



基于机器学习的渤、黄海夜间海表温度融合技术研究

张洁, 林志佳, 蔡文博

Machine learning-based fusion technique for sea surface temperature in the Bohai and Yellow Seas

Zhang Jie, Lin Zhijia, Cai Wenbo

在线阅读 View online: <https://doi.org/10.12284/hyxb20260006>

您可能感兴趣的其他文章

Articles you may be interested in

一种机器学习海面风场快速融合的方法

Sea surface wind field smart fusion base on machine learning method

海洋学报. 2022, 44(11): 144–158 <https://doi.org/10.12284/hyxb2022137>

基于机器学习的海洋浮标传感器观测数据的偏差校正方法

Machine learning-based bias correction method for ocean buoy sensor observations

海洋学报. 2025, 47(10): 146–154 <https://doi.org/10.12284/hyxb2025083>

顾及空间自相关特征的机器学习水深反演方法研究

Retrieving shallow bathymetry by integrating spatial autocorrelation features with machine learning

海洋学报. 2022, 44(11): 159–169 <https://doi.org/10.12284/hyxb2022033>

基于Himawari-8卫星的逐时次海表温度融合

Hourly sea surface temperature fusion based on Himawari-8 satellite

海洋学报. 2021, 43(1): 137–146 <https://doi.org/10.12284/hyxb2021011>

结合SAM视觉分割模型与随机森林机器学习的无人机影像盐沼植被“精灵圈”提取

Extraction of salt-marsh vegetation “fairy circles” from UAV images by the combination of SAM visual segmentation model and random forest machine learning algorithm

海洋学报. 2024, 46(5): 116–126 <https://doi.org/10.12284/hyxb2024048>

基于机器学习的黄河口邻近海域MODIS叶绿素a浓度遥感反演

Machine learning-based remote sensing retrieval model for MODIS chlorophyll a concentration in adjacent waters of the Huanghe River Estuary

海洋学报. 2025, 47(7): 109–122 <https://doi.org/10.12284/hyxb2025088>

关注微信公众号, 获得更多资讯信息

张洁, 林志佳, 蔡文博. 基于机器学习的渤、黄海夜间海表温度融合技术研究[J]. 海洋学报, 2026, 48(2): 131–143, doi:10.12284/hyxb20260006

Zhang Jie, Lin Zhijia, Cai Wenbo. Machine learning-based fusion technique for sea surface temperature in the Bohai and Yellow Seas[J]. Haiyang Xuebao, 2026, 48(2): 131–143, doi:10.12284/hyxb20260006

基于机器学习的渤、黄海夜间海表温度融合技术研究

张洁, 林志佳, 蔡文博*

(国家海洋环境预报中心, 北京 100081)

摘要: 基于搭载于相应卫星平台的 MODIS 和 AMSR2 传感器观测的夜间海表温度 (Sea Surface Temperature, SST) 数据, 通过构建反向传播神经网络 (BPNN)、随机森林 (RF) 和卷积神经网络 (CNN) 3 种机器学习模型, 开展海表温度数据融合技术研究。在模型输入设计层面, 提出两种差异化方案: 基础方案仅包含经纬度与原始 SST 数据, 增强方案则在此基础上引入时间参数, 进而形成 BP_without_time、BP_with_time、RF_without_time、RF_with_time、CNN_without_time、CNN_with_time 共 6 种融合方案。实验测试结果显示, 在 3 种机器学习模型中, CNN 展现出最为优异的性能表现, 而 RF 模型的性能相对较弱。在 3 种模型的对比测试中, 融入时间参数的增强方案均显著优于未包含时间参数的基础方案。基于 2023–2024 年浮标实测数据的验证结果表明, 融合后的 SST 数据精度略低于 AMSR2_SST, 但相较于 MODIS_SST, 其精度提升效果显著。融合数据的逐月覆盖率较原始数据大幅提升, 2023–2024 年最低覆盖率从 MODIS 的 19.79%、AMSR2 的 32.10% 提升至 49.56% 以上。同时, 高分辨率融合结果能捕捉更精细的温度分布特征, 较 10 km 分辨率的 AMSR2 数据提供更丰富的空间细节。

关键词: 海表温度融合; 机器学习; BPNN; RF; CNN

中图分类号: P731.11

文献标志码: A

文章编号: 0253-4193(2026)02-0131-13

1 引言

海表温度 (SST) 是一个非常重要的海洋环境要素, 代表海洋表层水体的温度^[1]。卫星遥感观测的海表温度, 通常指海洋表层极薄一层的温度 (又称为“皮温”; 而海洋学研究中提及的“表层水温”, 若未特别说明, 可指海洋表面以下 1~20 m 深度的平均水温 (需通过浮标、CTD 等现场观测获取)。具体定义可能因测量方法或应用场景而略有差异。SST 对气候系统、生态系统、人类社会等多个领域都有重要的影响意义。SST 是地球能量平衡的重要指标, 海洋吸收了约

90% 的全球变暖额外热量, SST 的长期升高直接反映温室效应的影响, 因此 SST 是评估全球变暖趋势的核心参数^[2]。同时, SST 分布差异驱动洋流和大气环流 (如信风、季风), 影响厄尔尼诺-南方涛动 (ENSO)^[3]、台风等代表性气候现象的生成。SST 与海洋生态系统也密切相关, 持续升高的 SST 导致珊瑚白化^[4]。2016 年, 澳大利亚大堡礁珊瑚遭受最严重“白化”现象, 北端受灾面积高达 67%, 正是由于大堡礁附近海水表面温度比以往同时期月平均温度升高超过 0.56℃^[5]。SST 对人类社会与经济活动也具有重要的影响, 例如, 具有代表性的海洋灾害——海洋热浪^[6],

收稿日期: 2025-08-22; 修订日期: 2026-01-06。

基金项目: 山东省海洋生态环境与防灾减灾重点实验室 2023 年度开放基金“基于变尺度 Kriging 方法的渤、黄海海表温度融合技术研究” (202301)。

作者简介: 张洁 (1990—), 女, 安徽省滁州市人, 工程师, 研究方向为海洋遥感。E-mail: 915114492@qq.com

* 通信作者: 蔡文博, 高级工程师, 研究方向为计算机应用。E-mail: caiwb@nmefc.cn

海洋热浪对海洋渔业养殖带来巨大的威胁^[7]。海洋热浪与 SST 直接相关,通过监测 SST 的变化可提前预测海洋热浪的发生,在军事与国防方面, SST 影响声呐探测效率,直接影响潜艇活动和水下作战^[8]。因此,综上所述,研究 SST 是我们充分了解海洋,开发利用海洋资源,海洋防灾减灾至关重要的一部分。

目前,人们获取 SST 的途径主要有以下 4 种:岸基、海基、空基、天基。岸基是指通过在沿海、岛屿或固定平台(如码头、灯塔、海洋观测塔)上部署设备(分为接触式测温设备如固定式测温站和非接触式遥感如红外辐射计和地波雷达等),对近岸海域表层水温进行连续或定点监测的技术手段。岸基观测具有高时空分辨率、长期稳定性强、维护便捷等优势,但也受限于覆盖范围(通常为近岸数千米至数十千米)。海基观测海水温度是通过部署在海洋主体区域的移动或固定平台(浮标或者船舶)直接测量表层水温的技术手段,具有高精度、连续性强、覆盖范围广等特点^[9],是构建全球海洋观测网的核心组成部分,但其空间覆盖率不足。岸基与海基观测均属于现场观测范畴。空基观测 SST 是指通过飞机、无人机等航空平台搭载传感器,从空中对 SST 进行遥感测量的技术。其兼具高时空分辨率和灵活机动性,是 SST 观测的重要补充。但是空基观测 SST 面临成本过高与空域管理的限制,有人机每小时成本超 1 万美元,无人机空域申请流程复杂。天基观测海表温度是指利用卫星搭载的遥感传感器,从太空对海洋表层温度进行全球性、连续性监测的技术。作为当前 SST 观测数据的主要来源,天基观测具有覆盖广、时效强、成本低的优势,是构建全球海洋监测网络的基石^[10]。天基观测 SST 的传感器类型主要有红外辐射计和微波辐射计^[11],二者的区别见表 1。

由表 1 可知,红外和微波观测 SST 都有各自的优势和不足:红外观测 SST 的空间分辨率高,但是容易受云层和水汽干扰;微波观测 SST 对云层、光照等光学干扰因素不敏感,但是它的分辨率较低。因此,如何充分利用红外、微波的优点,将二者观测的 SST 融

表 1 红外辐射计和微波辐射计观测 SST 的区别

Table 1 Differences between SST observations by infrared and microwave radiometers

传感器类型	工作波段	测量深度	优势	局限性
红外辐射计	热红外 (8~14 μm)	表皮层 (几微米)	空间分辨率高	受云层、水汽干扰
微波辐射计	微波 (6~37 GHz)	次表层 (1~5 mm)	穿透云层	分辨率较低

合,成为学者们研究的一个热点。目前,红外、微波 SST 融合的主流方法是客观分析法(Objective Analysis, OA)^[12]、最优插值(Optimal Interpolation, OI)^[13]、数据同化(Data Assimilation, DA)^[14]和贝叶斯方法(Bayesian Fusion)^[15]。

客观分析法是基于统计插值(如克里金法),假设 SST 空间分布满足特定统计模型(如高斯相关函数),通过最小化插值误差协方差生成连续场。该方法计算简单便捷,很适合用于一些业务化系统。但是协方差函数参数(如相关尺度)需人工设定,易引入主观偏差。最优插值方法的原理是结合观测数据与背景场(如数值模式预报),通过最小化分析误差协方差矩阵,动态加权多源数据。该方法通过背景场引入物理约束(如地转流关系),保留海洋动力特征,同时显式分离观测误差与背景场误差,实现误差量化。但是,计算复杂度很高,协方差矩阵求逆耗时,需近似算法(如递归 OI);假设误差为高斯分布且时空平稳,难以处理非线性过程;背景场误差直接影响融合结果。数据同化是将多源观测逐步融入数值模型,通过模型动力学约束调整状态变量(如 SST、海流),实现“观测-模式”协同优化。主要包括变分(VAR)和卡尔曼滤波(KF),VAR 缺点是背景误差的方差和观测误差的协方差需要主观设置,KF 则需要在融合前进行尺度变换,会带来新的误差。而且计算成本极高,例如 4D-Var 需迭代求解伴随方程,超算资源依赖性强。贝叶斯方法是基于贝叶斯定理,将 SST 视为随机场,通过先验分布(如历史数据统计)与似然函数(观测不确定性)计算后验概率分布。贝叶斯融合方法主要为贝叶斯层次模型(BHM)和贝叶斯最大熵(BME),二者都需要先验知识作为输入。

近年来,随着神经网络、机器学习、深度学习等技术的发展,越来越多的学者开始探索将这些新兴技术用于海表温度融合^[16-17]。相比传统的统计或物理模型,深度学习方法在处理非线性关系、挖掘高维特征以及适应复杂时空变化方面展现出更强的能力。例如,卷积神经网络(CNN)被用于提取 SST 图像中的空间特征,循环神经网络(RNN)及其变体如长短期记忆网络(LSTM)则被用于建模时间序列中的动态变化。此外,生成对抗网络(GAN)和自编码器等生成模型也被应用于 SST 数据的重构与补充,提升了数据的空间覆盖率和精度。与此同时,多源遥感数据的融合也成为研究热点。通过融合来自不同传感器(如 MODIS、AVHRR、AMSR2 等)的 SST 数据,结合深度学习模型,可以有效提升数据的时空分辨率和一致性。一些

研究还尝试引入注意力机制、图神经网络等先进结构,以更好地捕捉海洋温度场中的复杂依赖关系。

本研究基于红外辐射计(MODIS)和微波辐射计(AMSR2)观测的SST,利用3种机器学习模型(BPNN、RF和CNN),开展黄、渤海夜间SST融合技术研究,在模型输入要素中,加入时间维度,探究时间维度对MODIS_SST和AMSR2_SST融合的影响,利用浮标观测的SST对6种融合方案(BP_without_time、BP_with_time、RF_without_time、RF_with_time、CNN_without_time、CNN_with_time)2023年和2024年SST融合结果进行对比验证。

2 数据与方法

2.1 研究区域

研究区域选择的是渤、黄海(31°~41°N, 117°~127°E)。相比如纬度较低的南海和东海,该海域SST具有显著的季节性变化特征:春季(3~5月),随着太阳辐射增强,SST开始从冬季低温回升,渤海南部和黄海中部升温较快,平均温度从3月的5~8℃升至5月的12~16℃;夏季(6~8月),SST达到全年峰值,渤海大部分区域水温升至22~26℃,黄海中部为24~28℃,南部受黄海暖流余脉影响可能更高;秋季(9~11月),太阳辐射减弱,冷空气南下加速SST下降,9月渤海和黄海温度保持在18~22℃,11月降至10~14℃;冬季(12月至翌年2月),SST降至全年最低,渤海北部近岸可能结冰,水温0~2℃;黄海北部2~5℃,南部受黄海暖流影响可达6~10℃^[18-19]。综上所述,可以发现,渤、黄海区域的SST在一个自然年内变化幅度较大,因此能较好的判断本文提出的融合算法的准确性。

2.2 数据

2.2.1 MODIS_SST

本算法采用的是MODIS(Moderate Resolution Imaging Spectroradiometer,中分辨率成像光谱仪)每日夜间SST数据,空间分辨率是4 km,从NASA Ocean-Color平台上下载(网址:<https://oceancolor.gsfc.nasa.gov/>)^[20]。MODIS搭载于NASA的Terra(1999年发射)和Aqua(2002年发射)卫星,用于全球环境监测。主要利用热红外波段(波段31: 10.78~11.28 μm,波段32: 11.77~12.27 μm)反演海表温度。

2.2.2 AMSR2_SST

AMSR2(Advanced Microwave Scanning Radiometer 2)是日本宇宙航空研究开发机构(JAXA)开发的被动微波辐射计,搭载于GCOM-W卫星上,自2012年

7月开始运行。AMSR2提供全球被动微波(PMW)亚表层(即1 mm)SST数据。AMSR2_SST使用在垂直极化下测量的6.9 GHz和10.7 GHz亮度温度(BT)计算的。垂直极化受风的影响小于水平极化。本研究使用垂直极化下6.9 GHz BT反演的SST,因为10.7 GHz数据在高纬度和冷水中的辐射噪声和地球物理误差比6.9 GHz更大。10 km分辨率AMSR2_SST数据可以从日本宇宙航空研究开发机构(<https://gcom-w1.jaxa.jp>)下载^[21]。为了与MODIS观测的夜间SST相匹配,选择的是AMSR2的降轨观测的SST。

2.2.3 浮标数据

收集研究区域内的浮标观测SST数据,用于对融合的SST进行检验,共15个浮标。

2.3 方法

2.3.1 数据预处理

对MODIS、AMSR2观测的SST都需要进行数据预处理,主要包括两个部分。第一部分是基于数据自带的质量控制(QC)标志符进行质量筛选,通过解析卫星数据产品中内置的质量编码,依据预设的高质量标准,筛选并保留观测质量达到最优等级的SST像元点,剔除因云覆盖、大气干扰、传感器噪声等导致的低质量或无效数据。然后,由于漂流浮标测量海面以下约20 cm处的海水温度,而卫星海表温度则从海面以下1 μm(红外)到1 mm(微波)处获取。虽然漂流浮标和卫星海表温度的测量深度略有不同,但漂流浮标测量结果与卫星海表温度之间存在很强的相关性。在这种情况下,有必要尽量减少卫星和现场SST之间测量深度不同造成的偏差。因此,在MODIS海表温度基础上增加了0.17℃以减弱冷表皮效应,并从AMSR2海表温度中减去0.03℃以缓解夜间的暖偏差^[22]。

2.3.2 数据集构建

2.3.2.1 数据匹配

需要将MODIS观测的SST(用MODIS_SST表示)和AMSR2观测的SST(用AMSR2_SST表示)按照经纬度一一对应起来。如果直接将所有的MODIS观测点和AMSR2观测点一一匹配,则处理起来计算量太大,速度很慢。利用二者的分辨率,直接计算MODIS_SST点对应AMSR2观测的经纬度行列号,再依据行列号找到AMSR2_SST即可。若出现一个AMSR2网格对应多个MODIS网格的情况时,则选择距离AMSR2网格点最近的那个MODIS网格的SST值进行匹配。

2.3.2.2 异常值去除

对匹配得到的数据需要进行异常值去除操作,这一步可以提高数据质量,避免异常值降低模型精度。计算 AMSR2_SST 和 MODIS_SST 之间的差异,得到一个差异值集。如果该差异值超出差异集平均值的 3 个标准差,则将其标记为异常值并去除。

2.3.2.3 归一化处理

经过上述处理,可以得到(MODIS_SST, AMSR2_SST, latitude, longitude)数据集。由于这 4 个属性数据范围差异较大,如果直接进入模型训练,结果会难收敛。因此,需要对数据进行归一化处理,提高神经网络的训练速度和性能。

$$y = \frac{(y_{\max} - y_{\min})(x - x_{\min})}{x_{\max} - x_{\min}} + y_{\min}, \quad (1)$$

式中, x 表示原始数据,即待归一化的输入值,本研究中的 MODIS_SST、AMSR2_SST、latitude、longitude; y 表示归一化后的数据;本研究中设置归一化后的目标范围是[0,1],因此, $y_{\min}=0$, $y_{\max}=1$ 。

2.3.2.4 数据集

最终,经过上述处理得到数据集,收集 2023 年至 2024 年,两年共 1 341 525 个样本,按照 80% : 20%,划分训练集和测试集。

2.3.3 模型训练与评估

本研究选择了 3 种机器学习方法(BPNN、CNN 和 RF)进行训练。3 种模型的数据分配都是 80% 的数据集用于训练,20% 的数据集用于测试,为了测试 3 种模型的性能,3 种模型的训练集和测试集都是完全一致的。

值得注意的是,为了探究时间维度在 sst_modis、sst_amsr2 二者的非线性变化关系构建中的作用,输入要素中增加了时间属性,分别是:

$$\text{time1} = \cos(\text{day_of_year}/365.25), \quad (2)$$

$$\text{time2} = \sin(\text{day_of_year}/365.25). \quad (3)$$

因此输入层的 5 个要素是(sst_amsr2, latitude, longitude, time1, time2),输出(sst_amsr2_modis)。

2.3.3.1 BPNN 模型

反向传播神经网络(BPNN)的概念是由 Rumelhart 和 McClelland 于 1986 年提出。它是一种基于误差反向传播算法训练的多层前馈神经网络,是应用最广泛的神经网络模型之一。在外部输入样本的学习过程中,神经网络通过不断调整网络的连接权重,使网络输出接近预期输出。由于其在回归任务中具有很强的学习能力,BPNN 模型也被广泛应用于卫星遥感要素反演领域^[23-24]。

BP 模型构建了一个由全连接层与激活函数交替组成的前馈神经网络,包含 4 个线性变换层和 3 个非线性激活层,形成深度特征提取能力。BP 模型框架如下(图 1):输入层接受维度为 5 的特征向量,表示每个样本包含 5 个输入变量。随后接入 3 个隐藏层,第一隐藏层包括 16 个神经元,第二隐藏层包括 8 个神经元,第三隐藏层包括 4 个神经元,且都通过 ReLU 激活函数引入非线性。

2.3.3.2 CNN 模型

CNN(卷积神经网络)是一种复杂的深度学习方法,专为识别网格结构数据中的复杂模式而设计,尤其擅长预测连续值^[25-26]。在回归任务中,CNN 通过分层学习数据表示,生成一个连续值作为最终输出。通过系统地提取和组合特征,模型能熟练捕捉输入数据中的复杂模式和相关性,从而实现连续值的精确预测。

本研究的 CNN 模型使用的是 Pytorch 构建的。CNN 模型框架如下(图 2):从一个重塑层开始,为输入数据引入一个额外维度,使其适应适合一维卷积操作的格式。随后,模型包含两个一维卷积层(conv1D),每个后面跟着一个一维最大池化层(maxpooling1D)。这些卷积层擅长捕捉序列数据中的空间模式和分层特征;一维最大池化层则进行下采样,降低特征图的空间维度。在卷积层之后,使用一个展平层(flatten layer)将多维输出转换为一维数组,随后输入到密集

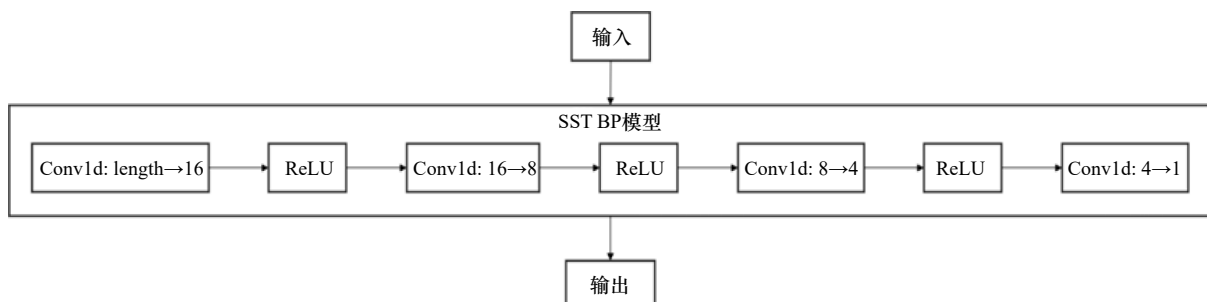


图 1 BP 模型网络结构

Fig. 1 BP model network structure

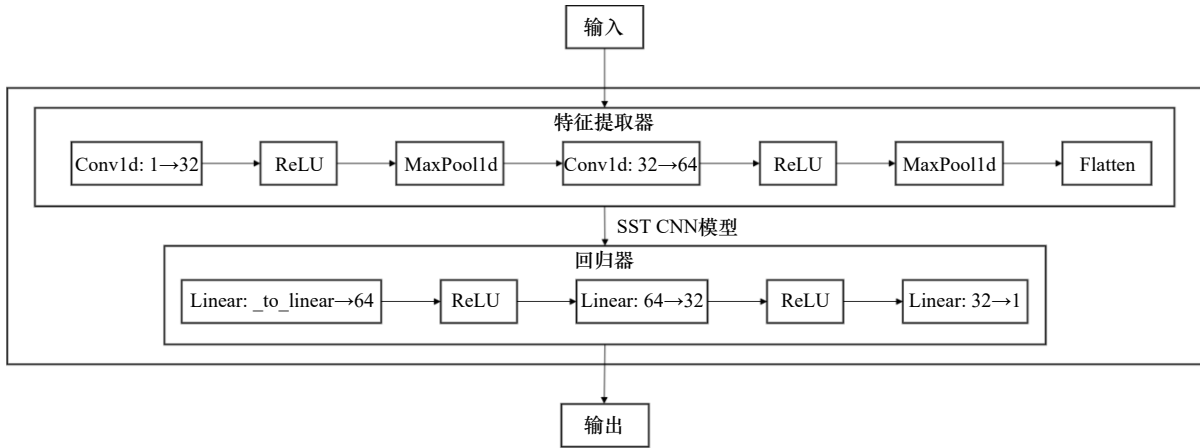


图2 CNN模型网络结构

Fig. 2 CNN model network structure

层(dense layers)进行更高级的特征处理。模型以一个具有128个单元和ReLU激活函数的密集层结束,接下来是一个用于回归的密集输出层——该层仅有1个单元,且没有激活函数。

2.3.3.3 RF模型

随机森林(RF)由Breiman^[27]于2001年提出,是一种监督学习算法。它是基于Breiman^[28]于1996年提出的自助聚合思想的集成学习算法。集成中的每棵树均由从训练数据中抽取的自助样本(bootstrap samples)构建而成。然后,将生成的树进行聚合以用于分类,或取平均以用于回归,从而提供聚合估计。

随机森林回归模型的核心是集成学习中“多棵决策树协同预测”的思想。RF模型网络结构图如图3所示,左侧的“输入特征数据”是模型的输入,对于本研究来说即训练数据集。采用Scikit-learn库中的Random Forest Regressor类实现随机森林模型初始化。n_estimators表示模型内部会构建200棵决策树,该参数是随机森林的核心参数,控制集成的规模,本研究设置n_estimators = 200。每棵决策树会基于Bootstrap

抽样,随机选取样本和特征独立训练,避免过拟合,增加多样性。特征数据会被传入每一棵决策树,每棵树输出一个预测值。由于本研究是回归任务,模型会将200棵决策树的预测值取平均值,作为最终的预测结果。

2.3.3.4 评估

本研究选用6个指标用于评估上述回归模型的优异程度。这6个指标分别为标准差(Standard Deviation, Std)、平均绝对误差(Mean Absolute Error, MAE)、均方误差(Mean Squared Error, MSE)、均方根误差(Root Mean Squared Error, RMSE)、偏差(Bias)和相关系数(R)。这些指标的计算公式如下,其中n为样本总数,真实值为 y_i ,预测值为 \hat{y}_i :

$$Std = \sqrt{\frac{1}{n} \sum_{i=1}^n (e_i - \bar{e})^2}, \quad (4)$$

其中 $e_i = y_i - \hat{y}_i$, $\bar{e} = \frac{1}{n} \sum_{i=1}^n e_i$ 。

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (5)$$

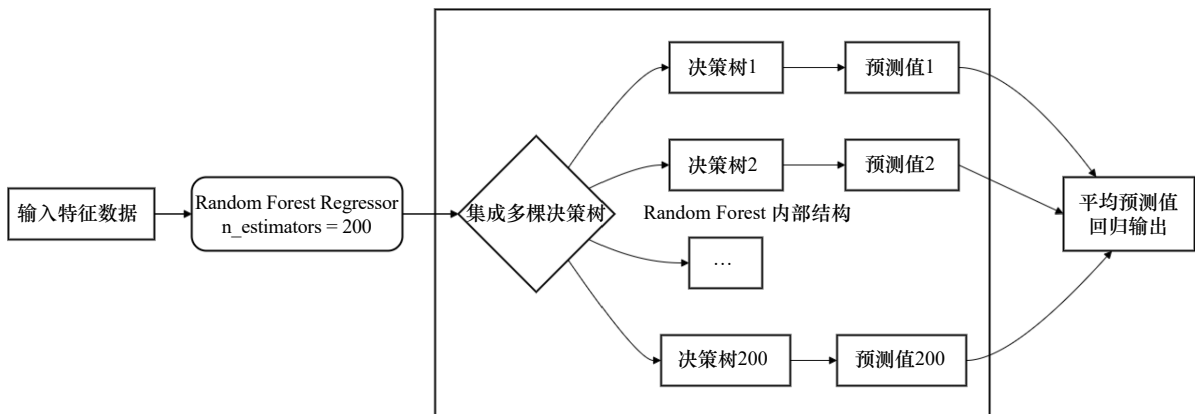


图3 RF模型网络结构

Fig. 3 RF model network structure

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (6)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (7)$$

$$\text{Bias} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i), \quad (8)$$

$$R = \frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2}}. \quad (9)$$

在测试数据上, 3 种模型的优异程度见表 2, 散点图见图 4。这 6 张图分别是 BP、RF 和 CNN 3 种模型两种输入(不包含时间维度和包含时间维度)得到的结果, 右侧的色标表示数值出现的频次。从图上可以看出, 3 种模型的 RMSE 都在 0.5℃ 以下, 其中 CNN 表现最好, 其次是 BP, 最后是 RF。加入时间维度, 对于 3 种模型结果都有改善, BP 模型改善幅度较小, 对 CNN 和 RF 模型改善程度较大。

2.4 利用浮标数据进行检验

浮标观测的 SST 数据的时间分辨率是 1 h, 由于

本论文研究目标是夜间 SST 融合, 因此, 不能直接将浮标 SST 用于检验, 需要对浮标 SST 进行筛选后, 再与卫星观测的 SST 进行匹配。筛选浮标 SST 的思路是对每一个位置实测浮标 SST 进行夜间(00:00–03:00)的数据分析, 若期间最大值 SST 与最小值 SST 之差小于 0.5°, 则认为这个地方的夜间温度是稳定的, 可以用于验证, 否则只要有时间差, 两者由于自身环境确实有变化, 不适宜评估卫星观测的海温。

同样采用上述 6 个统计指标利用浮标 SST 对原始 MODIS_SST、AMSR2_SST 和 6 种融合的 SST 进行检验评估。MODIS_SST 和 AMSR2_SST 的检验结果如图 5 所示, AMSR2 的 SST 观测质量优于 MODIS_SST。

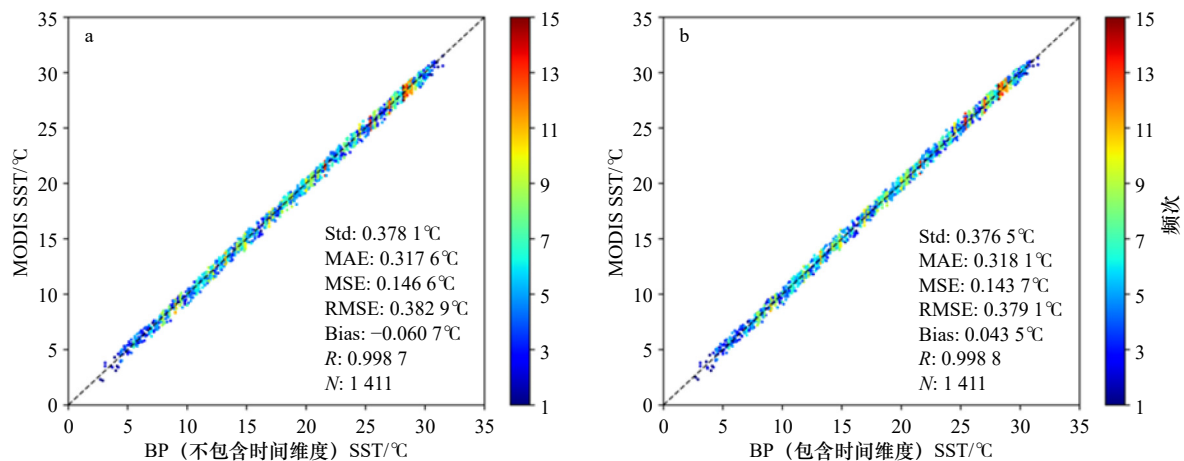
图 6 是 6 种融合方案的 SST 和浮标进行对比验证的检验结果, 3 种模型(BP、RF 和 CNN)中, CNN 表现最好, 在两种输入情况(包含时间和不包含时间维度)下, 不论是 BP、RF 或者是 CNN, 都是包含时间维度输入的方案最优, 包含时间输入的 CNN 模型的融合 SST 与浮标进行检验的 RMSE 可以达到 0.614 7℃。将 6 种融合 SST 的验证结果与原始 MODIS、AMSR2 SST 的验证结果进行对比, 发现融合 SST 的 RMSE 会

表 2 6 种融合方案在测试集上的验证结果

Table 2 Validation results of six fusion schemes on the test set

融合方案	Std/℃	MAE/℃	MSE/℃	RMSE/℃	Bias/℃	R
BP(不包含时间维度)	0.38	0.32	0.15	0.38	-0.06	0.998 7
BP(包含时间维度)	0.38	0.32	0.14	0.38	0.04	0.998 8
RF(不包含时间维度)	0.47	0.38	0.22	0.47	-0.02	0.998 3
RF(包含时间维度)	0.43	0.35	0.19	0.43	-0.04	0.998 5
CNN(不包含时间维度)	0.37	0.31	0.14	0.38	-0.06	0.998 8
CNN(包含时间维度)	0.34	0.28	0.12	0.35	-0.03	0.999 0

注: Std、MAE、MSE、RMSE、Bias 数据仅保留 2 位小数。



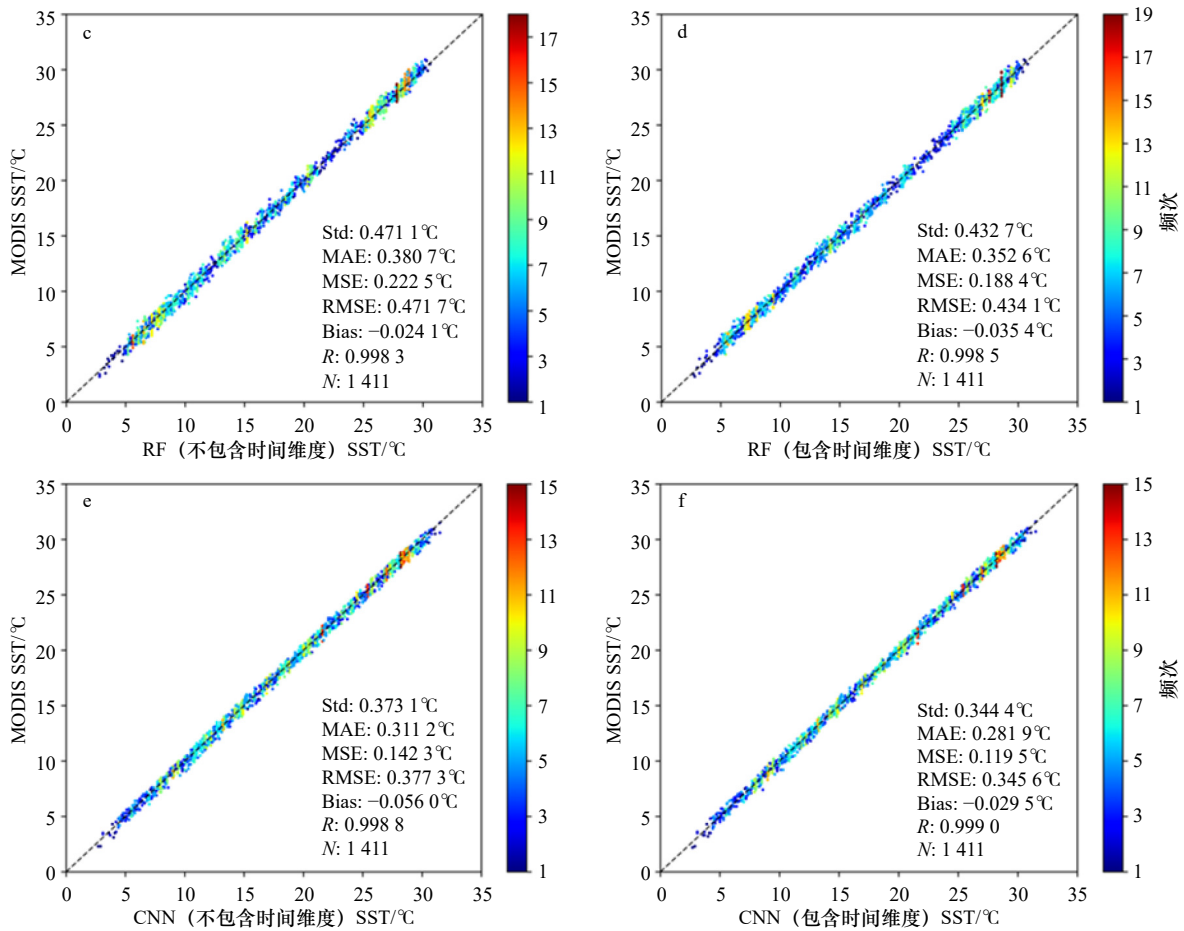


图4 6种融合方案在测试集上的验证结果散点图(N表示测试点数)

Fig. 4 Scatter plot of validation results for six fusion schemes on the test set (N represents the number of test points)

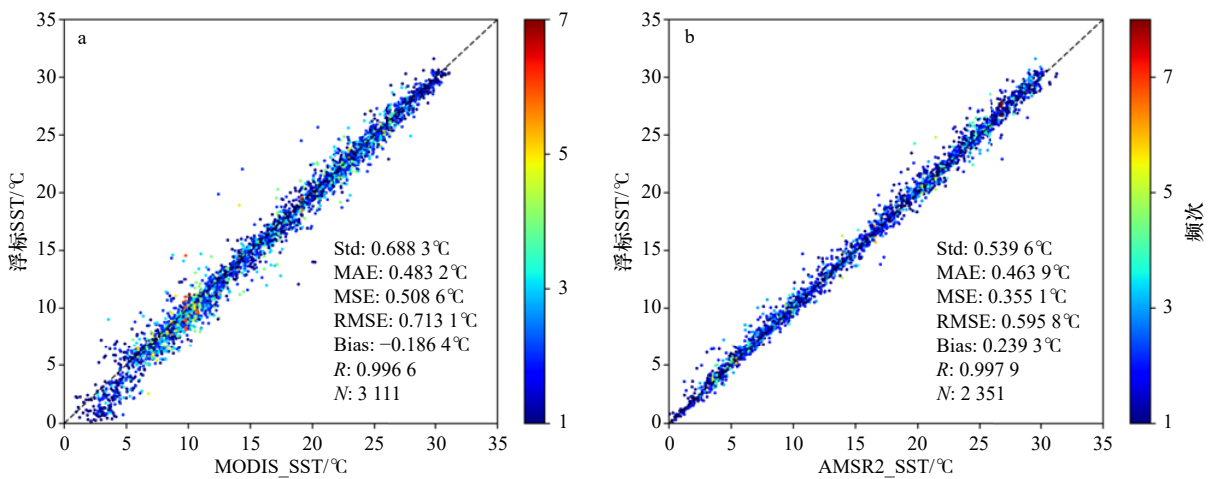


图5 浮标SST对原始MODIS_SST(a)、AMSR2_SST(b)的检验评估结果散点图(2023-2024年,N表示匹配点数)

Fig. 5 Scatter plot of the validation results of buoy SST for original MODIS_SST (a) and AMSR2_SST (b) (2023-2024, N represents the number of matching points)

比 AMSR2_SST 的 RMSE 高, 比 MODIS_SST 的 RMSE 低, 具体数值请见表 3。

3 结果与讨论

3.1 覆盖率

图 7 显示 2023 年、2024 年 MODIS_SST、AMSR2_SST 和本论文融合的 SST 的逐月覆盖率, 可知, MODIS 在冬季覆盖率较低, AMSR2 在夏季覆盖率较低。

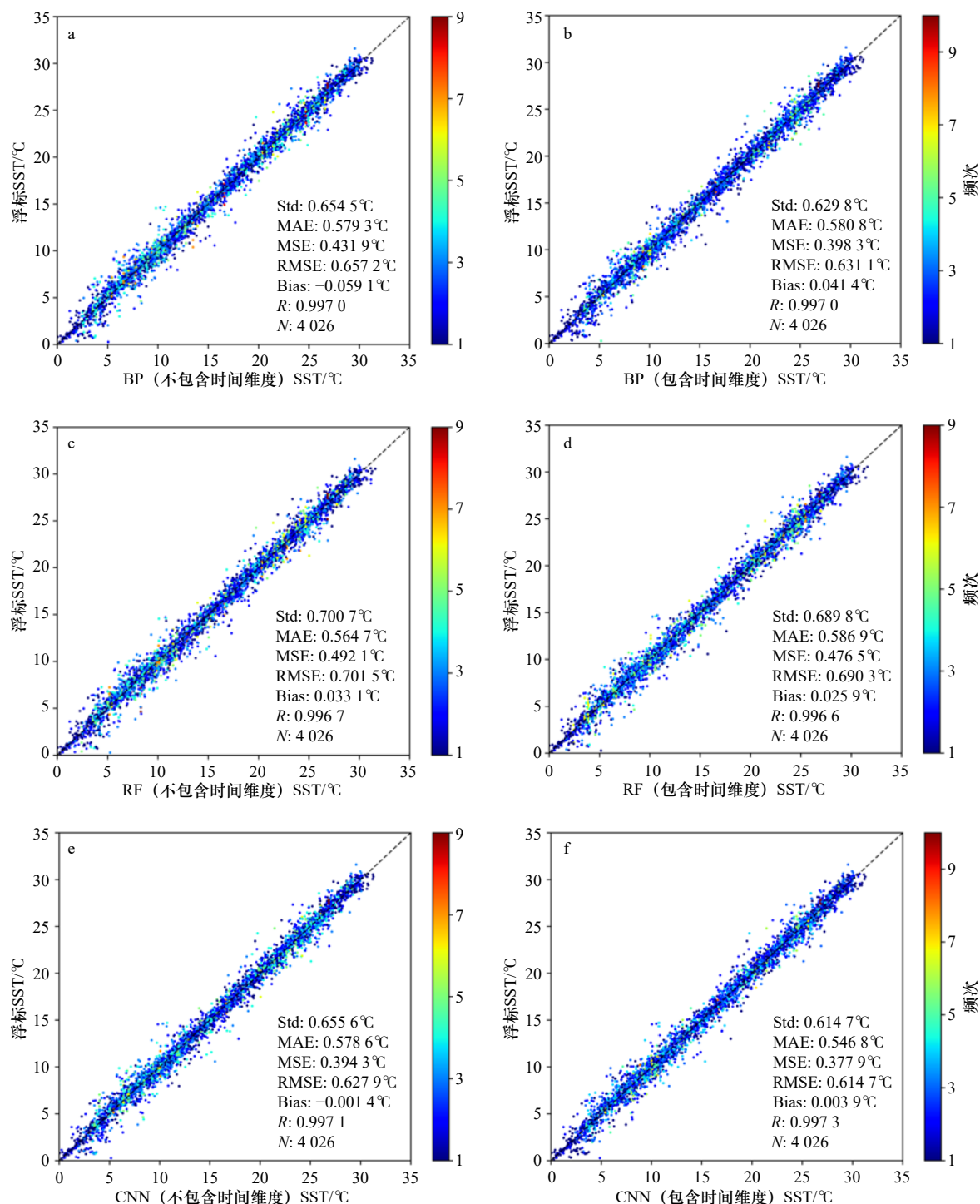


图 6 6 种融合方案 SST 的浮标检验结果散点图 (2023–2024 年, N 表示匹配点数)

Fig. 6 Scatter plot of buoy verification results for six fusion schemes SST (2023–2024, N represents the number of matching points)

这是因为, MODIS 属于光学传感器, 依赖可见光、近红外及短波红外波段探测海面信息。其信号无法穿透厚云层、海冰或大雾, 观测受天气和下垫面物理状态的直接限制。渤、黄海冬季的气候特点对光学传感器的限制显著, 主要包括海冰的广泛覆盖和高云量与低云、雾的干扰。渤、黄海 (尤其是渤海) 冬季受强冷空气影响, 气温可低至 -10°C 以下, 海面易形成海

冰, 1~2 月海冰覆盖率可达 30%~50%。海冰的高反射率会完全遮挡海水的信号, 导致 MODIS 无法获取冰区有效数据。同时, 冬季北方冷空气南下时, 易在渤、黄海形成大范围低云 (如层积云)、雾或降雪, 云层覆盖率常超过 70%。MODIS 的光学信号无法穿透厚云层, 被云层覆盖的区域会直接表现为“数据缺失”。

表3 原始 MODIS_SST 和 AMSR2_SST 以及 6 种融合方案 SST 的浮标检验结果 (2023–2024 年)
Table 3 Buoy validation results of original MODIS_SST, AMSR2_SST and SST from six fusion schemes (2023–2024)

	Std/°C	MAE/°C	MSE/°C	RMSE/°C	Bias/°C	R
MODIS_SST	0.69	0.48	0.51	0.71	-0.19	0.996 6
AMSR2_SST	0.54	0.46	0.36	0.60	0.24	0.997 9
BP_without_time_SST	0.65	0.58	0.43	0.66	-0.06	0.997 0
BP_with_time_SST	0.63	0.58	0.40	0.63	0.04	0.997 0
RF_without_time_SST	0.70	0.56	0.49	0.70	0.03	0.996 7
RF_with_time_SST	0.69	0.57	0.48	0.69	0.03	0.996 6
CNN_without_time_SST	0.66	0.58	0.39	0.63	-0.001 4	0.997 1
CNN_with_time_SST	0.61	0.55	0.38	0.61	0.003 9	0.997 3

注: Std、MAE、MSE、RMSE、Bias数据仅保留2位小数。

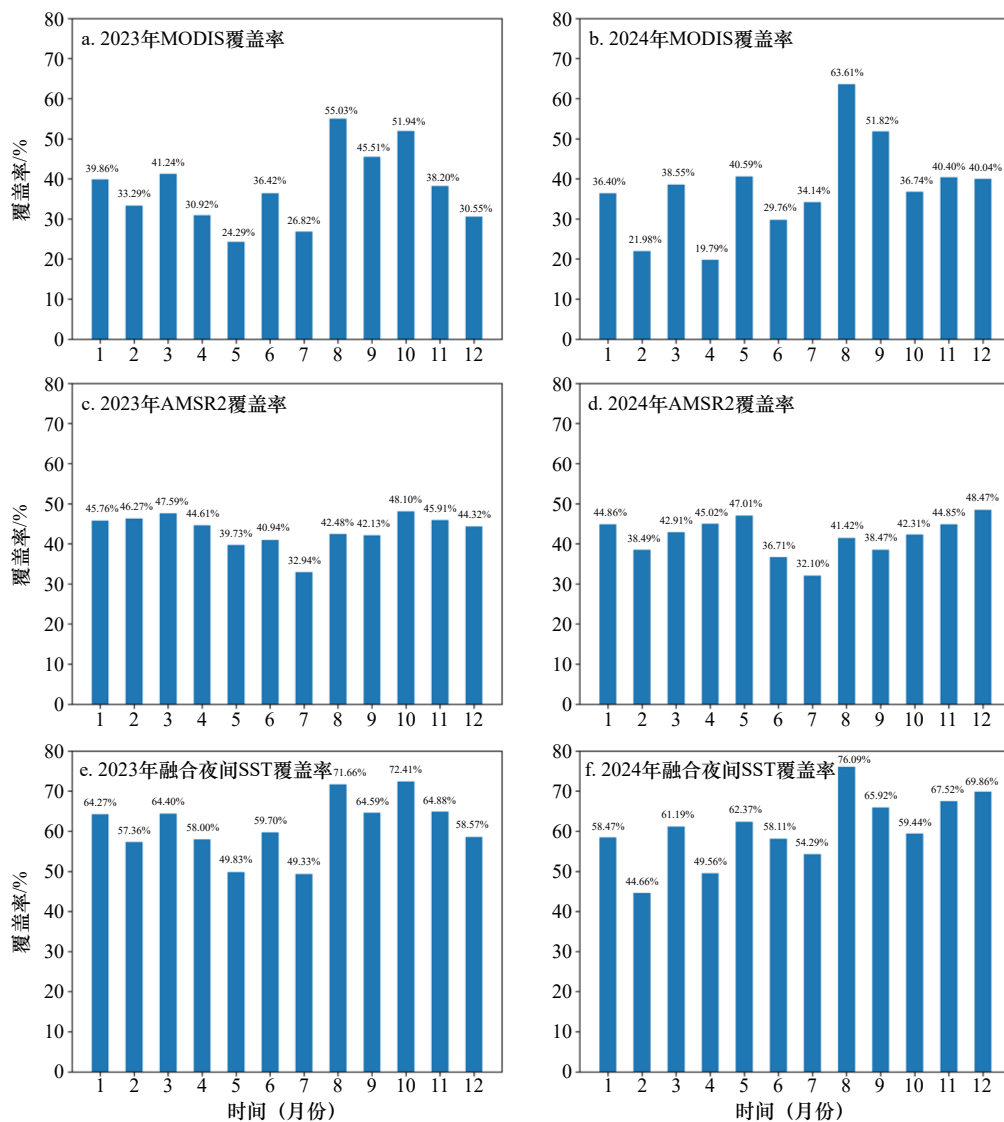


图7 MODIS_SST、AMSR2_SST 和融合 SST 三者的 2023 年和 2024 年逐月覆盖率

Fig. 7 Monthly coverage of MODIS_SST, AMSR2_SST, and fused SST for 2023 and 2024

AMSR2 属于微波传感器,通过微波波段探测。微波具有穿透云层、雾、小雨的能力,但信号易受大气水汽、海面粗糙度等因素影响(微波与这些介质的相互作用会改变辐射、散射特性)。AMSR2 在夏季覆盖率较低是因为大气水汽的强吸收作用。夏季受夏季风影响,渤、黄海区域气温高(25~30℃),海面蒸发旺盛,大气水汽含量显著增加(相对湿度常达80%以上)。AMSR2 的微波波段对水汽极为敏感,水汽会强烈吸收微波信号,导致信号衰减甚至丢失。即使微波能穿透云层,高水汽环境仍会大幅降低数据质量,使大量观测因未达质量标准而被剔除。

总体而言,MODIS_SST 的覆盖率最低,AMSR2_SST 的覆盖率较 MODIS_SST 高,融合的 SST 覆盖率最高。2023 年 MODIS_SST 覆盖率最低为 5 月份,只

有 24.29%,而融合的 SST 当月覆盖率提升为 49.83%;2024 年 MODIS_SST 覆盖率最低为 4 月份,只有 19.79%,融合的 SST 当月覆盖率提升为 49.56%。2023 年 AMSR2_SST 覆盖率最低为 7 月份,只有 32.94%,而融合的 SST 当月覆盖率提升为 49.33%;2024 年 AMSR2_SST 覆盖率最低为 7 月份,只有 32.10%,而融合的 SST 当月覆盖率提升为 54.29%。可以看出,融合 SST 可以显著提升 SST 覆盖率。

3.2 空间分辨率

融合的 SST 除了可以提高原始单独 SST 的覆盖率,还可以提高 AMSR2_SST 的空间分辨率,图 8 至图 11 是原始 MODIS_SST、AMSR2_SST 以及融合 SST 示例,分别选取 2024 年春、夏、秋、冬 4 个季节的其中 1 天 SST 为代表,融合模型选择的是包含时间输入

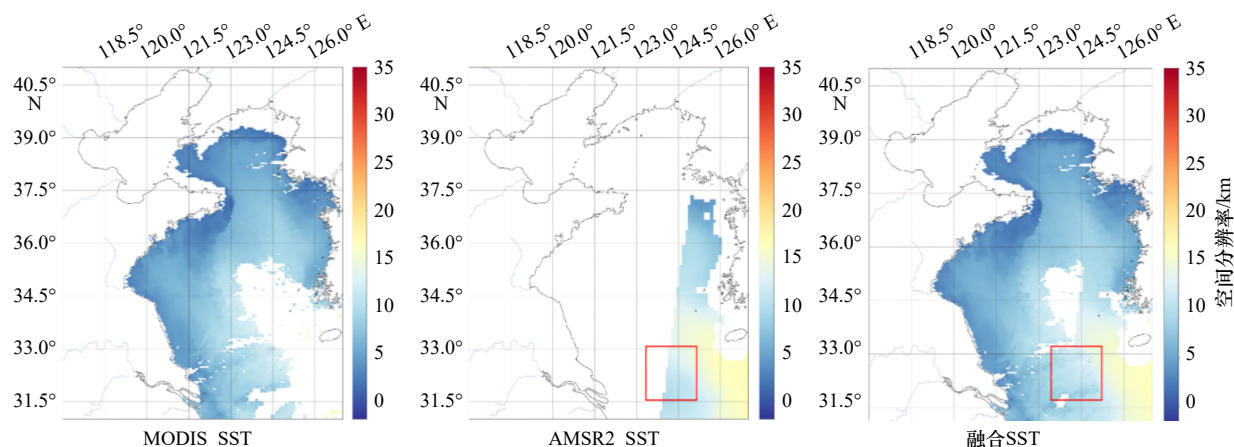


图 8 2024 年 2 月 10 日的 MODIS_SST、AMSR2_SST 和融合 SST 对比

Fig. 8 Comparison of MODIS_SST, AMSR2_SST, and fused SST on February 10, 2024

红框处表示融合 SST 显示出更丰富的空间分布变化特征

The red box indicates that the fused SST exhibits richer characteristics of spatial distribution variation

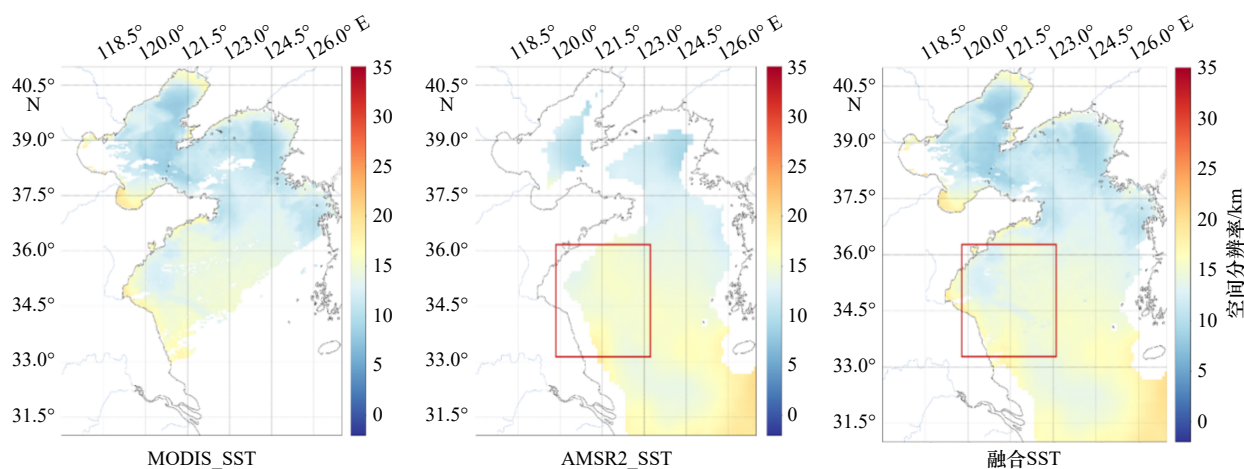


图 9 2024 年 5 月 12 日的 MODIS_SST、AMSR2_SST 和融合 SST 对比

Fig. 9 Comparison of MODIS_SST, AMSR2_SST, and fused SST on May 12, 2024

红框处表示融合 SST 显示出更丰富的空间分布变化特征

The red box indicates that the fused SST exhibits richer characteristics of spatial distribution variation

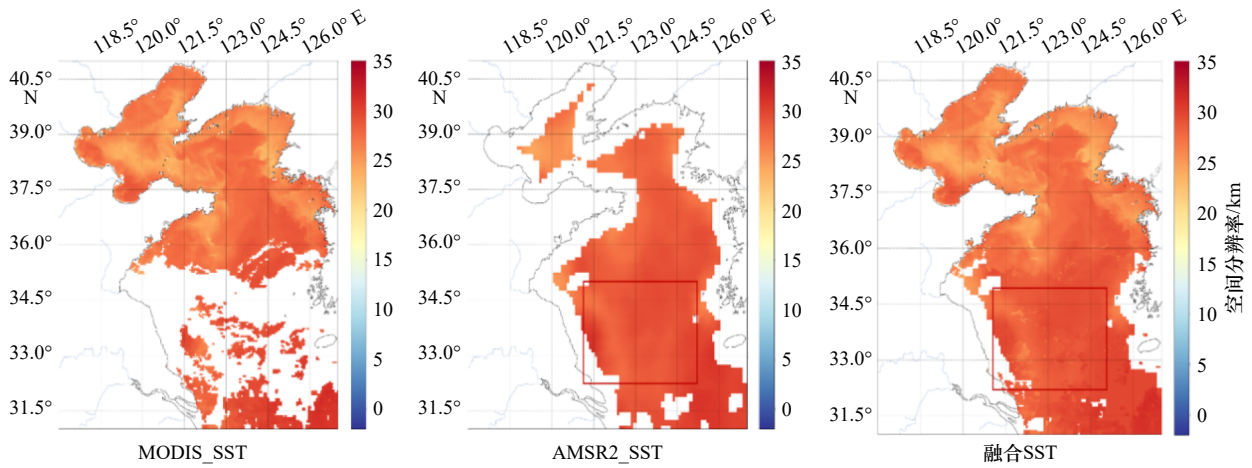


图 10 2024 年 8 月 28 日的 MODIS_SST、AMSR2_SST 和融合 SST 对比

Fig. 10 Comparison of MODIS_SST, AMSR2_SST, and fused SST on August 28, 2024

红框处表示融合 SST 显示出更丰富的空间分布变化特征

The red box indicates that the fused SST exhibits richer characteristics of spatial distribution variation

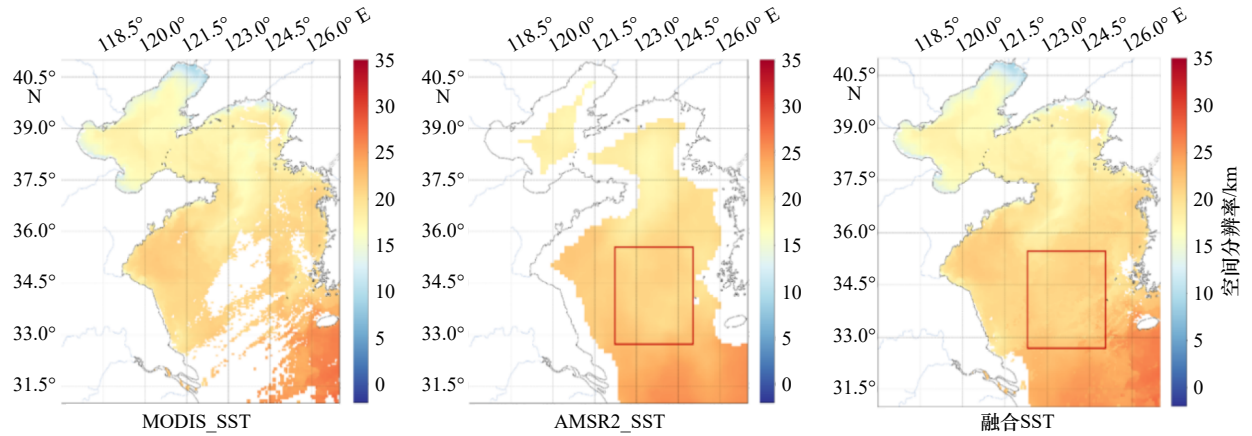


图 11 2024 年 10 月 24 日的 MODIS_SST、AMSR2_SST 和融合 SST 对比

Fig. 11 Comparison of MODIS_SST, AMSR2_SST, and fused SST on October 24, 2024

红框处表示融合 SST 显示出更丰富的空间分布变化特征

The red box indicates that the fused SST exhibits richer characteristics of spatial distribution variation

的 CNN 回归模型。由于提高了 AMSR2_SST 的分辨率(从原始的 AMSR2_SST 10 km 的空间分辨率提升到融合 SST 的 4 km 空间分辨率),融合 SST 显示出更丰富的空间分布变化特征(见图 8 至图 11 红框)。

3.3 精度评价

表 3 显示了原始 MODIS_SST 和 AMSR2_SST 以及 6 种融合方案 SST 的浮标检验结果,发现 6 种融合方案中 CNN_withtime 与浮标对比精度最高,但是仍然比原始 AMSR2_SST 与浮标对比的结果要差, RMSE 分别为 0.614 7°C 和 0.585 8°C。为了讨论本研究融合 SST 是否有意义,将研究区域内直接合并的 SST 与浮标进行对比验证(图 12)。可以发现,直接将 MODIS SST 和 AMSR2 SST 合并得到的 SST 与浮标进行对比的 RMSE 为 0.648 1°C,明显高于融合的 SST。相比于 MODIS_

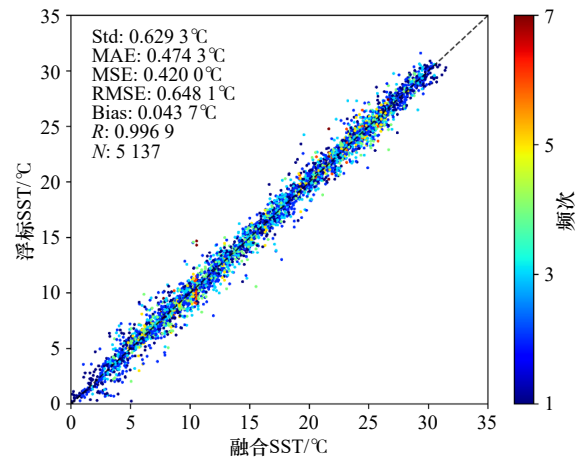


图 12 研究区域内直接合并的 SST 与浮标对比

Fig. 12 Comparison of directly merged SST and buoy data within the study area

SST, 融合后的 SST 提升了精度和覆盖率; 相比于 AMSR2_ SST, 融合后的 SST 提升了空间分辨率和覆盖率。

参考文献:

- [1] Martin S. 海洋遥感导论[M]. 蒋兴伟, 译. 北京: 海洋出版社, 2008.
Martin S. An Introduction to Ocean Remote Sensing[M]. Jiang Xingwei, trans. Beijing: Ocean Press, 2008.
- [2] Vytla V, Baduru B, Kolukula S S, et al. Forecasting of sea surface temperature using machine learning and its applications[J]. *Journal of Earth System Science*, 2025, 134(1): 25.
- [3] Lau N C, Nath M J. Impact of ENSO on SST Variability in the North Pacific and North Atlantic: seasonal dependence and role of extratropical sea-air coupling[J]. *Journal of Climate*, 2001, 14(13): 2846–2866.
- [4] Mellin C, Brown S, Heron S F, et al. *CoralBleachRisk*—global projections of coral bleaching risk in the 21st century[J]. *Global Ecology and Biogeography*, 2025, 34(2): e13955.
- [5] Oliver E C J, Benthuisen J A, Bindoff N L, et al. The unprecedented 2015/16 Tasman Sea marine heatwave[J]. *Nature Communications*, 2017, 8(1): 16101.
- [6] Chen T C, Kahru M, Landry M R, et al. Multi-trophic level responses to marine heatwave disturbances in the California current ecosystem[J]. *Ecology Letters*, 2024, 27(12): e14502.
- [7] Iglesias I S, Fiechter J, Santora J A, et al. Vertical distribution of mesopelagic fishes deepens during marine heatwave in the California current[J]. *ICES Journal of Marine Science*, 2024, 81(9): 1837–1849.
- [8] 胡秋良. 渤海海温、流季节变异研究及对海上军事活动的影响[D]. 青岛: 中国海洋大学, 2005. Doi: [10.7666/d.y828432](https://doi.org/10.7666/d.y828432).
Hu Qiuliang. Temperature and current seasonal variation of the Bohai Sea and Huanghai Sea and its effect of operation on sea[D]. Qingdao: Ocean University of China. Doi: [10.7666/d.y828432](https://doi.org/10.7666/d.y828432).
- [9] Roemmich D, Johnson G C, Riser S, et al. The Argo program: observing the global ocean with profiling floats[J]. *Oceanography*, 2009, 22(2): 34–43.
- [10] Reynolds R W, Rayner N A, Smith T M, et al. An improved in situ and satellite SST analysis for climate[J]. *Journal of Climate*, 2002, 15(13): 1609–1625.
- [11] Guan Lei, Kawamura H. SST availabilities of satellite infrared and microwave measurements[J]. *Journal of Oceanography*, 2003, 59(2): 201–209.
- [12] Dommenget D. An objective analysis of the observed spatial structure of the tropical Indian Ocean SST variability[J]. *Climate Dynamics*, 2011, 36(11/12): 2129–2145.
- [13] 张大明, 许东峰, 章本照, 等. 最优插值法及其在热带太平洋海表温度数据同化中的应用[J]. *海洋学研究*, 2005, 23(4): 1–7.
Zhang Daming, Xu Dongfeng, Zhang Benzao, et al. Optimal interpolation and its application to assimilation of SST data in the Tropic Pacific Ocean[J]. *Journal of Marine Sciences*, 2005, 23(4): 1–7.
- [14] Waters J, Lea D J, Martin M J, et al. Implementing a variational data assimilation system in an operational 1/4 degree global ocean model[J]. *Quarterly Journal of the Royal Meteorological Society*, 2015, 141(687): 333–349.
- [15] Guo Peng. Study on Bayesian hierarchal model-based SST data fusion methods[C]//Remote Sensing of the Ocean, Sea Ice, and Large Water Regions 2010. Toulouse, France: SPIE, 2010, 7825: 78250O. Doi: [10.1117/12.864912](https://doi.org/10.1117/12.864912).
- [16] Dong Wanqiu, Han Guijun, Li Wei, et al. A comparative evaluation of two bias correction approaches for SST forecasting: data assimilation versus deep learning strategies[J]. *Remote Sensing*, 2025, 17(9): 1602.
- [17] Ma Yuanzhe, Xie Bowen, Feng Zhongkun, et al. A deep learning-based hybrid model for improved SST prediction in the tropical Pacific Ocean[J]. *Journal of Oceanology and Limnology*, 2025, 43(6): 1709–1725.
- [18] 袁本坤, 黄蕊, 商杰, 等. 基于岸基观测数据的渤海沿岸海域表层温盐特征分析[J]. *海洋开发与管理*, 2015, 32(12): 31–34.
Yuan Benkun, Huang Rui, Shang Jie, et al. Analysis of sea surface temperature and salinity based on coastal observed data in the Bohai Sea[J]. *Ocean Development and Management*, 2015, 32(12): 31–34.
- [19] 鲍献文, 万修全, 高郭平, 等. 渤海、黄海、东海 AVHRR 海表温度场的季节变化特征[J]. *海洋学报*, 2002, 24(5): 125–133.
Bao Xianwen, Wan Xiuquan, Gao Guoping, et al. The characteristics of the seasonal variability of the sea surface temperature field in the Bohai Sea, the Huanghai Sea and the East China Sea from AVHRR data[J]. *Haiyang Xuebao*, 2002, 24(5): 125–133.
- [20] Qin Huiling, Chen Guixing, Wang Weiqiang, et al. Validation and application of MODIS-derived SST in the South China Sea[J]. *International Journal of Remote Sensing*, 2014, 35(11/12): 4315–4328.
- [21] Gentemann C L, Akella S. Evaluation of NASA GEOS-ADAS Modeled Diurnal Warming Through Comparisons to SEVIRI and AMSR2 SST Observations[J]. *Journal of Geophysical Research: Oceans*, 2018, 123(2): 1364–1375.
- [22] Jung S, Yoo C, Im J. High-resolution seamless daily sea surface temperature based on satellite data fusion and machine learning over Kuroshio Extension[J]. *Remote Sensing*, 2022, 14(3): 575.
- [23] Shi Xinjie, Duan Boheng, Ren Kaijun. A more accurate field-to-field method towards the wind retrieval of HY-2B scatterometer[J]. *Remote Sensing*, 2021, 13(12): 2419.
- [24] Cornford D, Nabney I T, Ramage G. Improved neural network scatterometer forward models[J]. *Journal of Geophysical Research:*

- [Oceans](#), 2001, 106(C10): 22331–22338.
- [25] Harbola S, Coors V. One dimensional convolutional neural network architectures for wind prediction. *Energy Conversion and Management*, 2019, 195: 70–75.
- [26] Kattenborn T, Leitloff J, Schiefer F, et al. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2021, 173: 24–49.
- [27] Breiman L. Random forests. *Machine Learning*, 2001, 45(1): 5–32.
- [28] Breiman L. Bagging predictors[J]. [Machine Learning](#), 1996, 24(2): 123–140.

Machine learning-based fusion technique for sea surface temperature in the Bohai and Yellow Seas

Zhang Jie, Lin Zhijia, Cai Wenbo

(National Marine Environmental Forecasting Center, Beijing 100081, China)

Abstract: This study is based on sea surface temperature (SST) data from MODIS and AMSR2 satellite observations. Three machine learning models—backpropagation neural network (BPNN), random forest (RF), and convolutional neural network (CNN)—were constructed to conduct research on SST data fusion technology. In terms of model input design, two differentiated schemes are proposed: the basic scheme includes only latitude, longitude, and raw SST data, while the enhanced scheme introduces a time parameter, resulting in six fusion schemes—BP_without_time, BP_with_time, RF_without_time, RF_with_time, CNN_without_time, and CNN_with_time. Experimental test results show that among the three machine learning models, CNN demonstrates the most outstanding performance, while the RF model performs relatively weakly. In comparative tests of the three models, the enhanced schemes incorporating time parameters significantly outperform the basic schemes without time parameters. Validation results based on 2023–2024 buoy measurement data indicate that the accuracy of the fused SST data is slightly lower than that of AMSR2_SST, but shows a significant improvement compared to MODIS_SST. The monthly coverage of the fused data has been significantly improved compared to the original data. The minimum coverage for 2023–2024 increased from 19.79% of MODIS and 32.10% of AMSR2 to over 49.56%. Additionally, the high-resolution fused results can capture more detailed temperature distribution characteristics, providing richer spatial details compared to the 10 km resolution AMSR2 data.

Key words: sea surface temperature fusion; machine learning; BPNN; RF; CNN