, 王继光, 蒋国荣, 等. 非线性模糊识别及其在海温异常检测中的应用[J]. 地球科学进展, 2002, 17(4): 470-476.
- ZHANG R, WANG J G, JIANG G R, et al. Non-linear fuzzy recognition and its application in identifying SST abnormality[J]. Advances in Earth Sciences, 2002, 17(4): 470-476.
- [29] WU A M, HSIEH W W, TANG B Y. Neural network forecasts of the tropical Pacific sea surface temperatures[J]. Neural Networks, 2006, 19(2): 145-154.
- [30] 董兆俊, 滕军, 王骥鹏. 基于相空间重构与模糊神经网络耦合的海温预测模型[J]. 热带海洋学报, 2008, 27(4): 73-76.
- DONG Z J, TENG J, WANG J P. Application of phase space reconstruction and ANFIS model in SST forecasting[J]. Journal of Tropical Oceanography, 2008, 27(4): 73-76.
- [31] 张韧, 蒋国荣. 赤道东太平洋海温及 El Nino/La Nina 的反演及预测[J]. 解放军理工大学学报, 2000, 1(5): 7-12.
- ZHANG R, JIANG G R. Retrieve and prediction of equatorial east Pacific SST and El Nino / La Nina[J]. Journal of PLA University of Science and Technology, 2000, 1(5): 7-12.
- [32] 陈璇, 游小宝, 周广庆, 等. 基于权重调整的 BP 神经网络在 Nino 区海温预报中的应用[J]. 海洋预报, 2011, 28(5): 61-68.
- CHEN X, YOU X B, ZHOU G Q, et al. Application of BP neural network based on weight adjustment to the temperature forecast[J].

- Marine Forecasts, 2011, 28(5): 61-68.
- [33] LIU S Y, XU L Q, LI D L. Multi-scale prediction of water temperature using empirical mode decomposition with back-propagation neural networks[J]. Computers & Electrical Engineering, 2016, 49: 1-8.
- [34] 张韧. 非线性BP网络映射与赤道东太平洋海温预测[J]. 海洋通报, 2000, 19(4): 1-7.
- ZHANG R. Non-linear BP neural network mapping and prediction of equatorial east-Pacific SST[J]. Marine Science Bulletin, 2000, 19(4): 1-7.
- [35] 严明良. 数值产品释用方法在预报业务系统中的应用[C]//推进气象科技创新加快气象事业发展——中国气象学 2004 年年会论文集. 北京: 气象出版社, 2004.
- YAN M L, The Application of Numerical Product Interpretation Methods in Forecasting Operational Systems. Promoting Innovation in Meteorological Technology and Accelerating the Development of Meteorological Undertakings - Collection of Discussion Papers at the 2004 Annual Conference of China Meteorology. Beijing: China Meteorological Association, 2004
- [36] 韩玉康, 余丹丹, 申晓莹, 等. HYCOM 模式 SST 的预报误差订正[J]. 海洋预报, 2018, 35(3): 76-80.
- HAN Y K, YU D D, SHEN X Y, et al. Study on the correction of SST prediction of HYCOM[J]. Marine Forecasts, 2018, 35(3): 76-80.
- [37] 匡晓迪, 王兆毅, 张苗茵, 等. 基于 BP 神经网络方法的近岸数值海温预报释用技术[J]. 海洋与湖沼, 2016, 47(6): 1107-1115.
- KUANG X D, WANG Z Y, ZHANG M Y, et al. An interpretation scheme of numerical near-shore sea-water temperature forecast based on BPNN[J]. Oceanologia et Limnologia Sinica, 2016, 47(6): 1107-1115.
- [38] 何恩业, 杨静, 李尚鲁, 等. 基于双隐层 ANN 模型的叶绿素 a 浓度智能预报方法[J]. 海洋预报, 2021, 38(1): 43-53.
- HE E Y, YANG J, LI S L, et al. Intelligent prediction method for Chl-a based on the artificial neural network[J]. Marine Forecasts, 2021, 38(1): 43-53.
- [39] 张建华. 海温预报知识讲座——第三讲 海表温度长期数值预报[J]. 海洋预报, 2004, 21(3): 81-88.
- ZHANG J H. Lecture on SST Forecast: Lecture 3 Long term numerical prediction of sea surface temperature [J]. Marine Forecasts, 2004, 21(3): 81-88.
- [40] 谭桂容, 孙照渤, 赵振国. 我国东部夏季降水型与北半球大气环流和北太平洋海温的关系[J]. 南京气象学院学报, 1998, 21(1): 1-7.
- TAN G R, SUN Z B, ZHAO Z G. Relation of summer East-China rainfall patterns to northern circulations and SST[J]. Journal of Nanjing Institute of Meteorology, 1998, 21(1): 1-7.

## Technology and application of offshore SST prediction based on PCA-BP feature engineering

HE Enye<sup>1</sup>, LI Qiong<sup>2\*</sup>, ZHANG Yubai<sup>2</sup>, KUANG Xiaodi<sup>1</sup>, WANG Yuan<sup>2</sup>, ZHU Xianye<sup>2</sup>

(1. Key Laboratory of Marine Hazards Forecasting, National Marine Environmental Forecasting Center, Ministry of Natural Resources, Beijing 100081, China; 2. Shandong Marine Forecast and Hazard Mitigation Service, Qingdao 266104, China)

**Abstract:** An intelligent forecasting model of offshore sea surface temperature (SST) at single-point based on PCA-BP feature engineering has been established in this paper by combining principal components analysis (PCA) and back propagation (BP) neural networks. The model has been tested and implemented using meteorological numerical forecast products and continuous in-situ observation data of on-line SST monitor in Rongcheng coastal waters, Shandong Province. The operational results in 2021 show that this forecasting model has the advantages of less memory occupation, faster running speed and lower forecasting error. Compared with the numerical forecasting of offshore basic units and empirical forecasting, the 24-hour root mean square error of SST from the intelligent forecasting model decreases by 1.0 °C、0.8 °C, and the root mean square prediction error decreases by 12%~14%. The forecasting errors for 48 and 72 hours also significantly decrease. In addition, the forecasting calculation time is less than 10 s, and the forecasting time is further extended by 3 day to 144 h.

**Key words:** SST prediction; principal component analysis; neural network; feature engineering; interpretation technology