基于海洋气候数据集的区域海平面变化非线性预测

赵 健1, 蔡瑞阳1, 孙伟富2, 杨俊钢2

(1. 中国石油大学(华东)海洋与空间信息学院,山东 青岛 266580; 2. 自然资源部 第一海洋研究所,山东 青岛 266061)

摘要:本文基于中国首套长时间序列、高精度、高时空一致性的全球海洋气候数据集产品,利用 1993 年1月至 2015 年 12月的山东半岛近海海平面异常数据,构建了基于集合经验模式分解(EEMD)和长短 期记忆神经网络(LSTM)的海平面非线性变化组合预测模型。EEMD 可以得到海平面异常的各周期项、 线性趋势及残差部分,LSTM 模型可对其进行逐个预测并重构得到最终的海平面异常预测结果。 EEMD-LSTM 组合模型海平面异常预测的均方根误差仅为 25.87 mm,取得了令人满意的效果。基于该 组合模型预测 2016—2025 年山东半岛近海海平面上升速率将达到 3.54 mm·a⁻¹。

关键词:海平面变化;气候数据集;集合经验模式分解;长短期记忆神经网络;预测 中图分类号:P228 文献标识码:A 文章编号:1000-3096(2023)4-0069-10 DOI:10.11759/hykx20210128002

海平面变化所受影响因子较多,是非线性、非平 稳性且多尺度变化的时间序列,需要长时间且连续 的数据集来监测其变化过程[1],利用传统的统计预 报或机器学习进行区域海平面变化的准确预测存在 一定的技术难度^[2]。近年来,以经验模式分解 (empirical mode decomposition, EMD)为代表的数据 分解方法和神经网络模型开始应用于海平面变化预 测。EMD 是由 Huang 等^[3]提出的一种基于数据驱动 的自适应分解方法,集合经验模式分解(ensemble empirical mode decomposition, EEMD)通过叠加高斯 白噪声进行多次分解来消除 EMD 分解存在的模态 混叠现象^[4]。近年来各种深度学习算法不断演进,循 环神经网络较为适合处理时间序列数据,但存在梯 度消失、梯度爆炸等问题,且若采用 tanh、relu 为激 活函数,则无法处理长时间序列。为解决梯度消失和 爆炸等问题以及更好的预测时间序列数据,循环神 经网络逐渐转变为长短期记忆神经网络(long and short-term memory, LSTM), 如今 LSTM 在长期依赖 信息学习及预测方面应用广泛^[5]。

中国目前海平面变化研究多局限于季节性信号 和低频信号丰富的南海、东海海域,黄、渤海海域的 海平面变化研究相对较少^[6]。张吉等^[7]利用气象耦合 模式模拟了 RCP4.5 情景下 21 世纪南海海平面变化 趋势,王国松等^[8]构建了多模态融合的神经网络模 型预测南海海域海平面变化,赵健^[9]从波形理论出 发重构海平面数据,进行了南海及东海海平面变化 监测。本文利用海洋气候数据集中的海平面异常数 据,构建基于 EEMD 与 LSTM 的组合预测模型,并 应用到山东半岛近海海平面非线性变化预测中,可 为区域海平面变化预测研究提供参考。

1 海洋气候数据集

海 洋 气 候 数 据 集 (climate data records, CDRs)^[10]是"可用于确定气候变率和气候变化,具 有足够时间长度、一致性和连续性的长时间序列数 据,是开展全球气候变化研究的有效数据"。利用 海洋 CDRs 数据研究海洋变化是应对全球气候变 化重要组成部分,对于时空动态变化范围较大的 海洋水体可以较好的揭示其在季节、年际、年代际 尺度上的变化^[11]。

为加强对中国中高纬度海域海平面变化的研究^[12], 本文利用自然资源部第一海洋研究所研制的中国首套 长时间序列、高精度、高时空一致性的海洋气候数据集 中经数据融合后获取的格网化海平面异常(sea level anomaly, SLA)数据,对山东半岛近海(34°00'N~38°00'N,

收稿日期: 2021-01-28; 修回日期: 2021-03-02

基金项目: 国家重点研发计划项目(2016YFA0600102)

[[]Foundation: The National Key Research and Development Program of China, No. 2016YFA0600102]

作者简介:赵健(1981—),男,山东齐河人,博士,副教授,主要研究 方向为海洋空间信息技术与探测,E-mail:zhjianupc@163.com

119°00'E~123°00'E)海平面变化非线性趋势进行分析预测。海平面异常数据可以很好地捕捉海洋变化的特征,能对海平面变化趋势进行准确的监测和评价。本文所采用的 SLA 数据时间范围为 1993 年 1月—2015 年 12月,时间分辨率为 1 个月,空间分辨率为 25 km,与验潮站数据比较其精度指标为 0.2 mm·a^{-1[11]}。考虑到不同纬度数据密度差异,在计算区域平均的总体数值时,采用纬度加权平均方法:

$$\bar{h}_{k} = \frac{\sum_{i} \sum_{j} h_{ijk} \cos(\varphi_{j})}{\sum_{i} \sum_{j} \cos(\varphi_{j})}, \qquad (1)$$

式中, h_{ijk} 为对应地理位置(i, j)和时间 k 的数据值; φ_i 为数据点的纬度。

区域平均后的山东半岛近海 1993—2015 年海平 面异常时间序列如图 1 所示,线性拟合可得该时间 段山东半岛近海海平面上升速率为 3.23 mm·a⁻¹。

2 非线性预测方法

2.1 EEMD 分解

原始序列的极值分布不均匀,或存在混合信号时,进行 EMD 分解会产生上下包络线拟合误差,导致筛选出的分量是固有模式(intrinsic mode functions, IMFs)和异常事件或相邻时间尺度的固有模式组合,即模态混叠现象。海平面异常数据在采集过程中不可避免存在仪器及噪声的干扰,利用 EMD 分解信号会出现模态混叠等问题,降低预测精度。





Wu 等^[4]在 2005 年提出一种以噪声为辅助手段 的数据分析方法即 EEMD 分解。它将原始信号加入 若干次白噪声,把信号和噪声的组合作为待分解信 号进行 EMD 处理,考虑到白噪声的零均值特性, 对分解结果进行多次平均得到真实分量。EEMD 本 质是一种叠加高斯白噪声的多次 EMD 分解^[13],步 骤如下:

步骤 1: 在序列 x(t)加入服从正态分布的白噪 声, 有:

$$x_i(t) = x(t) + n_j(t) \tag{2}$$

步骤 2: 对新序列做 x(t)进行 EMD 分解, 求出 K 个 IMF 分量 $c_i(t)$ 和一个剩余分量 $r_n(t)$:

$$x(t) = \sum_{i=1}^{k} c_i(t) + r_n(t), \qquad (3)$$

式(3)中,固有模式分量 c₁,c₂,…,c_k 分别是原始信号 中包含不同时间尺度的成分,且时间尺度依次增大;

 $r_n(t)$ 代表原始信号的趋势项。

步骤 3: 重复 *M* 次步骤 1 和步骤 2, 每次加入新的白噪声序列;

步骤 4: 求 *M* 次分解后的 IMF 分量的整体平均, 作为序列 *x_i(t)* 的 IMF 分量,表示为:

$$c_{j}\left(t\right) = \frac{1}{M} \sum_{i=1}^{M} c_{ij}\left(t\right), \qquad (4)$$

式(4)中, $c_{ij}(t)$ 表示第 *i* 次加入高斯白噪声后,分解得 到的第 *j* 个 IMF。

2.2 LSTM 网络

LSTM 是一种特殊的循环神经网络,可以学习 长期依赖信息^[14]。它采用记忆单元代替 RNN 隐含层 的神经单元,在处理长序列输入时有更强的梯度处 理能力。LSTM 网络由 LSTM 结构单元串接而成,每 个结构单元决定之前的信息和状态是否保留,具体 结构如下^[15]: 1)决定从细胞状态丢弃什么信息。通过忘记门读 取 *h*_{t-1}和 *x*_t,输出数值 *f*_t。

$$f_t = \sigma[W_f(h_{t-1}, x_t) + b_f].$$
 (5)

2)决定往细胞状态存放什么信息。首先通过输入 门决定更新信息 i_t ,然后通过 tanh 层创建一个新的 候选值向量 \tilde{C}_t 。

$$i_t = \sigma[W_i(h_{t-1}, x_t) + b_i],$$
 (6)

$$C_t = \tanh[W_c(h_{t-1}, x_t) + b_c].$$
 (7)

3)将旧细胞状态 $C_{t-1} = f_t$ 相乘,再加上 $i_t \times C_t$, 根据决定更新状态的程度进行变化。

$$C_t = f_t \times C_{t-1} + i_t \times C_t \,. \tag{8}$$

4)决定输出什么信息。首先,通过输出门决定输出信息 *O_t*,然后通过 tanh 层将其和输出信息 *O_t* 相乘
 得到确定输出部分 *h_t*。

$$O_t = \sigma[W_o(h_{t-1}, x_t) + b_o],$$
 (9)

$$h_t = O_t \times \tanh(C_t), \qquad (10)$$

式(5)~(10)中, x_t 、 h_t 分别为t时刻网络的输入与输出; f_t 、 i_t 、 O_t 、 C_t 分别为遗忘门、输入门、输出门和记忆细胞的输 出; W_x 为权重矩阵; b_x 为偏置; σ 为 sigmoid 激活函数。

3 基于 EEMD 和 LSTM 的组合预测 模型

EEMD-LSTM 组合预测模型主要包括: SLA 序列 的 EEMD 分解、各 IMF 分量的 LSTM 预测以及预测结 果的重构等三部分。首先通过 EEMD 分解将海平面异 常 SLA 时间序列中具有物理含义的信号提取出来,得 到频域稳定的固有模式分量 IMFs;再利用 LSTM 算法 对得到的各 IMF 分量分别建立预测模型;最后将各 IMF 分量预测结果进行重构叠加,得到最终的海平面 异常预测结果。

3.1 海平面异常时间序列 EEMD 分解

对山东半岛近海 1993—2015 年间的 SLA 时间序 列进行 EEMD 分解,得到 7 个子序列和一个残差序 列共 8 个 IMF 分量,如图 2(a)~(h)所示。借助白噪声 的集合扰动进行显著性分析^[16]发现 IMF7 未通过检 验,表明该分量包含真实信号的可能性小,为白噪 声的可能性较大,建立预测模型时应舍弃。IMF1~ IMF6 频率稳定,分别对应 3.97 月、11.96 月、22.92 月、30.56 月、137.5 月(约 11 a)和 91.67 月(约 7 a)







图 2 山东半岛近海海平面异常时间序列经验模态分解分解结果 Fig. 2 EEMD results of the SLA time series around the Shandong Peninsula

周期,表现为明显的季节变化、年际变化和 2~3 a 周 期变化,基本包含原始序列的显著周期^[17]。表明 EEMD 分解可以将原始非线性的海平面异常时间序 列分解成多个平稳的子序列,同时保留有物理含义 的信号。

3.2 LSTM 模型预测

对 EEMD 分解后的 IMF 分量进行预测时,将 1993 年 1 月至 2011 年 5 月的数据作为训练数据(约 占数据总量的 80%), 2011 年 6 月至 2015 年 12 月的 数据作为检验数据(约占数据总量的 20%)。在构建 LSTM 网络对 IMF 分量进行预测时,需选取不同的 学习率、迭代次数、衰落因子等超参数,结合各实验 结果选择达到最优预测效果的超参数设置。图 3 为 IMF1-6 分量及剩余分量在 LSTM 网络检验阶段的预 测值与原始 IMF数据对比,实线为 IMF 分量在检验阶 段的原始数据,虚线为 LSTM 预测结果。从图 3 可以 看出,随着 IMF 分量趋于平稳模型的预测精度逐渐提 高,预测效果相对较差的高频分量 IMF1,由于存在 大量突变点造成预测难度较大,但其总体振幅及走向 与原始数据基本保持一致,整体预测效果符合预期。



Marine Sciences / Vol. 47, No. 4 / 2023





图 3 检验阶段(2011 年 6 月—2015 年 12 月)各信号分量预测结果与学习样本对比 Fig. 3 Comparison of the predictions of IMF components with the testing data in the testing period (2011.06–2015.12)

4 结果与分析

4.1 预测方法对比分析

为评价 EEMD-LSTM 组合模型的预测效果,将 仅采用 LSTM 模型进行预测与 EEMD-LSTM 组合模 型预测结果进行比对。为避免与原始观测数据比对 时产生偏差,首先对 IMF 分量进行归一化处理,再 将 IMF 分量预测结果叠加重构,最终得到 EEMD-LSTM 组合模型预测结果,两种方法检验数据的预 测结果如图 4 所示。可以看出仅采用 LSTM 预测结 果的峰值和谷值与原始海平面异常观测数据较为 接近,但由于模型存在滞后导致误差相对较大;而 EEMD-LSTM 组合模型明显地改善了上述滞后性, 与原始 SLA 观测数据更为吻合。即采用组合模型



图 4 LSTM 算法预测与 EEMD-LSTM 组合模型在检验阶段(2011 年 6 月—2015 年 12 月)预测结果对比 Fig. 4 Comparison of predictions between the LSTM model and the EEMD-LSTM combined model (2011.06—2015.12) 注: LSTM 为长短期记忆网络模型; EEMD-LSTM 为经验模态分解与长短期记忆网络组合模型

预测效果明显优于仅利用 LSTM 网络对原始 SLA 观 测数据进行预测。海平面异常观测数据经平稳及降 噪处理后,再依据信号的不同变化特征进行多尺度 模拟预测效果更好。但由于海平面变化的非平稳性 和非线性,海平面异常时间序列中存在大量的突变 点和极值点,预测误差偏大的点较易出现在时间序 列变化的极值点上。

为定量分析两种方法的预测精度,采用相关系数 (*R*)、平均绝对误差(*M*_{AE})和均方根误差(*R*_{MSE})等三个指标进行预测结果的定量评价^[18],结果如表1所示。从表1可以看出,EEMD-LSTM组合模型预测结果相关系数*R*较LSTM网络直接预测提高0.1,而*R*_{MSE}与*M*_{AE}分别降低21.51 mm和15.73 mm,组合模型预测精度显著提高。

表 1 LSTM 网络预测与 EEMD-LSTM 组合模型预测误 差统计

 Tab. 1
 Statistics of predictions with the LSTM model and the EEMD-LSTM combined model

模型	R	R _{MSE} /mm	$M_{\rm AE}/{ m mm}$
LSTM	0.84	47.38	36.47
EEMD-LSTM	0.94	25.87	20.74

注:LSTM 为长短期记忆网络模型;EEMD-LSTM 为经验模态分解与 长短期记忆网络组合模型;MAE 为平均绝对误差;RMSE 为均方根误差

4.2 EEMD-LSTM 组合模型应用

本文采用 EEMD-LSTM 组合预测模型,对山东 半岛近海未来 10 年的区域海平面变化趋势进行预测, 结果如图 5 所示。图 5 中 1993—2015 年的海平面异 常时间序列由实线表示,未来十年(2016—2025 年)的 海平面异常预测结果由虚线表示,线性拟合得到 2016—2025 年山东半岛近海海平面上升速率将增大 至 3.54 mm/a。《2021 年中国海平面公报》^[19]显示 1980—2021年,黄海沿海海平面上升速率为3.2 mm/a, 预计未来 30 年,山东沿海海平面将上升 60~160 mm。 本文利用海洋气候数据集研究山东半岛近海海平面 变化趋势,预测结果与上述公报研究结论基本相符, 与基于同时间段卫星测高数据的研究结果也较为一 致^[12],表明该组合预测模型具有很好的应用价值。

5 结论

论文基于 EEMD-LSTM 组合预测模型进行区域 海平面变化趋势的预测研究,能够解决 EMD 分解算 法所带来的模态混叠现象,保留原始序列的物理信 息,同时改善 LSTM 网络滞后现象,极值问题得到 缓解,预测结果更加接近原始观测数据,在区域海 平面变化预测中具有较好的应用价值。



图 5 基于 EEMD-LSTM 组合模型的山东半岛近海 2016—2025 年海平面异常预测结果 Fig. 5 Predicted SLAs with the EEMD-LSTM combined model from 2016 to 2025 around the Shandong Peninsula

论文使用中国首套高精度、高时空一致性的 CDRs 数据,对山东半岛近海未来十年海平面变化进行了预测, 结果表明未来十年山东半岛沿海海平面上升速率将增大 至 3.54 mm/a⁻¹,与中国海平面公报研究结果相符,与基 于同时间段卫星测高数据的研究结果也较为一致^[12],验 证了海洋 CDRs 数据和 EEMD-LSTM 组合预测模型的有 效性。同时发现海平面变化预测研究对极值模拟预测能力要求较高^[20],今后可考虑加入海平面的空间特性及其影响因子并开展极值预测方面的研究。

致谢:感谢自然资源部第一海洋研究所提供的海洋气候数据集 CDRs,感谢审稿专家的意见和建议。

参考文献:

- [1] ARMIN A, XIAO L. Estimating sea level rise around Australia using a new approach to account for low frequency climate signals. 2020, 65(10): 2324-2338.
- [2] 赵健, 樊彦国, 张音. 基于 EEMD-BP 组合模型的区域海平面变化多尺度预测[J]. 系统工程理论与实践, 2019, 39(10): 2713-2722.
 ZHAO Jian, FAN Yanguo, ZHANG Yin. Multi-scale

prediction of regional sea level variations based on EEMD- BP combined model[J]. System Engineering Theory and Practice, 2019, 39(10): 2713-2722.

- [3] HUANG N, SHEN Z, LONG S, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis[J]. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 1998, 454(1971): 903-995.
- [4] WU Z, HUANG N. Ensemble empirical mode decomposition: A noise assisted data analysis method[J]. Advances in Adaptive Data Analysis, 2009, 1(1): 1-41.
- [5] 杨丽, 吴雨茜, 王俊丽, 等. 循环神经网络研究综述[J]. 计算机应用, 2018, 38(S2): 1-6, 26.
 YANG Li, WU Yuxi, WANG Junli, et al. Research on recurrent neural network[J]. Journal of Computer Applications, 2018, 38(S2): 1-6, 26.
- [6] ZHAO J, FAN Y, MU Y. Sea level prediction in the Yellow Sea from satellite altimetry with a combined least squares-neural network approach[J]. Marine Geodesy, 2019, 42(4): 344-366.
- [7] 张吉, 左军成, 李娟, 等. RCP4.5 情景下预测 21 世纪南海海平面变化[J]. 海洋学报(中文版), 2014, 36(11):
 21-29.

ZHANG Ji, ZUO Juncheng, LI Juan, et al. Sea level variations in the South China Sea during the 21st century under RCP4.5[J]. Haiyang Xuebao, 2014, 36(11): 21-29.

- [8] WANG G, WANG X, WU X, et al. Multimodal fusion for sea level anomaly forecasting[P/OL].2020-06-15 [2021-01-25]. https://arxiv.org/abs/2006.08209v1.
- [9] 赵健. 中国南海及东海近海测高海平面变化监测与 预测[J]. 测绘学报, 2020, 49(5): 670. ZHAO Jian. Sea level change monitoring and prediction using satellite altimetry in south China Sea and east China sea coastal ocean[J]. Acta Geodaetica et Cartographica Sinica, 2020, 49(5): 670.
- [10] 梁顺林, 唐世浩, 张杰, 等. 全球气候数据集生成及 气候变化应用研究[J]. 遥感学报, 2016, 20(6): 1491-1499.

LIANG Shunlin, TANG Shihao, ZHANG Jie, et al. Production of the global climate data records and applications to climate change studies[J]. Journal of Remote Sensing, 2016, 20(6): 1491-1499.

- [11] 张杰,孙伟富,杨俊钢,等.海洋气候数据集生成与分析简介[J].海洋科学进展,2019,37(2):325-331.
 ZHANG Jie, SUN Weifu, YANG Jungang, et al. Brief introduction to the first global ocean climate data records(CDRs)of China[J]. Advances in Marine Science, 2019, 37(2): 325-331.
- [12] ZHAO J, CAI R, FAN Y. Prediction of sea level nonlinear trends around Shandong peninsula from satellite altimetry[J]. Sensors, 2019, 19(21): 4770.
- [13] WANG T, ZHANG M, YU Q, et al. Comparing the applications of EMD and EEMD on time-frequency analysis of seismic signal[J]. Journal of Applied Geophysics, 2012, 83: 29-34.
- [14] HOCHREITERS S, SCHMIDHUBER J. Long short-term memory. Neural Computation. 1997, 9(8): 1735-1780.
- [15] 彭燕, 刘宇红, 张荣芬. 基于 LSTM 的股票价格预测 建模与分析[J]. 计算机工程与应用, 2019, 55(11): 209-212.
 PENG Yan, LIU Yuhong, ZHANG Rongfen. Modeling and analysis of stock price forecast based on LSTM[J]. Rong fen Computer Engineering and Applications, 2019, 55(11): 209-212.
- [16] 易彩. 高速列车轮对轴承状态表征与故障诊断方法研究[D]. 成都:西南交通大学, 2015.
 YI Cai. State Characterization and fault diagnosis research on wheel bearing of high-speed train[D]. Chengdu: Southwest Jiaotong University, 2015.
- [17] 王慧,刘克修,张琪,等.中国近海海平面变化与 ENSO 的关系[J].海洋学报(中文版), 2014, 36(9): 65-74.
 WANG Hui, LIU Kexiu, ZHANG Qi, et al. The relationship between sea level change of China's coast and ENSO[J]. Haiyang Xuebao, 2014, 36(9): 65-74.
- [18] 赵健, 樊彦国, 丁宁. 基于最小二乘与径向基函数神 经网络的海平面变化预测[J]. 海洋科学, 2018, 42(5): 92-97.
 ZHAO Jian, FAN Yanguo, DING Ning. Sea level anomaly forecasting using least square and the radial basis function neural network[J]. Marine Sciences,

2018, 42(5): 92-97.
[19] 自然资源部. 2021年中国海平面公报[Z]. 北京:自然资源部, 2022.
Ministry of Natural Resources of the People's Republic of China. China sea level bulletin 2021[Z]. Beijing: Ministry of Natural Resources of the People's Republic of China, 2022.

[20] PENG H, FANG H. Understanding and reconstructing the coastal sea level variations along the western boundary of the North Pacific[J]. Geoscience Letters: Official Journal of the Asia Oceania Geosciences Society(AOGS), 2020, 7: 1-10.

Prediction of regional sea-level nonlinear change based on ocean climate data records

ZHAO Jian¹, CAI Rui-yang¹, SUN Wei-fu², YANG Jun-gang²

College of Oceanography and Space Informatics, China University of Petroleum, Qingdao 266580, China;
 First Institute of Oceanography, MNR, Qingdao 266061, China)

Received: Jan. 28, 2021

Key words: sea level change; climate data records; ensemble empirical mode decomposition; long short-term memory; prediction

Abstract: China's first global ocean climate data records exhibiting long time-series data, high accuracy, and spatiotemporal consistency were used to study regional nonlinear sea-level changes. Using the sea-level anomaly (SLA) data around the Shandong Peninsula from January 1993 to December 2015, we established a combined model based on ensemble empirical mode decomposition (EEMD) and the long and short-term memory (LSTM) approach to forecast nonlinear sea-level trends around the Shandong Peninsula. Herein, period terms, noise terms and residuals (trend item) are individually obtained from EEMD, forecasted using an LSTM neural network, and then reconstructed to obtain the trends in sea-level change. The EEMD–LSTM combined model may be a valuable approach in the prediction of sea-level change as indicated by the minimum 25.59 mm root mean square error values in the SLA prediction obtained during the testing period. The model predicts that, for the period 2016–2025, the rate of sea-level rise around the Shandong Peninsula will increase to 3.54 mm·a^{-1} .

(本文编辑:杨 悦)