

于振龙, 许东峰, 姚志雄, 等. 基于多变量 LSTM 神经网络模型的 PDO 指数预测研究[J]. 海洋学报, 2022, 44(6): 58–67, doi:10.12284/hyxb2022047

Yu Zhenlong, Xu Dongfeng, Yao Zhixiong, et al. Research on PDO index prediction based on multivariate LSTM neural network model[J]. Haiyang Xuebao, 2022, 44(6): 58–67, doi:10.12284/hyxb2022047

基于多变量 LSTM 神经网络模型的 PDO 指数预测研究

于振龙¹, 许东峰^{1,2*}, 姚志雄^{1,2}, 杨成浩^{1,2}, 刘松楠¹

(1. 自然资源部第二海洋研究所 卫星海洋环境动力学国家重点实验室, 浙江 杭州 310012; 2. 浙江省海洋科学研究院, 浙江 杭州 310012;; 3. 自然资源部海洋空间资源管理技术重点实验室, 浙江 杭州 310012)

摘要: 利用 1921–2020 年的海平面气压、海平面高度、热含量数据以及海冰密集度作为太平洋年代际振荡 (Pacific Decadal Oscillation, PDO) 指数的预报要素, 建立了关于 PDO 指数时间序列预测的多变量长短期记忆 (Long Short Term Memory, LSTM) 神经网络模型, 对比分析了 2011–2020 年不同时间序列预测模型的 PDO 指数预测结果, 最后利用多变量 LSTM 神经网络模型实现了 2021–2030 年的 PDO 指数预测。结果显示, 多变量 LSTM 神经网络模型的预测值与观测值经过交叉验证后的平均相关系数和均方根误差分别为 0.70 和 0.62; PDO 未来 10 年将一直处于冷位相, PDO 神经网络指数出现两次波动, 于 2025 年出现最小值。相比于其他时间序列预测模型, 本文采用的多变量 LSTM 神经网络模型预测结果误差小、拟合效果好, 可以作为一种新型的预测 PDO 指数的手段。

关键词: PDO 指数; LSTM 神经网络模型; 时间序列预测

中图分类号: P732.6

文献标志码: A

文章编号: 0253-4193(2022)06-0058-10

1 引言

太平洋年代际振荡 (Pacific Decadal Oscillation, PDO) 是北太平洋年代际尺度上最主要的海表面温度 (Sea Surface Temperature, SST) 异常模态, PDO 的概念最早由 Mantua 等^[1] 于 1997 年首次提出, PDO 作为一种太平洋气候变化强信号, 对太平洋及其周边地区气候变化有着重要影响。在 PDO 暖相位期间, 冬季阿留申低压加强, 北太平洋呈现“马蹄形”的海表面温度异常, 北太平洋中部偏冷, 热带中东太平洋和北美沿岸偏暖; 而在 PDO 冷相位期间情况相反, 北太平洋中部偏暖, 热带中东太平洋和北美沿岸偏冷。PDO 指

数被定义为太平洋 20°N 以北海表面温度异常场经过经验正交函数分析 (Empirical Orthogonal Function Analysis, EOF) 第一模态的时间系数, 其正、负分别代表了 PDO 的暖、冷位相。PDO 作为主要的年代际气候变异之一, 对太平洋动力环境、渔业收获^[1-3] 和我国的降雨^[4-5]、气温^[6-7]、农作物产量^[8] 以及南海的盐度变化^[9-10] 等有重要影响, 所以 PDO 的预测就显得尤为重要。

由于 PDO 的复杂性以及全球变暖背景下 PDO 的可预测性正逐渐降低^[11], 预测 PDO 的相位变化对于人们来说仍具有很大的挑战性。前人做过很多关于 PDO 形成机理以及 PDO 指数重建的研究: Mantua 等^[1-2] 重点分析了 PDO 的时间、空间特征, 并定义 PDO

收稿日期: 2021-06-02; 修订日期: 2021-08-03。

基金项目: 大洋“十三五”项目 (DY135-E2-3-02); 浙江省财政一般公共预算: 浙江海平面上升影响分析 (330000210130313013006); 国家重点基础研究发展计划 (“973”计划) 项目 (2014CB441501)。

作者简介: 于振龙 (1997—), 男, 山东省滨州市人, 主要从事物理海洋方面的研究。E-mail: yzll127@yeah.net

* 通信作者: 许东峰 (1966—), 男, 福建省莆田市人, 研究员, 主要从事大洋环流和海气相互作用方面的研究。E-mail: xudongfengsio@sio.org.cn

年代际指数是由 PDO 指数 5 年滑动平均获得。Yasuda^[12]利用树木年轮和太平洋珊瑚数据对 PDO 指数进行了气候重建。MacDonald 和 Case^[13]利用加利福尼亚和阿尔伯塔的柔枝松的水文敏感年轮年表重建了公元 993–1996 年的 PDO 指数时间序列,同时评估了 PDO 强度和周期的长期变动特征。赵传湖等^[14]利用东亚地区高分辨率的气候代用指标,通过双重检验的逐步回归方法和小波分析等方法,重建了近 400 年来的北太平洋 PDO 指数,并分析了 PDO 周期的演变特征。结果表明,北太平洋 PDO 具有准 20 年和准 50~70 年的变化周期。Chen 等^[15]通过模型数据比较,探讨了千年尺度 PDO 变率,并利用北太平洋副热带海温的东西向差异以及东亚地区与 PDO 相关的降水异常重建了 PDO 指数。Schneider 和 Cornuelle^[16]指出, PDO 不是单一模式,而是多个物理过程的组合, PDO 的位相转变源于 Niño 3.4 的强迫作用以及阿留申低压和黑潮-亲潮延伸体海区变化的累积,并进一步估算了这些影响因子对 PDO 的相对贡献。与 Schneider 和 Cornuelle^[16]的结论一致,Newman 等^[17]分析了驱动 PDO 的因素,改进了以前的 PDO 经验模型^[18]。Huang 和 Wang^[19]利用海表面高度、海平面压力和海冰密集度,通过“年际增量法”将变量的年际增量作为预测对象,实现了对 PDO 指数的后报。Lu 等^[20]利用一种气候网络分析的新技术来捕获 PDO 相位转变的前兆。通过计算 PDO 空间格局冷区和暖区之间的联系,在相位转变前几年,发现冷暖区域之间的负相关逐渐增强。该信号从 1890 年到 2000 年捕获了 6 个 PDO 相位转变,提前预警时间平均为(6.5±2.3)年。

PDO 是多个物理过程综合作用的结果。其中,大气强迫超前于海表温度变率,是影响北太平洋温度变率的主要因素。阿留申低压通过影响北太平洋地区水汽输送^[21],间接影响北太平洋温度,是驱动 PDO 变率的主要大气强迫之一; PDO 相位的改变影响风场^[22]和温度场,进而带来海表面高度的变化^[23];热含量(Ocean Heat Content, OHC)是表征海洋热状态的一个重要参数^[24],海洋主要的热量存储于 0~700 m 水层^[25-26]。大西洋多年代际振荡(Atlantic Multi-Decadal Oscillation, AMO)可以通过大气遥相关影响太平洋副热带热量分布,进而影响 PDO^[27-28];北极涛动(Arctic Oscillation, AO)在 PDO 低频变化中起着重要的作用, AO 增强会导致阿留申低压增强,进而通过北太平洋海气相互作用影响 PDO^[29]。格陵兰海冰可以通过影响 AO 来影响 PDO 变率,海冰密集度可以用来表征海冰的变化^[30]。综上所述,海平面气压、海表面高

度、热含量和海冰密集度均可作为 PDO 指数的预报要素进行建模。

随着人工智能技术的发展,机器学习被应用于各个领域的研究^[31-33],神经网络成为了一种新型的研究手段。对于时间序列的预测应用最多的是循环神经网络(Recurrent Neural Network, RNN)模型,由于 RNN 模型结构比较简单,使用的反向传播(Back Propagation Through Time, BPTT)算法不能适应长时间序列数据的训练,容易产生梯度爆炸以及梯度消失的问题,无法实现长期记忆的效果。因此,需要一个含有存储单元、存储记忆功能的神经网络模型。1997 年长短期记忆(Long Short Term Memory, LSTM)神经网络模型由 Hochreiter 和 Schmidhuber^[34]提出。LSTM 神经网络模型是 RNN 模型的改进,通过增加记忆单元里的控制门限来解决 RNN 模型短期记忆的问题,从而可以高效地处理长时间序列信息,实现对复杂函数的建模。目前,国内外利用 LSTM 神经网络模型对 ENSO 预报的研究比较多^[35-37],与其他统计预报模型相比, LSTM 神经网络模型在长时间序列上具有较强的预报能力。由于年代际变化信号的预测比较困难,部分学者主要利用气候模型实现对 PDO 指数的后报试验^[38-39],利用深度学习的研究还比较少。本文旨在前人研究基础上,将海平面气压、海表面高度、海冰密集度作为 PDO 指数的预报要素,新增 0~700 m 水层的热含量作为预报因子,建立 LSTM 神经网络模型,实现对 PDO 指数的预测,为 PDO 指数的预测研究提供一种新的手段。

2 研究方法

2.1 LSTM 神经网络模型

LSTM 神经网络模型在本质上是 RNN 模型的改进版本,其与 RNN 模型最大的区别是增加了 3 个逻辑控制单元(即控制门,包括输入门、输出门和遗忘门)来解决 RNN 模型短期记忆问题,通过改变神经网络的记忆单元与其他部分连接的权值大小来控制信息流的输入、输出以及记忆单元的状态。其中输入门(input gate, 记为 i_t)用来控制信息是否能够被记录到记忆单元中;输出门(output gate, 记为 o_t)控制当前时间节点记忆单元中的信息是否能够流入到当前隐藏层中;遗忘门(forget gate, 记为 f_t)选择性遗忘上一时刻记忆单元的信息并累加到当前时刻记忆单元中。LSTM 神经网络模型结构示意图如图 1 所示,其中, X_t 为当前状态下数据的输入, h_t 为当前隐藏层, σ 和 \tanh 为当前记忆单元中两个不同的激活函数,“×”

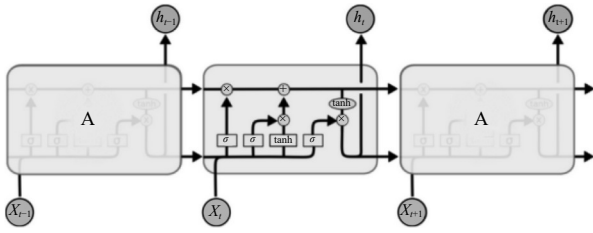


图 1 LSTM 神经网络模型结构示意图

(来源于文献 [34])

Fig. 1 Schematic diagram of LSTM neural network model structure (from reference [34])

和“+”分别代表矩阵相乘和矩阵相加运算。

在 LSTM 神经网络模型的训练过程中, 首先模型把 t 时刻的数据特征信息输入到输入层中, 经过激活函数得到输出结果。然后将输出结果、 $t-1$ 时刻的记忆单元存储信息和 $t-1$ 时刻隐藏层输出结果输入到当前 LSTM 神经网络模型结构处理单元中, 通过 3 个控制门限的处理后, 输出结果将进入下一输出层和隐藏层, 最后计算反向传播误差, 更新各个递归连接权重的大小。在 t 时刻 LSTM 神经网络模型每个记忆单元中的数据过程如式(1)~(6)所示:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \quad (1)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \quad (2)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \quad (3)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c), \quad (4)$$

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t, \quad (5)$$

$$h_t = o_t \times \tanh(C_t), \quad (6)$$

式中, f_t 、 i_t 、 o_t 为 3 个控制门限; W_f 、 W_i 、 W_o 分别代表了其对应的控制门限的递归连接权重; b_f 、 b_i 、 b_o 是通过对应的控制门限误差值; h_t 代表 t 时刻隐藏层输出值; C_t 代表当前神经元记忆状态的记忆单元, 具有随着训练不断存储、读取、重置和更新长时间信息的能力。控制门结构由激活函数与点乘组成, 公式中的激活函数 σ 在 $0 \sim 1$ 之间取值, \tanh 在 $-1 \sim 1$ 之间取值。点乘决定了记忆单元可以传送过去的信息量。激活函数的计算公式如式(7)和式(8)所示:

$$\sigma(x) = \frac{1}{1 + e^{-x}}, \quad (7)$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \quad (8)$$

2.2 精度评估

利用 PDO 预测值与观测值时间序列的均方根误差 (RMSE)、平均绝对误差 (MAE) 和平均绝对百分比误差 (MAPE) 评价模型预测值与观测值的精度大小, 计算公式为

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (9)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (10)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|, \quad (11)$$

式中, $\hat{y}_i = \{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n\}$ 为 PDO 指数模型预测值; $y_i = \{y_1, y_2, \dots, y_n\}$ 为 PDO 指数实际值。同时对预测值和实际值的时间序列数据进行了相关性分析, 计算皮尔逊相关系数 R , 计算公式为

$$R = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (12)$$

3 数据来源及处理

本文采用多个预报要素建立多变量 LSTM 神经网络模型预测 PDO 指数, 实验流程如图 2 所示。为了能够预测更长时间序列的 PDO 指数, 选取了 1921–2020 年 100 年的预报要素数据资料。除了海冰密集度数据以外, 其他预报要素都选取了相同的研究区域 ($20^\circ \sim 60^\circ \text{N}$, $120^\circ \text{E} \sim 100^\circ \text{W}$)。

首先, PDO 指数本文采用的是美国国家海洋和大气管理局 (National Oceanic and Atmospheric Administration, NOAA) 基于第五代版本 ERSST 扩展重建的 NCEI PDO 指数数据 (<https://www.ncdc.noaa.gov/teleconnections/pdo/>); 海平面气压 (Sea Level Pressure, SLP) 数据源自哥白尼气候变化服务 (Copernicus Climate Change Service, C3S, <https://cds.climate.copernicus.eu/>) 的 CMIP5 气候模式月平均数据, 该数据由 NOAA 的 GFDL-CM2p1 模型计算而来, 水平分辨率为 $2.5^\circ \times 1.5^\circ$; 海表面高度 (Sea Surface Height, SSH) 数据同样源自 C3S 的 CMIP5 模式数据, 水平分辨率为 $1^\circ \times 1^\circ$; 海冰密集度 (Sea Ice Concentration, SIC) 数据源自英国气象局哈德莱中心的全球海冰和海表面温度数据集 HadISST, 水平分辨率为 $1^\circ \times 1^\circ$ 。

热含量长时间序列数据获取相对困难, 本文利用热含量计算公式获取 1921–2020 年的热含量时间序列, 所采用的数据为哈德莱中心的 EN4 再分析温盐数据集。该数据集融合了 WOD9、GTSP、ASBO、GDACs 等数据集, 并进行了质量控制。水平分辨率为 $1^\circ \times 1^\circ$, 时间分辨率为 1 个月, 深度共 42 层, 最深达 5 000 m 以上。热含量计算方法为

$$\text{OHC} = \int_{-h}^0 S \rho C_p T dz, \quad (13)$$

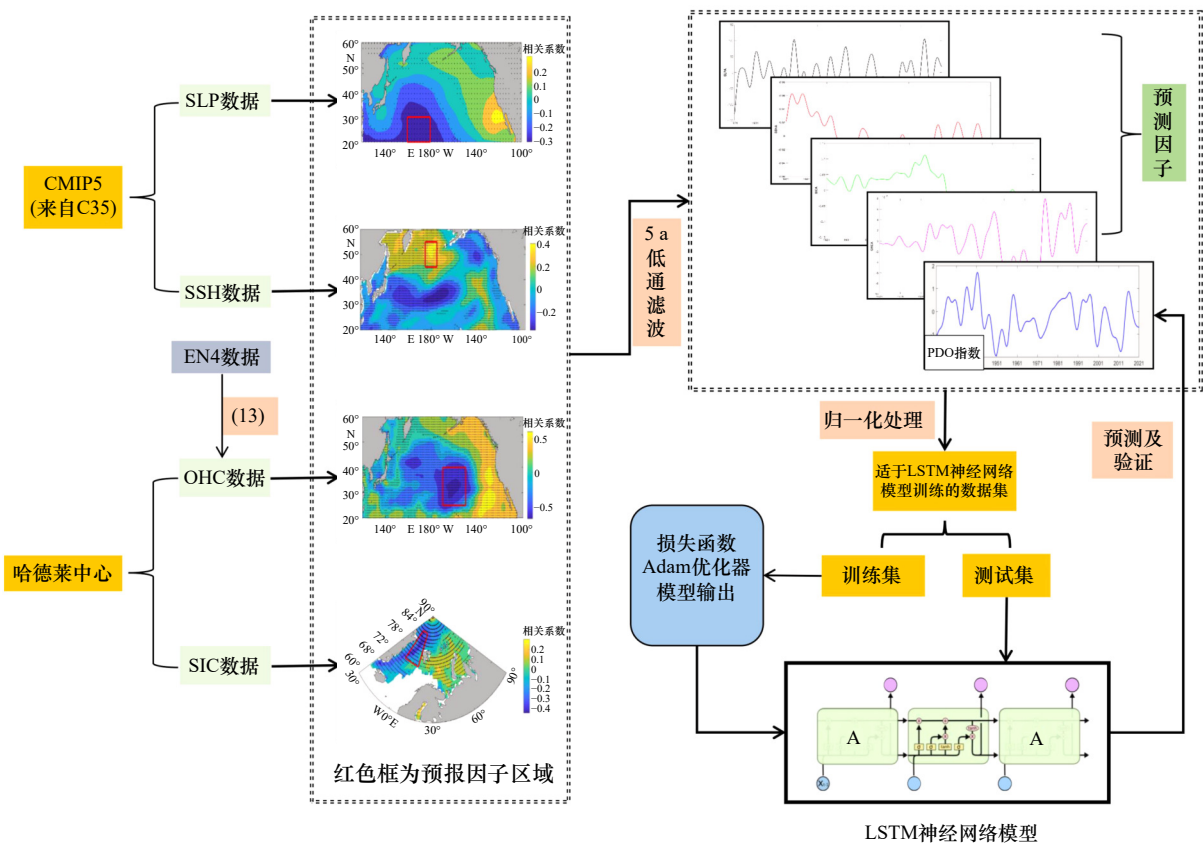


图2 实验流程图

Fig. 2 Experimental flowchart

式中, OHC 为热含量, 单位: J; S 为数据网格的面积大小; ρ 为海水的密度; C_p 为定压比热容; T 为海水温度; h 为热含量计算水深 (本文中 h 取为 700 m)。

其次, 采用相关性分析计算预报要素研究区域内的所有网格点时间序列与 PDO 指数的相关系数, 根据相关系数选取 4 个预报区域, 以下称“预报因子”, 其中, SLP 选取的是 $20^{\circ}\sim 30^{\circ}\text{N}$, $160^{\circ}\text{E}\sim 180^{\circ}$ 海域范围; SSH 选取 $45^{\circ}\sim 55^{\circ}\text{N}$, $175^{\circ}\text{E}\sim 175^{\circ}\text{W}$ 海域范围; $0\sim 700\text{ m}$ 的 OHC 选取 $25^{\circ}\sim 40^{\circ}\text{N}$, $150^{\circ}\sim 170^{\circ}\text{W}$ 海域范围; SIC 选取 $75^{\circ}\sim 85^{\circ}\text{N}$, $10^{\circ}\text{W}\sim 10^{\circ}\text{E}$ 海域范围。100 年的数据时间尺度会大大降低预报要素与 PDO 指数的相关性, 预报要素与 PDO 指数的相关系数空间分布及预报因子区域范围如图 3 所示, 预报因子在区域内每个数据格点的时间变化与 PDO 指数均满足 99% 置信水平的相关性。

最后, 计算预报因子的 1921–2020 年的时间序列 (异常), 由于 PDO 是年代际尺度上气候内部变率, 周期通常为 $20\sim 30$ 年, 本文将 PDO 指数数据利用 Butterworth 低通滤波器进行了 5 年的低通滤波处理, 去掉高频变化信号得到 PDO 指数, 4 个预报因子的数据处理与 PDO 指数相同, 这样不仅能够让 LSTM 神经

网络模型预测更长时间尺度的 PDO 指数序列, 同时可以提升模型训练时的训练效率。图 4 中从上到下依次是 PDO 指数、海平面气压异常、海平面高度异常、 $0\sim 700\text{ m}$ 热含量异常以及海冰密集度异常数据。

4 模型训练与调参

本文全部数据采用归一化处理以消除量纲和数量级带来的影响, 有利于提高预测模型的收敛速度和精度。多变量 LSTM 神经网络模型预测时, 预测的精度对模型参数的要求较大, 不同的参数组合类型对模型的预测精度影响较为明显。LSTM 神经网络模型的性能取决于模型中超参数的设置, 超参数是在模型训练前手动设定的训练变量, 包括激活函数、优化器、批大小、训练轮数和隐藏层节点数。为了提高模型的训练效率, 本文提前确定不重要的超参数: 模型设定训练轮数为 300 轮; 学习率设置为 0.005; 使用自适应矩估计 (Adaptive Moment Estimation, Adam) 优化器不断更新训练网络权重以使损失最小。由于数据量大且为复杂矩阵, 故调用 GPU 进行训练加速。为了防止 LSTM 神经网络模型出现过拟合的情况, 通过在全连接层后添加随机失活 (dropout) 层对在正相传递或

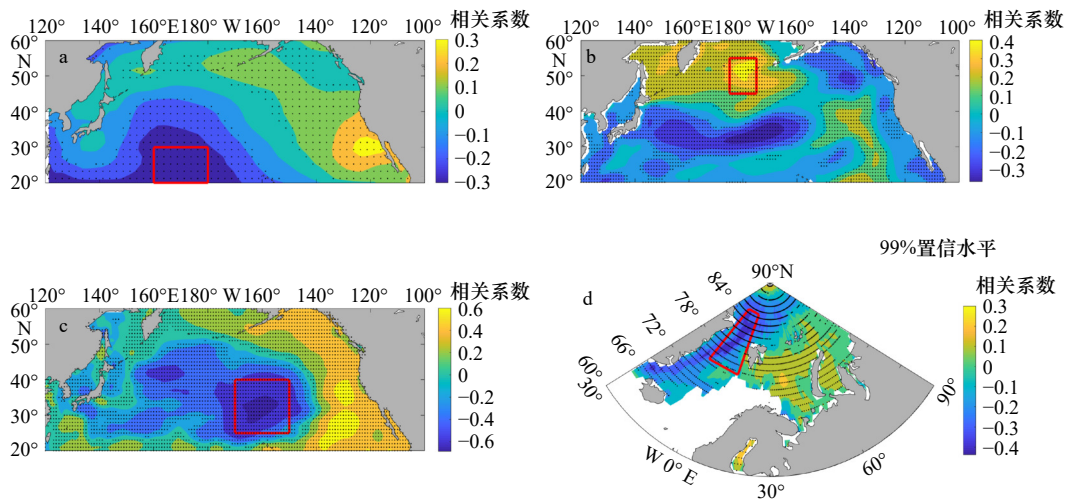


图 3 1921–2020 年预报要素与 PDO 指数的相关系数

Fig. 3 Correlation coefficients of forecast elements and PDO index from 1921 to 2020

a. 海平面气压; b. 海平面高度; c. 0~700 m 热含量; d. 海冰密集度。其中, 红色框为预报因子, 加点区域为超过 99% 的置信水平
 a. Sea level pressure; b. sea surface height; c. heat content of the 0–700 m layer; d. sea ice concentration. The red box is the predictor, and the dotted area is more than 99% confidence level

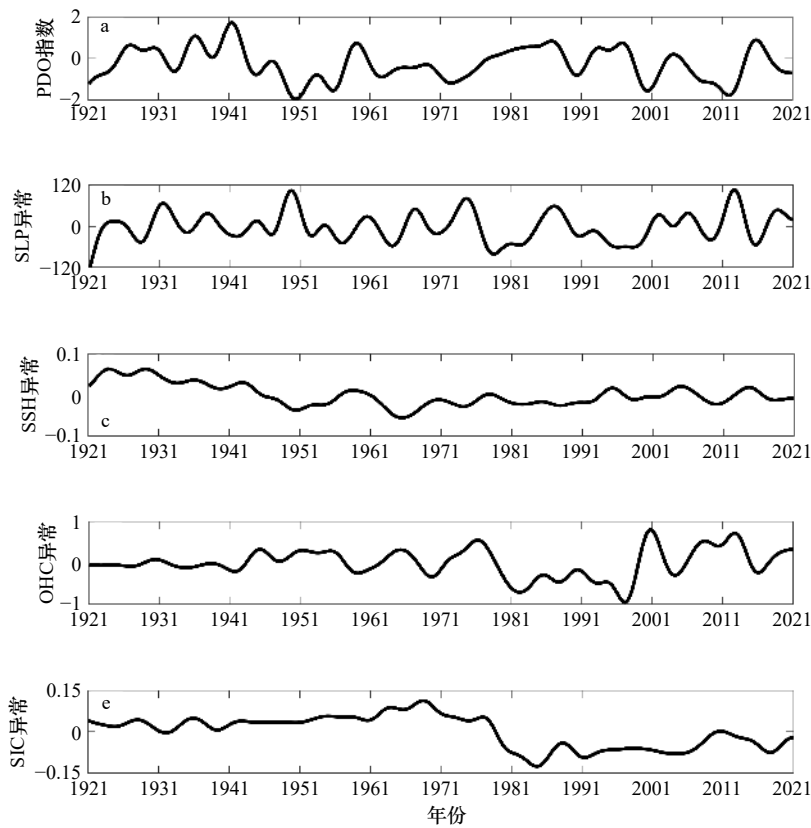


图 4 太平洋年代际振荡 (PDO) 指数以及预报因子时间序列

Fig. 4 Pacific decadal oscillation index and predictor time series

a. 太平洋年代际振荡 (PDO) 指数; b. 海平面气压 (SLP) 异常; c. 海平面高度 (SSH) 异常; d. 热含量 (OHC) 异常; e. 海冰密集度 (SIC) 异常
 a. Pacific decadal oscillation index; b. sea level pressure anomaly; c. sea surface height anomaly; d. ocean heat content anomaly; e. sea ice concentration anomaly

权值更新过程中的 LSTM 神经网络模型记忆单元进行优化, 随机失活率的大小设置为 0.2。

关于模型最优参数的确定, 本文对隐藏层节点数、

批大小以及学习率下降周期 3 个超参数进行了组合调参, 确保训练精度的同时也降低了模型训练的复杂程度。采用控制变量的方法, 将隐藏层节点数分别取 120

层、240层和360层,批大小取60、120和180,学习率下降周期取75、150和225轮,共27种参数组合。每种参数组合在模型建立时都需要通过交叉验证(Cross Validation, CV),交叉验证是一种流行的具备鲁棒性

的模型性能评估技术。本文建立的是关于PDO指数时间序列的多变量多步LSTM神经网络模型,训练集和测试集在逻辑上存在时间先后顺序,采用“递增时间窗交叉验证法”对模型进行交叉验证,方法如图5所示。

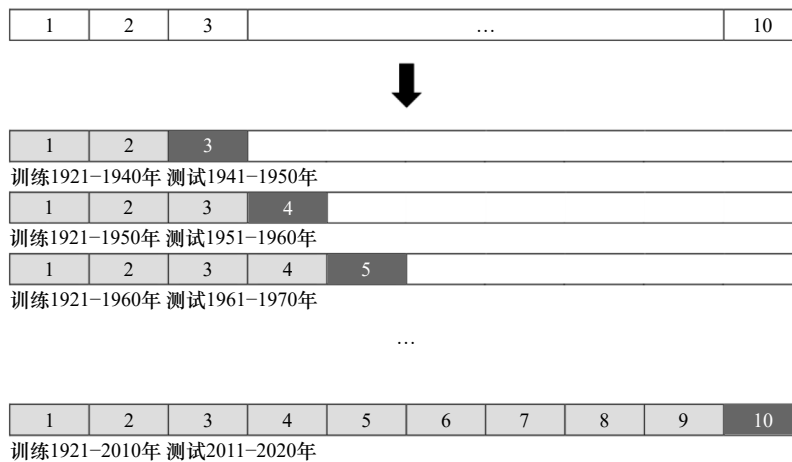


图5 递增时间窗交叉验证法

Fig. 5 Incremental time window cross-validation

该交叉验证方法将PDO指数数据等分成10份,考虑到PDO周期在20年以上,本文将最小训练集设置为20年。首先将前20年(1921-1940年)作为模型的训练集,后10年(1941-1950年)作为模型的测试集;将前30年(1921-1950年)作为模型的训练集,后10年(1951-1960年)作为模型的测试集;以此类推,直到将前90年(1921-2010年)作为模型的训练集,后10年(2011-2020年)作为模型的测试集。然后分别计算测试集中预测值与实际值的相关系数及均方根误差,模型训练5次后取平均结果。最终将交叉验证的结果取平均作为该参数组合的交叉验证结果,通过对比不同参数组合的模型平均相关系数以及均方根误差大小,确定模型的最优参数并利用最优参数建立LSTM神经网络模型,模型训练10次取平均作为最终预报结果,最终将1921-2020年的预报因子数据作为输入,实现对未来10年的PDO指数预测。

5 结果分析

多变量LSTM神经网络模型调参的过程实际上是模型寻找靶心的过程,经过交叉验证,根据测试集与训练集相关系数(R)以及均方根误差(RMSE)确定模型的最优参数,可以最大程度地保证模型的靶心处于最佳位置。模型调参结果如表1所示,拟合效果比较优异,所有参数组合测试集平均相关系数都在0.6~0.7之间。其中,隐藏层节点数为360,批大小为180,学习率下降周期为75轮的模型的平均相关系数

为0.6972,均方根误差为0.6221,选取该参数组合作为模型训练的最优参数,已在表1中粗体显示。

图6为根据最优参数建立的LSTM模型10次训练的平均拟合及预测结果。1921-2010年为模型训练集拟合结果,2011-2020年为模型测试集的拟合结果。黑色曲线代表NOAA提供的PDO指数观测值;红色曲线代表多变量LSTM神经网络模型训练结果(预测值)。PDO指数预测值与观测值的变化趋势基本一致,在峰值和谷值处的差距较大。2013-2019年的预测值比观测值偏小,其余时间的预测值比观测值偏大,总体来看,LSTM神经网络模型的拟合效果比较理想;蓝色曲线代表LSTM神经网络模型对未来10年(2021-2030年)的低频PDO指数预测结果。预测结果显示,PDO未来10年将一直处于冷位相,北太平洋中部出现SST正异常,热带中东太平洋和北美沿岸出现SST负异常,PDO指数出现两次波动,于2025年出现最小值。

如图7所示,本文对比了不同模型对2011-2020年PDO指数的预测结果,包括单一变量LSTM神经网络模型,统计回归模型以及多变量支持向量回归(Support Vector Regression, SVR)模型。其中,单变量LSTM神经网络模型是由PDO指数原始时间序列(来自NOAA)构建,经过与多变量LSTM神经网络模型相同的交叉验证方法确定模型最优参数组合,并训练10次取平均作为最终结果;统计回归模型采用多

表 1 交叉验证平均结果

Table 1 Cross-validation average result

隐藏层 节点数/层	批大小	学习率 下降周期/轮	测试集 相关系数	测试集 RMSE	
120	60	75	0.638 8	0.684 2	
		150	0.654 9	0.679 4	
		225	0.654 3	0.718 3	
	120	75	75	0.634 7	0.673 1
			150	0.630 9	0.680 0
			225	0.614 6	0.695 6
	180	75	75	0.661 8	0.648 8
			150	0.605 3	0.714 4
			225	0.624 1	0.686 3
60		75	75	0.650 1	0.653 4
			150	0.680 0	0.613 4
			225	0.637 4	0.649 9
120	75	75	0.656 7	0.633 0	
		150	0.626 4	0.660 0	
		225	0.659 3	0.618 3	
	180	75	75	0.683 3	0.615 0
			150	0.685 4	0.623 9
			225	0.653 1	0.633 4
360	60	75	0.684 5	0.614 5	
		150	0.674 9	0.624 1	
		225	0.677 7	0.666 4	
	120	75	75	0.684 7	0.615 8
			150	0.684 3	0.625 2
			225	0.659 2	0.656 5
	180	75	75	0.697 2	0.622 1
			150	0.690 1	0.635 4
			225	0.690 8	0.608 9

元回归,与多变量 LSTM 神经网络模型采用相同的预报因子作为变量;多变量 SVR 模型变量的选取与本文采用的多变量 LSTM 神经网络模型相同。结果显示,多变量 LSTM 神经网络模型拟合效果优于其他模型,特别是在峰值和谷值处更接近实际值。图 8 为不同模型的误差精度评估结果,多变量 LSTM 神经网络模型 RMSE=0.358 0, MAE=0.298 4, MAPE=0.406 9;用 PDO 指数原始时间序列构建的单变量 LSTM 神经网络模型 RMSE=0.447 2, MAE=0.342 4, MAPE=0.529 1;

统计回归模型 RMSE=0.466 1, MAE=0.409 9, MAPE=0.590 1;多变量 SVR 模型 RMSE=0.484 1, MAE=0.424 2, MAPE=0.629 9。从对比结果可以看出多变量 LSTM 神经网络模型预测效果最好,误差最小,与实际值拟合效果最好,其次是单变量 LSTM 神经网络模型、统计回归模型、多变量 SVR 模型。

6 结论

利用 1921–2020 年的海平面气压、海平面高度、热含量数据以及海冰密集度作为 PDO 指数的预报要素,通过空间相关性分析选取 PDO 指数预报因子,进一步将预报因子的时间序列数据进行 5 年低通滤波处理,通过交叉验证选取最优参数组合,建立关于 PDO 指数预测的多变量 LSTM 神经网络模型,对比分析不同模型 2011–2021 年的 PDO 指数预测结果,进而利用模型实现了对 2021–2030 年的 PDO 指数的预测。通过此次研究,本文得到如下结论。

(1) 基于多个预报因子数据建立的多变量 LSTM 神经网络模型可以有效地实现对 PDO 指数的预测,模型经过交叉验证后平均相关系数和均方根误差分别为 0.70 和 0.62,预测结果较为理想。

(2) 多变量 LSTM 神经网络模型预测效果要优于多变量 SVR 模型、统计回归模型以及单变量 LSTM 神经网络模型,其 PDO 指数预测结果误差小、拟合效果好,可以作为一种新型的预测 PDO 指数的手段。

(3) 本文预测了未来 10 年的低频 PDO 指数结果,结果显示, PDO 未来 10 年将一直处于冷位相,北太平洋中部将出现 SST 正异常,热带中东太平洋和北美沿岸出现 SST 负异常, PDO 指数出现两次波动,于 2025 年出现最小值。

PDO 作为年代际气候变化信号,本文建立的多变量的 LSTM 人工神经网络模型实现了 PDO 指数的位相预测。前人的研究^[19]虽然能够有效向后预报 1975–2009 年的 PDO 指数(相关系数为 0.81),但是其算法并不能实现对未来年代际时间尺度上 PDO 指数的预报。本研究优势在于采用机器学习方法,采用更大空间范围的预报因子,一定程度上克服了预测结果对 PDO 的依赖,实现了对未来 10 年 PDO 指数的预报。然而本研究的交叉验证试验结果与实测值相关系数在 0.6~0.7 之间,低于 Huang 和 Wang^[19]预测的相关性,产生这种结果的原因可能是其预测对象为 PDO 冬季(12 月、1 月、2 月)数据,而本研究选取所有月份的数据;其次,超参数的选取具有一定主观性,

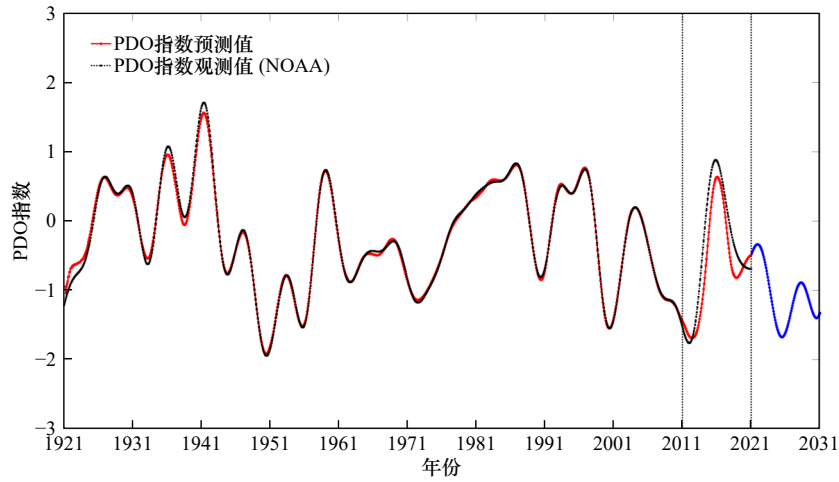


图6 LSTM网络模型预测值及观测值

Fig. 6 LSTM neural network model prediction results and true values

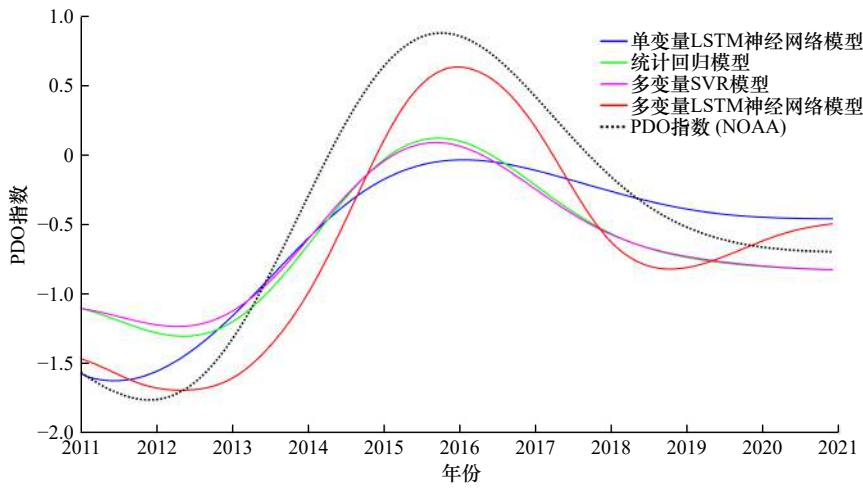


图7 2011-2020年不同模型的预测结果

Fig. 7 Forecast results of different models from 2011 to 2020

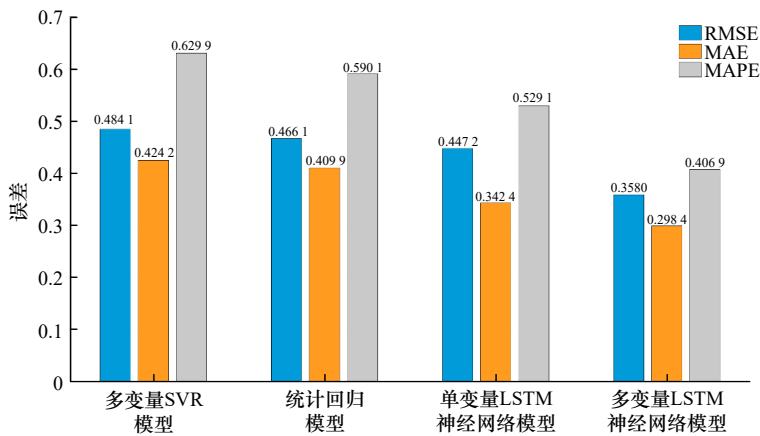


图8 不同模型预测的精度评估结果

Fig. 8 Accuracy evaluation results of different model predictions

在今后的研究中可以适当增加参数组合的数量来减小这种误差;最后,长时间尺度理想的预测结果要求

模型用更长时间的训练集,在更长时间尺度上的预测精度有待提高。

参考文献:

- [1] Mantua N J, Hare S R, Zhang Yuan, et al. A Pacific interdecadal climate oscillation with impacts on salmon production[J]. *Bulletin of the American Meteorological Society*, 1997, 78(6): 1069–1080.
- [2] Mantua N J, Hare S R. The Pacific decadal oscillation[J]. *Journal of Oceanography*, 2002, 58(1): 35–44.
- [3] Liu Dan, Tian Yongjun, Ma Shuyang, et al. Long-term variability of piscivorous fish in China Seas under climate change with implication for fisheries management[J]. *Frontiers in Marine Science*, 2021, 8: 581952.
- [4] Song Wanjiao, Wang Qiguang. CP El Niño and PDO variability affect summer precipitation over East China[J]. *Advances in Meteorology*, 2020: 2747194.
- [5] Bi Shuoben, Qu Ying, Bi Shengjie, et al. Multi-scale impacts of the Pacific SST and PDO on the summer precipitation of North-Central China from 1870 to 2002[J]. *Theoretical and Applied Climatology*, 2018, 132(3): 953–963.
- [6] Zhang Guwei, Zeng Gang, Li Chun, et al. Impact of PDO and AMO on interdecadal variability in extreme high temperatures in North China over the most recent 40-year period[J]. *Climate Dynamics*, 2020, 54(5): 3003–3020.
- [7] Xue Xu, Chen Wen, Chen Shangfeng, et al. PDO modulation of the ENSO impact on the summer South Asian high[J]. *Climate Dynamics*, 2018, 50(3/4): 1393–1411.
- [8] Xu Yifei, Li Te, Shen Shuanghe, et al. Effects of cyclic variability in Pacific decadal oscillation on winter wheat production in China[J]. *International Journal of Climatology*, 2021, 41(4): 2239–2252.
- [9] Nan Feng, Yu Fei, Xue Huijie, et al. Freshening of the upper ocean in the South China Sea since the early 1990s[J]. *Deep-Sea Research Part I: Oceanographic Research Papers*, 2016, 118: 20–29.
- [10] Zeng Lili, Wang Dongxiao, Xiu Peng, et al. Decadal variation and trends in subsurface salinity from 1960 to 2012 in the northern South China Sea[J]. *Geophysical Research Letters*, 2016, 43(23): 12181–12189.
- [11] Li Shujun, Wu Lixin, Yang Yun, et al. The Pacific Decadal Oscillation less predictable under greenhouse warming[J]. *Nature Climate Change*, 2020, 10(1): 30–34.
- [12] Yasuda I. The 18.6-year period moon-tidal cycle in Pacific Decadal Oscillation reconstructed from tree-rings in western North America[J]. *Geophysical Research Letters*, 2009, 36(5): L05605.
- [13] MacDonald G M, Case R A. Variations in the Pacific Decadal Oscillation over the past millennium[J]. *Geophysical Research Letters*, 2005, 32(8): L08703.
- [14] 赵传湖, 弓泓, 李猛. 近400年来北太平洋年代际振荡变化[J]. *中国科技论文*, 2015, 10(21): 2528–2533.
Zhao Chuanhu, Gong Hong, Li Meng. Variations of Pacific decadal oscillation over the past 400 years[J]. *China Sciencepaper*, 2015, 10(21): 2528–2533.
- [15] Chen Chunzhu, Zhao Wenwei, Zhang Xiaojian. Pacific Decadal Oscillation-like variability at a millennial timescale during the Holocene[J]. *Global and Planetary Change*, 2021, 199: 103448.
- [16] Schneider N, Cornuelle B D. The forcing of the Pacific decadal oscillation[J]. *Journal of Climate*, 2005, 18(21): 4355–4373.
- [17] Newman M, Alexander M A, Ault T R, et al. The Pacific decadal oscillation, revisited[J]. *Journal of Climate*, 2016, 29(12): 4399–4427.
- [18] Newman M. Interannual to decadal predictability of tropical and North Pacific sea surface temperatures[J]. *Journal of Climate*, 2007, 20(11): 2333–2356.
- [19] Huang Yanyan, Wang Huijun. A possible approach for decadal prediction of the PDO[J]. *Journal of Meteorological Research*, 2020, 34(1): 63–72.
- [20] Lu Zhenghui, Yuan Naiming, Yang Qing, et al. Early warning of the Pacific Decadal Oscillation phase transition using complex network analysis[J]. *Geophysical Research Letters*, 2021, 48(7): e2020GL091674.
- [21] Xiong Yating, Chen Qiuyu, Ren Xuejuan. Influence of boreal winter intraseasonal variation of Aleutian Low on water vapor transport and Atmospheric rivers[J]. *Atmosphere*, 2019, 10(2): 49.
- [22] Zeng Zhenzhong, Ziegler A D, Searchinger T, et al. A reversal in global terrestrial stilling and its implications for wind energy production[J]. *Nature Climate Change*, 2019, 9(12): 979–985.
- [23] Casey K S, Adamec D. Sea surface temperature and sea surface height variability in the North Pacific Ocean from 1993 to 1999[J]. *Journal of Geophysical Research*, 2002, 107(C8): 3099.
- [24] 杜金博, 李春, 刘征宇. 基于热含量探究PDO的三维结构及信号传播特征[J]. *海洋气象学报*, 2020, 40(1): 1–11.
Du Jinbo, Li Chun, Liu Zhengyu. Study on three-dimensional structure and signal propagation features of PDO based on heat content[J]. *Journal of Marine Meteorology*, 2020, 40(1): 1–11.
- [25] Cheng Lijing, Trenberth K E, Fasullo J, et al. Improved estimates of ocean heat content from 1960 to 2015[J]. *Science Advances*, 2017, 3(3): e1601545.
- [26] Cheng Lijing, Trenberth K E, Palmer M D, et al. Observed and simulated full-depth ocean heat-content changes for 1970–2005[J]. *Ocean Science*, 2016, 12(4): 925–935.
- [27] Sun Cheng, Kucharski F, Li Jianping, et al. Western tropical Pacific multidecadal variability forced by the Atlantic multidecadal oscillation[J]. *Nature Communications*, 2017, 8(1): 15998.

- [28] Gong Zhanqiu, Sun Cheng, Li Jianping, et al. An inter-basin teleconnection from the North Atlantic to the subarctic North Pacific at multidecadal time scales[J]. *Climate Dynamics*, 2020, 54(1): 807–822.
- [29] Sun Jianqi, Wang Huijun. Relationship between Arctic Oscillation and Pacific Decadal Oscillation on decadal timescale[J]. *Chinese Science Bulletin*, 2006, 51(1): 75–79.
- [30] Lindsay R W, Zhang J. The thinning of Arctic sea ice, 1988–2003: have we passed a tipping point?[J]. *Journal of Climate*, 2005, 18(22): 4879–4894.
- [31] Zheng Gang, Li Xiaofeng, Zhang Ronghua, et al. Purely satellite data-driven deep learning forecast of complicated tropical instability waves[J]. *Science Advances*, 2020, 6(29): eaba1482.
- [32] Nieves V, Radin C, Camps-Valls G. Predicting regional coastal sea level changes with machine learning[J]. *Scientific Reports*, 2021, 11(1): 7650.
- [33] Haq D Z, Novitasari D C R, Hamid A, et al. Long short-term memory algorithm for rainfall prediction based on El-Niño and IOD data[J]. *Procedia Computer Science*, 2021, 179: 829–837.
- [34] Hochreiter S, Schmidhuber J. Long short-term memory[J]. *Neural Computation*, 1997, 9(8): 1735–1780.
- [35] Guo Yanan, Cao Xiaoqun, Liu Bainian, et al. El Niño index prediction using deep learning with ensemble empirical mode decomposition[J]. *Symmetry*, 2020, 12(6): 893.
- [36] Broni-Bedaiko C, Katsriku F A, Unemi T, et al. El Niño-Southern Oscillation forecasting using complex networks analysis of LSTM neural networks[J]. *Artificial Life and Robotics*, 2019, 24(4): 445–451.
- [37] Cao Xiaoqun, Guo Yanan, Liu Bainian, et al. ENSO prediction based on Long Short-Term Memory (LSTM)[J]. *IOP Conference Series: Materials Science and Engineering*, 2020, 799(1): 012035.
- [38] Chikamoto Y, Kimoto M, Ishii M, et al. An overview of decadal climate predictability in a multi-model ensemble by climate model MIROC[J]. *Climate Dynamics*, 2013, 40(5/6): 1201–1222.
- [39] Volodin E, Gritsun A. Simulation of observed climate changes in 1850–2014 with climate model INM-CM5[J]. *Earth System Dynamics*, 2018, 9(4): 1235–1242.

Research on PDO index prediction based on multivariate LSTM neural network model

Yu Zhenlong¹, Xu Dongfeng^{1,2}, Yao Zhixiong^{1,2}, Yang Chenghao^{1,2}, Liu Songnan¹

(1. State Key Laboratory of Satellite Ocean Environmental Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou 310012, China; 2. Zhejiang Institute of Marine Sciences, Hangzhou 310012, China; 3. State Key Laboratory of Marine Space Resource Management Technology, Ministry of Natural Resources, Hangzhou 310012, China)

Abstract: A multivariate long short term memory (LSTM) neural network model was developed for the Pacific decadal oscillation (PDO) index time series prediction using sea level pressure, sea level height, ocean heat content data and sea ice concentration from 1921 to 2020 as forecast elements of the PDO index. The PDO index prediction results of different time series from 2011 to 2020 were compared and analyzed, and finally the PDO index forecasting from 2021 to 2030 is realized by using the multivariate LSTM neural network model. The results show that the average correlation coefficient and root mean square error of the predicted value and the observed value of the multivariate LSTM model after cross-validation are 0.70 and 0.62, respectively. PDO will remain in the cold phase in the next ten years, and the PDO index will fluctuate twice, there will be a minimum in 2025. Compared with other time series forecasting models, the multivariate LSTM neural network model used in this paper has less error in forecasting results and good fitting effect, which can be used as a new method of predicting PDO index.

Key words: PDO index; LSTM neural network model; time series forecasting