

基于SARIMA模型的近岸海表温度短期预报研究

赵强, 王擎宇, 舒志光

(自然资源部宁波海洋中心, 浙江 宁波 315012)

摘要: 基于石浦海洋站实测数据,采用周期性自回归积分滑动平均方法(SARIMA)构建了逐时海表温度短期预报模型,根据观测数据的周期特征和模型预报误差比选确定了模型参数。结果表明:与采用逐时观测数据作为输入的模型相比,采用逐0.5 h内插数据构建的SARIMA模型的预报结果与实测数据间的相位更为一致,预报误差更小,但进一步将输入数据的时间分辨率提高,72 h逐时预报精度提升不明显;研究还发现模型预报误差总体随输入数据时长的减小而增大;采用366 d逐0.5 h数据构建的SARIMA(2,0,2)(2,1,0)₂₄模型的预报结果较优,0~24 h、24~48 h、48~72 h预报的平均绝对误差分别为0.176 °C、0.350 °C、0.520 °C,相应的均方根误差分别为0.217 °C、0.396 °C、0.567 °C。

关键词: 周期性自回归积分滑动平均方法;统计预报;海表温度;预报

中图分类号: P731.31 **文献标识码:** A

文章编号: 1003-0239(2024)01-0042-08

0 引言

近岸海表温度(Sea Surface Temperature, SST, 简称海温)与渔业养殖、滨海旅游等人类生产生活息息相关,是海洋预报的主要工作内容之一。目前常用的海温预报方法主要有三类^[1]:经验预报、数值预报和统计预报。经验预报是根据海温变化的周期性和持续性、结合天气系统特征进行预报,预报质量多取决于预报员的工作经验,预报产品通常为日平均海温或日最高、最低海温,难以满足日益增长的精细化海洋预报需求。数值预报则基于海洋动力学方程组,以气温、短波辐射、感热、风、潮等气象和海洋预报要素为强迫场驱动海温初始场,获得高时空分辨率的海温预报结果。但由于受到地形分辨率、观测数据、初始场和强迫场质量、模型参数取值不确定等因素限制,海温数值预报结果的准确度在近岸往往会下降,需要结合各类释用技术提高预报结果的可靠性^[2-3]。统计预报则利用回归分析、聚类分析、主成分分析、相似分析等时间序列分析

方法^[4],通过海温数据自身特征或海温与气温、气压等不同数据间的关系和规律探究数据的变化趋势。它是一种客观统计方法,计算效率相对较高,特别适合在具备观测基础的单点开展。

自回归积分滑动平均(Autoregressive Integrated Moving Average, ARIMA)是一种经典的时间序列分析和预测方法^[5],它综合了自回归(Autoregressive, AR)和滑动平均(Moving Average, MA)的特点,通过引入差分计算,满足了AR和MA对时间序列平稳性的要求,被广泛应用于社会和自然科学的各个领域。周期性自回归积分滑动平均(Seasonal Autoregressive Integrated Moving Average, SARIMA)是在ARIMA的基础上引入与数据周期性相关的参数,更适用于具有显著周期特征数据的预测分析^[6]。

在海洋领域,ARIMA及相关方法已被应用于赤潮和海温的季节和年际趋势预测研究^[7-10]以及潮位和浪高的单步短时临近预测^[11-12],而其在短期预报特别是近岸海温短期预报方面的准确度尚不明

收稿日期: 2022-11-01。

基金项目: 国家重点研发计划(2018YFC1407000)。

作者简介: 赵强(1982-),男,工程师,博士,主要从事海洋数据分析和数值模拟工作。E-mail: zhaoqiang@ecs.mnr.gov.cn

确。目前,国内各海洋预报机构面向公众提供的近岸海温预报多为基于人工经验或数值预报得到的单点预报,预报时效多为24 h,预报产品为日平均海表温度或日最高、最低海表温度。本文利用石浦海洋站实测海表温度,采用SARIMA方法构建了一个预报时效为72 h的逐时海表温度预报模型,计算了模型预报结果误差,探讨了输入数据的时间分辨率和时长对预报误差的影响。研究结果对近岸海温的短期预报工作具有参考作用。

1 数据来源

本文所使用的数据为石浦海洋站2020年1月1日—2021年12月31日的逐时海表温度(见图1)。石浦海洋站位于浙江省宁波市象山县石浦镇东门岛,为国家标准海洋站。海温测量采用YZY4型温度传感器,测量范围为 $-5\sim 50\text{ }^{\circ}\text{C}$,测量精度为 $0.1\text{ }^{\circ}\text{C}$ 。

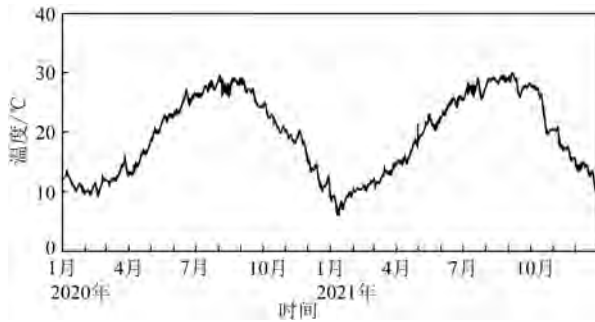


图1 石浦海洋站2020年1月1日—2021年12月31日逐时海表温度

Fig.1 Hourly SST at Shipu Station from Jan 1, 2020 to Dec 31, 2021

2 SARIMA模型构建

2.1 SARIMA模型

SARIMA模型通常写作 $\text{SARIMA}(p,d,q)(P,D,Q)_s$,其中 p,d,q 分别表示自回归、差分和滑动平均的阶数,而 P,D,Q 则表示周期为 s 的周期性自回归、差分和滑动平均的阶数。模型的数学表达式为^[13]:

$$\Phi_p(B^s)\phi_p(B)(1-B^s)^d(1-B)^d y_t = c + \Theta_q(B^s)\theta_q(B)\varepsilon_t \quad (1)$$

式中: y_t 为 t 时刻的时间序列值; ε_t 为 t 时刻的白噪声

序列值; c 为常数项; B 为延迟算子; B^s 表示将 y_t 在时间上向后移 s 次,即 $B^s y_t = y_{t-s}$; $\phi_p(B)$ 和 $\theta_q(B)$ 分别表示 p 阶自回归多项式和 q 阶滑动平均多项式; $\Phi_p(B^s)$ 和 $\Theta_q(B^s)$ 分别表示周期为 s 的 P 阶周期性自回归多项式和 Q 阶周期性滑动平均多项式。

由此可见,一个SARIMA模型包含 p,d,q,s,P,D,Q 共7个参数,这些参数的确定过程称为模型定阶。

2.2 观测数据前处理

2.2.1 平稳性检验

平稳性是指时间序列在未来的一段时期内能沿着现有的形态持续下去、不存在随时间变化的趋势性或者周期性,即序列的均值和方差不随时间发生明显变化^[14]。若时间序列为非平稳,需要通过差分将其转化为平稳时间序列,以满足自回归和滑动平均模型的计算要求。

时间序列的平稳性检验可以通过数据曲线直观判断,也可以通过ADF检验(Augmented Dickey-Fuller Test)^[15]等单位根检验方法进行判断。若存在单位根,则说明时间序列存在趋势性或周期性,是非平稳的。近岸海温在长周期上受季节变化的影响,在短周期上受气温日变化和潮汐等周期性因素的影响,显然是非平稳的。ADF检验的结果也显示其存在单位根,因此需要进行差分运算以消除其周期性。

2.2.2 非白噪声检验

非白噪声检验是对时间序列的自相关性进行检验,通常采用Ljung-Box Q检验方法^[5]。若通过检测,则时间序列为非白噪声,即一个时刻的观测值会对某时刻之后的观测值产生影响,意味着该时间序列是可预测的;否则,该时间序列为白噪声,不可预测。对石浦站逐时海表温度数据的一阶差分进行非白噪声检验,结果显示其为非白噪声,即可通过数据的自身特征进行预测。

2.3 模型定阶

$\text{SARIMA}(p,d,q)(P,D,Q)_s$ 模型有7个参数需要确定。周期参数 s 可以通过分析时间序列的特征来确定。对逐时海表温度数据做调和分析,前10位成分的周期见表1,其中包括1个年周期SA,1个半年周

期 SSA, 2 个月周期 MSM、MM, 2 个半月周期 MSF、MF, 1 个日周期 S1, 3 个半日周期 M2、H2、S2。短期逐时预报不考虑长周期, 因此 s 取 12。近岸逐时海温为非平稳时间序列, 通过 1 阶差分或 1 阶周期性差分都可将其转变为平稳时间序列, 因此差分参数 d 和 D 可以取 0 或 1。自回归参数 p 和 P 、滑动平均参数 q 和 Q 的取值可先根据时间序列的自相关系数 (Auto-Correlation Function, ACF) 拖尾和偏自相关系数 (Partial Auto-Correlation Coefficient, PACF) 截尾的性质进行初步判断, 之后利用贝叶斯信息准则 (Bayesian Information Criterion, BIC)、热力图、误差评估等方法结合网格搜索进行确定。

表 1 石浦海洋站 2 a 逐时海表水温调和结果
Tab. 1 Harmonic analysis result of the 2-year hourly SST at Shipu Station

分潮周期名	周期/h	振幅/℃
SA	8 764.24	9.21
SSA	4 382.12	0.63
MSM	763.48	0.32
MM	661.29	0.21
MSF	354.37	0.15
S1	24.00	0.12
MF	327.86	0.11
M2	12.42	0.09
H2	12.40	0.06
S2	12.00	0.06

由于本文的目的是构建一个短期预报模型, 因此对 SARIMA 模型赋以不同的参数值, 并比较各组参数值下的模型预报结果误差, 最终选择其中预报误差最小的一组参数作为最终参数。模型的输入数据为 366 d 的逐时海温, 输出数据为之后 0~72 h 的逐时海温预报结果, 通过计算模型预报误差进行比较评估。为减少评估结果的偶然性和季节性并充分利用 2 a 的海温实测数据, 本文开展了 360 组预报模型计算, 每组输入数据的开始时间皆以 1 d 为间隔后延, 即第 1 组的输入数据为 2020 年 1 月 1 日 00 时 (北京时, 下同)—12 月 31 日 24 时的逐时海温, 计算 2021 年 1 月 1 日 00 时—1 月 3 日 24 时的预报误差, 第 2 组的输入数据为 2020 年 1 月 2 日 00 时—2021 年 1 月 1 日 24 时的逐时海温, 计算 2021 年 1 月 2 日 00 时—1 月 4 日 24 时的预报误差, 以此类推。每组都重新计算模型中的自回归系数和滑动平均系数, 共得到 360 组预报误差。每组误差包括平均绝对误差 (Mean Absolute Error, MAE)、平均绝对百分比误差 (Mean Absolute Percentage Error, MAPE) 和均方根误差 (Root Mean Square Error, RMSE), 并按照 0~24 h、24~48 h、48~72 h 的预报时效进行划分。比较不同参数取值下模型预报误差的平均值, 结果见表 2, 受篇幅限制, 仅列出部分误差较小的模型参数。

从结果来看, 模型预报误差随预报时效的增加而明显增大, 但同一预报时效、不同参数的模型预

表 2 不同参数下 SARIMA 模型的预报误差
Tab. 2 Forecasting errors of SARIMA model under different parameters

参数	MAE/℃			MAPE/%			RMSE/℃		
	0~24 h	24~48 h	48~72 h	0~24 h	24~48 h	48~72 h	0~24 h	24~48 h	48~72 h
SARIMA(0, 0, 0)(2, 1, 2) ₁₂	0.201	0.390	0.561	1.162	2.264	3.246	0.248	0.447	0.623
SARIMA(0, 0, 1)(2, 1, 0) ₁₂	0.205	0.395	0.563	1.187	2.277	3.229	0.250	0.448	0.620
SARIMA(0, 0, 2)(2, 1, 1) ₁₂	0.217	0.409	0.576	1.256	2.355	3.298	0.261	0.459	0.628
SARIMA(1, 0, 0)(1, 1, 0) ₁₂	0.211	0.398	0.566	1.223	2.292	3.238	0.257	0.449	0.616
SARIMA(1, 0, 1)(1, 1, 0) ₁₂	0.202	0.383	0.551	1.170	2.201	3.154	0.248	0.434	0.602
SARIMA(1, 0, 2)(1, 1, 0) ₁₂	0.203	0.384	0.552	1.174	2.209	3.162	0.248	0.435	0.603
SARIMA(2, 0, 0)(1, 1, 0) ₁₂	0.203	0.386	0.554	1.178	2.216	3.170	0.249	0.436	0.605
SARIMA(2, 0, 1)(1, 1, 0) ₁₂	0.202	0.384	0.552	1.172	2.204	3.157	0.248	0.434	0.603
SARIMA(2, 0, 2)(1, 1, 0) ₁₂	0.205	0.388	0.554	1.192	2.237	3.181	0.251	0.438	0.605

报误差较为接近,这与近岸海温受潮位影响较强且变化相对缓慢有关。综合来看, SARIMA(1,0,1)(1,1,0)₁₂模型的预报误差最小。

3 结果和讨论

3.1 输入数据时间分辨率对预报结果的影响

SARIMA(1,0,1)(1,1,0)₁₂模型中的周期参数值12与该海域占主导地位的M2半日分潮的潮周期12.42 h存在差异,这导致了预报海温与实测海温之间的相位差。以2021年1月10日00时起 SARIMA(1,0,1)(1,1,0)₁₂模型0~72 h的预报结果为例(见图2),与实测海温相比,预报海温的相位随着预报时效的增加不断向前偏移。由于 SARIMA 模型的参数只能取整数,如果周期参数值取13,预报结果的相位则会随着预报时效的增加不断向后偏移。

调整输入数据的时间分辨率可以减小预报结果的相位偏移。利用线性插值将原始的逐时海温观测数据内插为逐0.5 h并作为 SARIMA 模型的输入数值,周期参数 s 设为25,构建 SARIMA(2,0,2)(2,1,0)₂₅模型并计算整点预报误差。相较于采用逐时输入数据的 SARIMA(1,0,1)(1,1,0)₁₂模型, SARIMA(2,0,2)(2,1,0)₂₅模型预报结果的相位与观测更为一致(见图2),预报误差也相应减小(见表3),0~24 h、24~48 h、48~72 h的MAE平均值分别较 SARIMA(1,0,1)(1,1,0)₁₂模型下降了12.87%、8.62%、5.63%, RMSE平均值下降了12.50%、8.76%、5.81%。

将输入数据的时间分辨率进一步提高到逐6 min并构建 SARIMA(10,0,10)(2,1,0)₁₂₄模型,相较于 SARIMA(2,0,2)(2,1,0)₂₅模型,预报MAE平均值仅下降了0.001~0.002 °C。究其原因,当采用逐时输入数据时,周期参数为12,对应预报结果的周期为

12 h,与M2分潮的潮周期相比,每个周期的相位差为0.42 h,对应24 h和72 h的相位差分别为0.84 h和2.52 h;当采用逐0.5 h输入数据时,周期参数为25,对应预报结果的周期为12.5 h,每个周期的相位差为0.08 h,对应24 h和72 h的相位差分别为0.16 h和0.48 h,受逐时预报结果时间分辨率的限制,低于0.5 h的相位差不会体现在预报结果和观测值的对比曲线上。同理,进一步提高输入数据的时间分辨率对于72 h逐时预报结果的改进不明显,而且随着输入数据时间分辨率的提高,计算数据量也会提高,计算效率相应下降。由此可见,以逐0.5 h数据作为模型输入已经可以满足72 h逐时海温预报的要求。

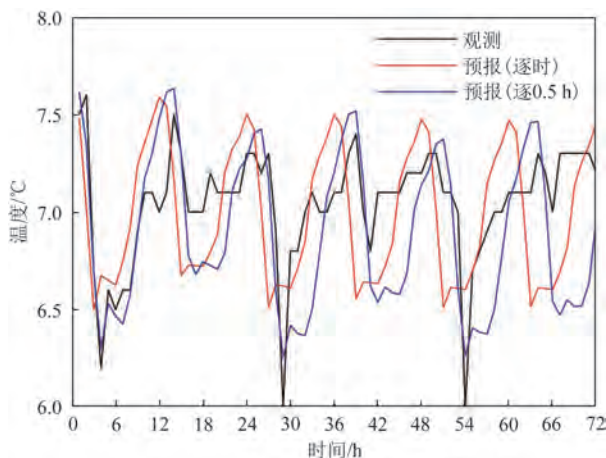


图2 2021年1月10日起0~72 h海温预报结果与观测的对比

Fig.2 The 0~72 h hourly SST observations and forecasts starting from Jan 10, 2021

3.2 输入数据时长对预报结果的影响

基于 SARIMA(2,0,2)(2,1,0)₂₅模型测试不同输入

表3 不同输入数据时间分辨率下 SARIMA 模型的预报误差

Tab.3 Forecasting errors of SARIMA model with different time resolution of inputs

模型	时间分辨率	MAE/°C			MAPE/%			RMSE/°C		
		0~24 h	24~48 h	48~72 h	0~24 h	24~48 h	48~72 h	0~24 h	24~48 h	48~72 h
SARIMA(1,0,1)(2,1,0) ₁₂	1 h	0.202	0.383	0.551	1.170	2.201	3.154	0.248	0.434	0.602
SARIMA(2,0,2)(2,1,0) ₂₅	0.5 h	0.176	0.350	0.520	1.012	2.005	2.982	0.217	0.396	0.567
SARIMA(10,0,10)(2,1,0) ₁₂₄	6 min	0.174	0.348	0.519	0.999	1.997	2.982	0.214	0.392	0.564

表4 不同输入数据时长下SARIMA(2,0,2)(2,1,0)₂₅模型的预报误差Tab.4 Forecasting errors of SARIMA(2,0,2)(2,1,0)₂₅ model at different input data lengths

数据时长/d	MAE/℃			MAPE/%			RMSE/℃		
	0~24 h	24~48 h	48~72 h	0~24 h	24~48 h	48~72 h	0~24 h	24~48 h	48~72 h
366	0.176	0.350	0.520	1.012	2.005	2.982	0.217	0.396	0.567
183	0.181	0.364	0.547	1.039	2.106	3.194	0.222	0.409	0.592
90	0.179	0.357	0.529	1.027	2.064	3.116	0.219	0.402	0.575
30	0.182	0.365	0.542	1.033	2.079	3.313	0.222	0.411	0.589
10	0.189	0.394	0.616	1.079	2.259	3.581	0.231	0.442	0.662
5	0.206	0.453	0.755	1.190	2.687	4.705	0.250	0.502	0.801

数据时长(366 d、183 d、90 d、30 d、10 d、5 d)对预报误差的影响,结果见表4。预报误差总体上是随输入数据时长的减少而增大的,且预报时效越长,误差增幅也越大。当输入数据时长为5 d时,0~24 h、24~48 h、48~72 h的预报MAE平均值分别为0.206℃、0.453℃、0.755℃。利用近岸海表温度变化存在周期性且相对缓慢的特点,采用一种极简的、低技巧的预报方式,即直接复制当前24 h的实测数据作为未来0~24 h、24~48 h、48~72 h的预报结果,其对应的MAE平均值分别为0.275℃、0.471℃、0.628℃。由此可见,当输入数据时长为5 d时,对于当前的预报点,SARIMA(2,0,2)(2,1,0)₂₅模型48~72 h的预报精度已经接近或低于上述低技巧的预报方式,不具备实际应用价值。

比较183 d和90 d输入数据时长的模型预报误差随时间的分布,并未发现183 d预报误差较高的原因,其是否为普遍规律还有待引入更多站点、更长时间序列的数据加以研究。

3.3 预报误差和预报结果评分

图3为采用SARIMA(2,0,2)(2,1,0)₂₅模型计算得到的石浦海洋站2021年1月1日—12月26日0~24 h表层海温逐时预报结果和误差。从图中可以看出,0~24 h逐时预报误差多在±0.5℃以下,其占比为93.84%,误差分布没有呈现显著的周期特征。较大的误差多出现在海温急剧变化的时段,如2021年11月8日,受强冷空气影响,实测海温在4 h内从20.4℃下降到17.8℃,降幅为2.6℃,而对应时段最大预报误差达到-3.3℃。由于没有考虑气象要素,

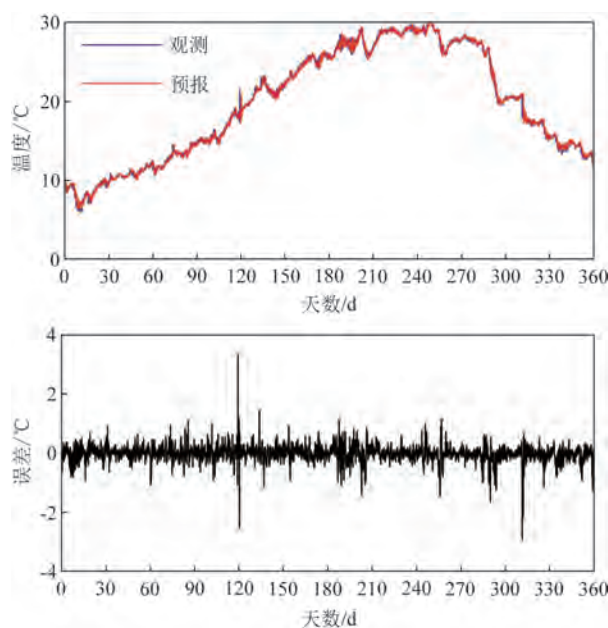


图3 SARIMA(2,0,2)(2,1,0)₂₅模型计算得到的石浦海洋站0~24 h表层海温逐时预报结果和误差

Fig.3 The 0~24 h hourly SST forecasts and the errors at Shipu Station from the SARIMA(2,0,2)(2,1,0)₂₅ model

SARIMA模型在气象条件发生短期剧烈变化或发展趋势发生反转时的预报精度相对较低,引入气象预报数据可能会改善这些时段的海温预报精度,但由于气象变化对近岸海温的影响具有复杂性^[16],也必然会引入新的误差,其预报效果还有待进一步研究。

图4为采用SARIMA(2,0,2)(2,1,0)₂₅模型计算得到的石浦海洋站2021年1月1日—12月26日日平均、日最高、日最低海温的预报误差图。日平均海

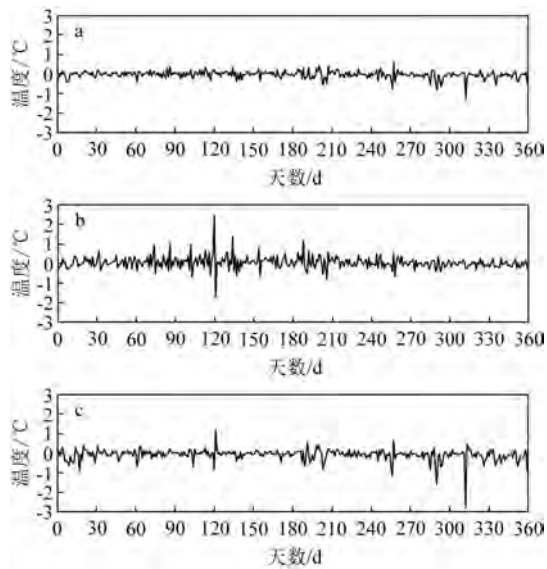


图4 SARIMA(2,0,2)(2,1,0)₂₅模型计算得到的石浦海洋站日平均(a)、日最高(b)、日最低(c)海温预报误差

Fig.4 The forecasting errors of daily average SST, daily maximum SST and daily minimum SST at Shipu Station from the SARIMA(2,0,2)(2,1,0)₂₅ model

温预报误差为 $-1.26 \sim 0.63$ °C, MAE为 0.13 °C;日最高海温预报误差为 $-1.73 \sim 2.45$ °C, MAE为 0.22 °C;日最低海温预报误差为 $-2.82 \sim 1.19$ °C, MAE为 0.18 °C。

按照《海洋预报结果准确性检验评估方法(GB/T 41165—2021)》^[17]的要求对日平均海温预报结果的绝对误差进行评估,按照“表层海水温度绝对误差不大于 2 °C时为满分 100 分;……每相差 0.1 °C扣 5 分”的标准, SARIMA(2,0,2)(2,1,0)₂₅模型 $0 \sim 24$ h日平均温度的预报质量得分在 360 d内皆为 100 分, $24 \sim 48$ h和 $48 \sim 72$ h预报质量的平均得分分别为 99.94 分和 99.49 分。鉴于现阶段近岸海温预报产品多为日最高、最低海温,本文也根据《海水浴场海洋环境预报技术导则》中的海温预报质量评分标准,对日最高、最低海温的预报质量进行综合评分。公式如下:

$$R = \left| \frac{A + B}{2} - C \right| \quad (2)$$

式中: A 和 B 分别表示预报的最低和最高海温; C 表示实测的日平均海温; R 表示海表水温预报绝对误差。质量评分根据 R 值所在区间确定。经计算,

SARIMA(2,0,2)(2,1,0)₂₅模型 $0 \sim 24$ h海温预报质量得分为 100 分、 95 分、 90 分及 90 分以下的天数占比分别为 75.28% 、 18.61% 、 4.44% 和 1.67% ,最低得分为 60 分,平均得分为 98.28 分; $24 \sim 48$ h和 $48 \sim 72$ h的预报平均得分为 93.69 分和 88.22 分。

3.4 与其它方法预报误差的比较

将 2021 年石浦站 SARIMA(2,0,2)(2,1,0)₂₅模型 $0 \sim 24$ h海温预报结果的日平均值与同期人工预报结果进行比较。SARIMA模型预报的MAE为 0.131 °C,最大绝对误差为 1.262 °C;人工预报的MAE为 0.401 °C,最大绝对误差为 2.633 °C。SARIMA模型预报结果优于人工预报结果的天数占比为 82.22% 。

李燕等^[18]基于逐步回归法,采用海温、气温、降水、风速等数据在沿海多个站点建立了单站日均海温短期预报模型, 1 d预报的MAE为 $0.177 \sim 0.544$ °C;匡晓迪等^[2]基于BP(Back Propagation)神经网络方法,采用气象数值预报结果、舟山站海温经验预报和海温观测数据对近岸海温数值预报结果进行释用, 1 d预报的MAE为 0.88 °C;林小刚等^[19]基于长短期记忆网络方法(Long Short-Term Memory Networks, LSTM),采用气象数值预报结果和日均海温观测数据建立了粤东 7 个站点的近岸海温预报模型, 1 d预报的MAE为 $0.24 \sim 0.40$ °C;王兆毅等^[20]基于偏差订正方法,采用海洋站和浮标的逐时海温数据对中国沿海 213 个基础预报单元的海温数值预报结果进行订正, $0 \sim 24$ h、 $24 \sim 48$ h、 $48 \sim 72$ h预报的MAE平均值分别为 0.17 °C、 0.30 °C和 0.38 °C。忽略预报区域和预报时段的差异, SARIMA模型在日均海温和逐时海温预报方面都可以获得较优的预报结果。

4 结论

周期性自回归积分滑动平均(SARIMA)是一种适用于具有显著周期特征的数据序列的客观统计分析方法。本文基于石浦海洋站 2020 年 1 月 1 日— 2021 年 12 月 31 日的逐时海表温度观测数据,采用SARIMA方法构建了 72 h逐时海温预报模型,根据观测数据的周期特征和模型预报误差确定了模型

参数。当采用逐时观测数据作为 SARIMA 模型输入时,由于模型的周期参数只能取整数值,与在该海域占主导地位的 M2 分潮的潮周期 12.42 h 存在差异,导致预报结果与实测数据之间存在相位差,且相位差随着预报时效的增加而不断增大。将模型输入数据的时间分辨率提高到逐 0.5 h,预报结果与观测数据的相位更为一致,预报误差更小。进一步提高输入数据的时间分辨率,对 72 h 逐时预报结果准确度的提升作用不明显。模型输入数据时长会对预报误差产生影响,总体上看,预报误差会随输入数据时长的减小而增大,且误差增幅随预报时效的增加而增大。经比选,采用 366 d 观测海温内插为逐 0.5 h 数据作为输入构建的 SARIMA(2,0,2)(2,1,0)₂₅ 模型的预报结果较优。2021 年 1 月 1 日—12 月 26 日石浦海洋站 0~24 h、24~48 h、48~72 h 预报时效的海温预报的 MAE 平均值分别为 0.176 °C、0.350 °C、0.520 °C,MAPE 平均值为 1.012%、2.005%、2.982%,RMSE 平均值为 0.217 °C、0.396 °C、0.567 °C,参照相关海温预报质量评分标准,对应时效日平均海温预报质量得分平均值为 100 分、99.94 分、99.49 分,日最高和最低海温预报质量综合得分平均值为 98.28 分、93.69 分、88.22 分。

总的来看,SARIMA 模型在日均海温和逐时海温预报方面都可以获得较优的预报结果。由于没有考虑气象要素,SARIMA 模型在气象条件发生短期剧烈变化或发展趋势发生反转时的预报精度相对较低,提高这些时段的预报精度是下一步的研究方向。

参考文献:

- [1] 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 water temperature prediction: lecture 1 overview of sea water temperature prediction[J]. Marine Forecasts, 2003, 20(4): 81-85.
- [2] 匡晓迪, 王兆毅, 张苗苗, 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.
- [3] 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.
- [4] 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 sea water temperature prediction: lecture 2 application of mathematical statistical methods on sea water temperature prediction[J]. Marine Forecasts, 2004, 21(1): 85-90.
- [5] BOX G E P, JENKINS G M, REINSEL G C. Time series analysis: forecasting and control[M]. 3rd ed. Englewood Cliffs, NJ: Prentice Hall, 1994.
- [6] LIU X L, LIN Z, FENG Z M. Short-term offshore wind speed forecast by seasonal ARIMA—A comparison against GRU and LSTM[J]. Energy, 2021, 227: 120492.
- [7] 俞肇元, 袁林旺, 谢志仁, 等. 基于 SSA 和 AR 模型的海面变化预测试验[J]. 海洋湖沼通报, 2007(4): 14-20.
YU Z Y, YUAN L W, XIE Z R, et al. Prediction experiment of sea-level change based on the SSA and AR model[J]. Transactions of Oceanology and Limnology, 2007(4): 14-20.
- [8] HALMOVA D, PEKAROVA P, PEKAR J, et al. The simulation of the water temperature rising using ARIMA models[J]. WSEAS Transactions on Heat and Mass Transfer, 2016, 11: 46-55.
- [9] 张莹, 谭艳春, 彭发定, 等. 基于 EEMD 和 ARIMA 的海温预测模型研究[J]. 海洋学研究, 2019, 37(1): 9-14.
ZHANG Y, TAN Y C, PENG F D, et al. Study on time series prediction model of sea surface temperature based on Ensemble Empirical Mode Decomposition and Autoregressive Integrated Moving Average[J]. Journal of Marine Sciences, 2019, 37(1): 9-14.
- [10] 徐丽丽, 余骏, 高鑫鑫, 等. 一种基于时间序列分析的赤潮预测新方法研究——以浙江海域为例[J]. 海洋预报, 2020, 37(5): 95-103.
XU L L, YU J, GAO X X, et al. Study on the red tide prediction based on time series analysis—A case study in Zhejiang sea area [J]. Marine Forecasts, 2020, 37(5): 95-103.
- [11] 张振全, 李醒飞, 杨少波. 基于 AR-SVR 模型的有效波高短期预测[J]. 太阳能学报, 2021, 42(7): 15-20.
ZHANG Z Q, LI X F, YANG S B. Short-term prediction of significant wave height based on AR-SVR model[J]. Acta Energetica Solaris Sinica, 2021, 42(7): 15-20.
- [12] 刘娇, 史国友, 朱凯歌, 等. 基于调和分析和 ARIMA-SVR 的组合潮汐预测模型[J]. 上海海事大学学报, 2019, 40(3): 93-99.
LIU J, SHI G Y, ZHU K G, et al. A combined tide prediction model based on harmonic analysis and ARIMA-SVR[J]. Journal of Shanghai Maritime University, 2019, 40(3): 93-99.
- [13] ARUNKUMAR K E, KALAGA D V, SAI KUMAR C M, et al. Forecasting the dynamics of cumulative COVID-19 cases (confirmed, recovered and deaths) for top-16 countries using statistical machine learning models: auto-Regressive Integrated Moving Average (ARIMA) and Seasonal Auto-Regressive Integrated Moving Average (SARIMA) [J]. Applied Soft

- Computing, 2021, 103: 107161.
- [14] 耿宏, 王伟, 邢承滨. 基于 ARMA 模型的潮位拟合及内插研究[J]. 海洋测绘, 2021, 41(5): 17-20.
- GENG H, WANG W, XING C B. Research on tide fitting and interpolation based on ARMA model[J]. Hydrographic Surveying and Charting, 2021, 41(5): 17-20.
- [15] DICKEY D A, FULLER W A. Likelihood ratio statistics for autoregressive time series with a unit root[J]. Econometrica, 1981, 49(4): 1057-1072.
- [16] 白彬人, 宋家喜. 中国近海沿岸海温多时间尺度变率及影响其变化的天气气候因素[J]. 海洋预报, 2005, 22(4): 78-88.
- BAI B R, SONG J X. Multi-temporal sea temperature variability in the coastal areas of China and their affecting atmospheric and climatic factors[J]. Marine Forecasts, 2005, 22(4): 78-88.
- [17] 国家市场监督管理总局, 国家标准化管理委员会. GB/T 41165-2021 海洋预报结果准确性检验评估方法[S]. 北京: 中国标准出版社, 2021.
- State Administration for Market Regulation, Standardization Administration. GB/T 41165-2021 Accuracy evaluation methods of marine forecast results[S]. Beijing: Standards Press of China, 2021.
- [18] 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.
- [19] 林小刚, 王兆毅, 李竞时, 等. 基于 LSTM 神经网络方法的粤东近岸海温预报[J]. 海洋预报, 2022, 39(5): 27-36.
- LIN X G, WANG Z Y, LI J S, et al. Sea temperature forecasting based on LSTM neural network along the coast of eastern Guangdong[J]. Marine Forecasts, 2022, 39(5): 27-36.
- [20] 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.

A study of short-term forecast of nearshore SST based on SARIMA model

ZHAO Qiang, WANG Qingyu, SHU Zhiguang

(Ningbo Marine Center, Ministry of Natural Resources, Ningbo 315012, China)

Abstract: Based on the Sea Surface Temperature (SST) data of Shipu Station, time-series model of Seasonal Auto-Regressive Integrated Moving Average (SARIMA) was used to construct a short-term forecasting model for hourly SST. Model parameters were determined according to the periodic of the data and the model forecasting errors. Compared to the model with original hourly input data, the model with interpolated half-hourly input data shows better performance, and the phases of the forecasts have a better consistent with the observations. Using higher temporal resolution of the input data shows no obvious improvement of the accuracy of the 72 h hourly SST forecasts. The results also show that the forecasting error increases with the reduction of the training data length. SARIMA(2, 0, 2) (2, 1, 0)₂₅ model with 366-day interpolated half-hourly SST data shows the best forecasting accuracy. The mean absolute errors of 0~24 h, 24~48 h and 48~72 h forecasts are 0.176 °C, 0.350 °C and 0.520 °C, the corresponding root mean square error are 0.217 °C, 0.396 °C and 0.567 °C, respectively.

Key words: Seasonal Auto-Regressive Integrated Moving Average; statistical forecast; sea surface temperature; forecast