

研究论文

海表面盐度的高精度预测模型

王颖超¹, 柳青青^{2*}, 李洪平¹, 赵红³

(1. 中国海洋大学 信息科学与工程学院, 山东 青岛 266100; 2. 青岛大学 商学院, 山东 青岛 266100;

3. 中国海洋大学 数学科学学院, 山东 青岛 266100)

摘要: 为了建立高精度的海洋表面盐度预测模型, 采用 BP 神经网络的方法, 针对 SMOS 卫星 level 1C 级亮度温度数据和辅助数据建立了一种海表面盐度预测模型, 以 ARGO 浮标观测值作为海表盐度实测值来检验新模型预测结果的准确度, 同时利用验证集对模型的精度进行验证。结果表明: 通过新模型预测的海表盐度 (SSS₀) 比 SMOS 卫星的 3 个粗糙度模型盐度产品 (SSS1, SSS2, SSS3) 精度高; SSS₀, SSS1, SSS2, SSS3 与 ARGO 浮标实测盐度 (SSS_{ARGO}) 的均方根误差分别为 0.847 3, 2.041 7, 2.028 8 和 2.080 5, 平均绝对误差分别为 0.755 3, 1.422 6, 1.421 6 和 1.456 6, SSS₀ 与 SSS_{ARGO} 的均方根误差和绝对平均误差值都明显小于 SSS1, SSS2 和 SSS3 与 SSS_{ARGO} 的; 由此可见, 建立的海表盐度预测模型精度较高。新模型为海表盐度的反演算法提供了新思路。

关键词: 海表盐度; BP 神经网络; SMOS 卫星; ARGO 浮标

中图分类号: P731.11

文献标识码: A

文章编号: 1671-6647(2021)01-0037-08

doi: 10.3969/j.issn.1671-6647.2021.01.005

引用格式: WANG Y C, LIU Q Q, LI H P, et al. High-precision prediction model for sea surface salinity[J]. Advances in Marine Science, 2021, 39(1): 37-44. 王颖超, 柳青青, 李洪平, 等. 海表面盐度的高精度预测模型[J]. 海洋科学进展, 2021, 39(1): 37-44.

海表面盐度 (Sea Surface Salinity, SSS) 在研究海洋循环、气候变化和全球水循环中均发挥着不可替代的作用, 因此研究 SSS 至关重要^[1]。2009 年, 欧洲航天局发射了 SMOS (Soil Moisture and Ocean Salinity) 卫星, 该卫星是一颗专门用来研究土壤湿度与海洋盐度的卫星^[2-3]; 2011 年, 美国国家宇航局和阿根廷国家空间活动委员会共同发射了 Aquarius/SAC-D 卫星, 为海表盐度的进一步研究提供了重要数据源^[4-5]。但是, 目前这些卫星已经达到使用年限。我国将计划于 2022 年发射海表盐度卫星, 为我国海表盐度卫星的研究提供新思路。

SMOS 卫星搭载的 L 波段微波辐射计, 能够实现对海表盐度和土壤湿度的大范围、全天候实时观测^[6]。为了获取高精度海表盐度, 学者们对其做了大量的研究和评估。例如: Kolodziejczyk 等^[7]提出了一种降低 SMOS 系统误差的方法, 通过调整驻留线和轨道方向之间的 SSS 相对变化, 提高了海表盐度的探测精度; 而 Boutin 等^[8]在 2018 年纠正了他的方法, 其修正的靠近陆地的 SMOS 海表盐度与 SMAP 测量的海表盐度更加一致; 金旭辰等^[9]基于 SMOS 数据, 采用广义相加模型和偏最小二乘法分析了水温对海表盐度的影响, 在平均水温约 16 °C 时的 SSS 均方根误差 (Root Mean Square Error, RMSE) 约为 0.9, 在平均水温 23 °C 时的 RMSE 约为 0.7, 在平均水温 30 °C 时的 RMSE 约为 0.4; Boutin 等^[10]针对远离海岸的宽阔海域, 将 SMOS 卫星的升轨盐度数据与 ARGO 浮标盐度数据进行比较, 结果表明, SMOS 卫星在宽阔海域的精度可以达到 0.3~0.5; Yin 等^[11]通过修正正演模型, 将风场和降雨的影响最小化, 从而提高盐度的反演精度; Zhao 等^[12]

收稿日期: 2019-03-29

资助项目: 国家自然科学基金项目——稳健主成分回归的数值方法研究 (11871444)

作者简介: 王颖超 (1994—), 女, 硕士, 主要从事卫星海洋遥感方面研究。E-mail: wangyingchao0509@163.com

* **通信作者:** 柳青青 (1984—), 女, 讲师, 主要从事卫星海洋遥感方面研究。E-mail: 1440689632@qq.com

(王 燕 编辑)

利用多元线性回归的方法来预测海表盐度,其预测的盐度值与 WOA13 的 RMSE 值为 0.21,提高了海表盐度的观测精度。

机器学习相关算法的应用领域越来越广泛,神经网络、K-means 算法和遗传算法等也被引用于海表盐度的反演模型中。BP(Back Propagation)神经网络作为人工神经网络中的一种,其计算过程由正向和反向组成,通过不断的迭代将误差降为最低,其优点较多,自适应性、自组织性、良好的学习能力、良好的联想能力、良好的容错能力和抗干扰能力。同时,BP 神经网络在理论和性能方面都已经成熟,并具有很强的非线性映射能力。除此之外,其网络结构具有可调节性,可根据研究的实际情况对隐藏层的数量进行调节,以达到研究的预期目标^[13]。

本文综合考虑海表面温度、亮温(包括水平和垂直两个方向的数据值)、波高、主波长、蒸发量和风速等因素对海表盐度的影响,采用 BP 神经网络方法建立了一种新型的海表盐度预测模型。对比分析本模型计算所得的海表盐度值(SSS0)和 SMOS 卫星 level 2(简称 L2)级的 3 个粗糙度模型盐度产品(SSS1, SSS2, SSS3)与 ARGO 实测盐度值(SSS_{ARGO}),以期为发射的海表盐度卫星提供技术储备。

1 数 据

1.1 SMOS 数据

本文使用 2 种 SMOS 卫星数据,即 L1C 级别的亮度温度数据和欧洲中期预报中心提供的辅助数据(ECMWF 数据),选取 2015 年 5 月 12—31 日的南半球海域卫星数据,入射角选择 $42^{\circ}\sim 43^{\circ}$ 。

1.2 ARGO 数据

ARGO 浮标采用一种沉浮式的自律式拉格朗日环流剖面观测浮标,测量海水温度、海水盐度和海流参数等多种海洋数据。本文利用 ARGO 浮标数据第一层(5 m)的盐度值,为了与 SMOS 数据进行时间与空间上的匹配,本文同样也使用 2015 年 5 月南半球的海表盐度数据。数据来源于网站 <http://www.ARG0.org.cn/>。此数据为经过国际上普遍认可的对 ARGO 数据处理模式,处理后可直接应用^[14]。

1.3 数据预处理

对于 SMOS L1C 级数据,需要通过 SMOS 卫星 L1C 级数据反演至 L2 级盐度数据所使用的数据处理软件 L2OS(L2 Ocean Salinity Operational Processor)进行处理,输出水滴密集度、主波长和海表温度等参数信息。软件主要包括预处理程序、主程序、后置处理和正演模型四部分。其中,预处理程序主要用于反演前对亮温数据进行异常值检测,去除受到射频干扰、太阳反辉区或者其他不明来源的外来污染的异常值,并且 L1C 级亮温数据的初始化工作也在预处理程序中进行。首先,对于下载的 Zip 文件进行解压获得“.DBL”和“.HDR”文件,该文件储存了该条带内所有网格点的编号、经纬度信息等;然后,修改软件配置文件里涉及的数据版本信息并检查数据的经纬度,生成 Job order 配置文件,每组数据对应一个配置文件;最后即生成 OTT 文件,并将 OTT 文件加入到 L2OS 中,输出参数信息。为了节约程序运行时间,只需对存在 ARGO 浮标的 SMOS 网格进行处理。此软件源于欧空局 SMOS 卫星官方网站 <https://smos.argans.co.uk/>。

1.4 匹配原则

将 SMOS 卫星 L1C 数据和 ECMWF 辅助数据与 ARGO 实测数据进行时间和空间上的匹配,并去除 NAN 无效数据。时间上,匹配的最大时间间隔为 1 天;空间上,按照 SMOS 数据的网格点进行匹配。除此之外,对同一入射角对应的不同亮温值进行求平均值处理,而对相同的天、小时、网格点数据进行标准化处理。随后,按照同样的匹配原则将其 L2 级数据进行匹配。在 2015 年 5 月 12—18 日时间范围内,共匹配有效数据 1 021 组,并将此数据集作为本次研究的建模数据;在 2015 年 5 月 19—31 日时间范围内,共匹配有

效数据 2 055 组,将此数据集作为新模型的验证数据集。

2 方 法

2.1 神经网络算法

BP(Back Propagation)神经网络是人工神经网络的一种,该网络模型主要包括三部分:输入层、隐藏层和输出层。在建模过程中,为了提高精度可以设置多个隐藏层并给其设定传递函数。BP 神经网络包括 2 个方向的传播,先是信号的正向传播(图 1a),信号到达输出层后,若网络输出值与期望值有偏差,则计算损失函数并进行误差信号的反向传播(图 1b)。由于层与层之间通过权重和偏置来连接,因此反向传播是通过不断地迭代来调整每一层各个神经单元的权重系数和偏置值,从而使得损失函数最小,直到达到所限定的迭代次数或者规定的误差为止。

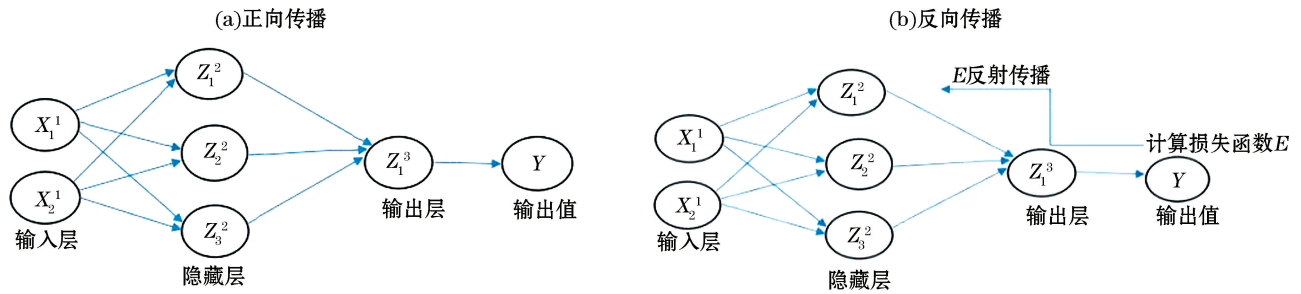


图 1 BP 神经网络组成及原理

Fig.1 Composition and schematic of BP neural network

BP 神经网络的计算公式:

$$a_i^l = \sum_{j=1}^n w_{ij}^l Z_j^{l-1} + b_i^l, \quad (1)$$

$$Z_j^l = f(a_i^l), \quad (2)$$

式中, a_i^l 中 l 为其所在的层数, i 为其所在层数的第 i 个神经元; j 为其所在层数的第 j 个神经元; n 为所在层数的神经元节点总数; w_{ij}^l 为权重(第 $l-1$ 层第 j 个神经元到第 l 层第 i 个神经元之间的连接权重), $w_{ij}^l = w_{ij}^l - \frac{\mu}{N} \sum_{k=1}^N \frac{\partial E_{(k)}}{\partial w_{ij}^l}$, 此处采用梯度下降法对权值和偏置进行更新, μ 为学习率(移动步长), N 为总层数, k 为

对应的层数, $E_{(k)}$ 为损失函数(设置 1/2 是为了后续方便计算), $E_{(k)} = \frac{1}{2} \sum_{i=1}^n (y_i^k - Z_i^k)^2$, 其中 y_i 为期望输出, Z_i 为相应神经元对应的模型输出值, 由于输出值可以是多维的, 因此 i 表示对应的第 i 个输出值(假设有 N 组数据输出值是 n 维的, y_i 表示输出值中的第 i 个神经元的期望值, Z_i 代表输出值中的第 i 个神经元的网络输出值); Z_j^l 为第 l 层的第 j 个神经元的激活值; b_i^l 为偏置(第 l 层第 i 个神经元之间的偏置), $b_i^l =$

$$b_i^l - \frac{\mu}{N} \sum_{k=1}^N \frac{\partial E_{(k)}}{\partial b_i^l}; f(\cdot) \text{ 表示每一层的传递函数。}$$

2.2 模型建立

利用匹配的 1 021 组数据进行建模, 由于不同参数的数值有很大差异, 所以为了避免大数据淹没小数据的情况发生, 利用 mapminmax 对数据进行归一化处理。经过不断的试验, 最终选定了隐藏层为 5 层, 且确定正向传播时每一层的神经元节点数分别为 9, 5, 5, 4 和 3, 对应的每一隐藏层的传递函数分别为 tansig, tansig, tansig, tansig 和 purelin, 训练函数选择 trainrp 函数。除此之外, 我们将模型的最大训练次数设为

1 000次,目标精度设为 0.000 01,学习率为 0.000 01。该模型训练 1 021 组数据用时 4 s,迭代 602 次,利用该模型得到的权值和偏置矩阵如表 1 所示。

表 1 模型输出的各层权值和偏置
Table 1 Weight and offset of the model output

权 重										偏 置		
$\mathbf{W}_1 =$	-0.347	-0.321	2.295	-0.319	0.914	-0.088	0.504			$\mathbf{b}_1 =$	1.995	
	3.225	0.261	-0.71	-2.117	-0.149	-1.1805	-0.763				-1.251	
	0.476	-0.830	1.240	-0.155	0.175	-0.750	-0.757				-0.774	
	-0.332	0.822	-1.472	-0.261	0.867	0.044	-2.184				-0.273	
	0.602	-0.661	-2.410	-0.971	0.916	-0.291	-0.793				-0.034	
	-1.752	4.581	5.506	-0.230	0.230	4.632	1.227				-0.742	
	-0.018	-0.083	-0.364	-1.669	0.622	3.621	0.536				-1.241	
	0.353	0.346	-0.001	0.812	0.071	-1.700	-1.744				1.768	
	1.536	0.519	-3.304	-1.877	0.566	1.009	-0.302				2.192	
$\mathbf{W}_2 =$	-0.940	0.091	0.560	-0.741	0.786	-0.675	-0.437	0.209	-0.050	$\mathbf{b}_2 =$	1.516	
	0.840	0.044	0.673	0.631	-1.454	-0.684	-0.654	0.715	0.272		-0.783	
	-0.823	0.384	-0.746	0.466	1.205	0.671	0.391	-0.360	-0.047		-0.048	
	-0.281	-0.101	0.635	-0.004	1.321	0.373	-0.050	1.441	-0.330		-0.250	
	-0.084	-0.950	0.189	-0.103	0.740	0.730	0.599	-0.435	0.705		1.566	
$\mathbf{W}_3 =$	0.001	1.432	-1.662	-0.425	0.965	$\mathbf{b}_3 =$					3.055	
	0.669	1.467	-1.462	-1.747	1.396						2.046	
	-1.330	1.026	-0.037	0.876	-0.252						0.196	
	-0.322	0.502	-1.286	-0.221	-1.213						-0.877	
	-1.614	0.488	-0.606	-11.547	-3.399						-1.956	
$\mathbf{W}_4 =$	0.674	1.271	1.152	1.063	0.093	$\mathbf{b}_4 =$					-1.687	
	-1.502	-1.057	-0.579	-0.785	0.545						0.182	
	0.057	0.342	1.166	0.820	1.322						0.627	
	-0.907	-0.241	1.586	2.258	-0.579						-1.514	
$\mathbf{W}_5 =$	-0.612	-0.275	-0.764	0.606	$\mathbf{b}_5 =$							0.202
	0.353	0.470	0.100	0.523								0.400
	0.723	0.692	-0.877	0.272								-0.261
$\mathbf{W}_6 = [-0.153 \quad 0.246 \quad 0.241]$										$\mathbf{b}_6 = [0.358]$		

3 模型验证与误差分析

利用匹配的 2 055 组数据作为模型的验证数据集,并按照与本文中模型建立相同的数据处理方法对验证数据集进行处理。利用式(1)、式(2)及表 1 中模型输出的权值和偏置矩阵进行计算。可得南半球海表盐度(SSS0)和 SSS3 以及 ARGO 的盐度值 SSS_{ARGO},将其对比(图 2)(随机抽取 1 000 组数据展示),结果表明,SSS0 与 ARGO 实测值更为接近,精度更高。

在模型验证部分,本文分别计算了 SMOS 的 3 个粗糙度模型盐度产品(SSS1,SSS2,SSS3)和本模型(SSS0)与 ARGO 实测盐度值 SSS_{ARGO}的误差,结果表明,SSS1,SSS2,SSS3 与 SSS_{ARGO}的误差值在[-1,1]

的数据分别占总数的 47%,49%和 49%;SSS0 与 SSS_{ARGO} 的误差值在 $[-1,1]$ 的数据约占总数的 80%;由此可知,本模型极大提高了海表盐度的预测精度。进一步计算了 SSS0,SSS1,SSS2,SSS3 与 SSS_{ARGO} 的均方根误差(RMSE)和平均绝对误差(Mean Absolute Error,MAE),如表 2 所示,结果表明:SSS0 与 SSS_{ARGO} 的 RMSE 和 MAE 都远小于其他 3 个盐度产品。

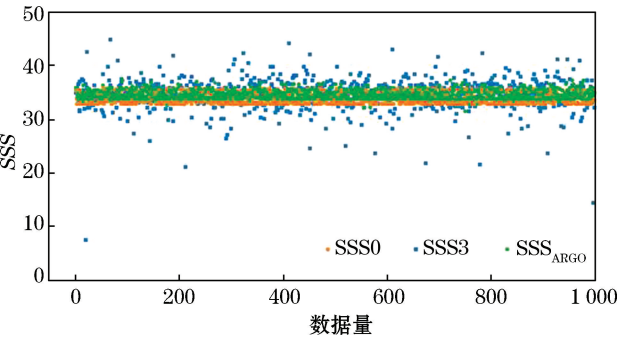


图 2 ARGO 实测盐度数据与 SSS0,SSS3 的对比
Fig.2 Comparison between SSS0, SSS3 and SSS_{ARGO}

表 2 SSS0, SSS1, SSS2, SSS3 与 ARGO 实测盐度值(SSS _{ARGO})的误差				
Table 2 The error of SSS0, SSS1, SSS2, SSS3 and ARGO (SSS _{ARGO}) measured salinity value				
误 差	SSS0 与 SSS _{ARGO}	SSS1 与 SSS _{ARGO}	SSS2 与 SSS _{ARGO}	SSS3 与 SSS _{ARGO}
RMSE	0.847 3	2.041 7	2.028 8	2.080 5
MAE	0.755 3	1.422 6	1.421 6	1.456 6

由于 SMOS 的 3 个粗糙度模型得出的盐度值极为相似,因此将 SSS3 作为 SMOS 粗糙度模型的代表盐度值。本文分别计算了 SSS0,SSS1,SSS2,SSS3 与 SSS_{ARGO} 的误差值。随机抽取验证数据中的 1 000 组数据进行分析,可见 SSS1,SSS2,SSS3 与 ARGO 浮标盐度值之间误差较大,其误差分布范围较广,而 SSS0 与 SSS_{ARGO} 误差值却集中分布在 0 附近,这表明二者之间误差较小(图 3);分析 SSS0,SSS3 与 SSS_{ARGO} 的误差,

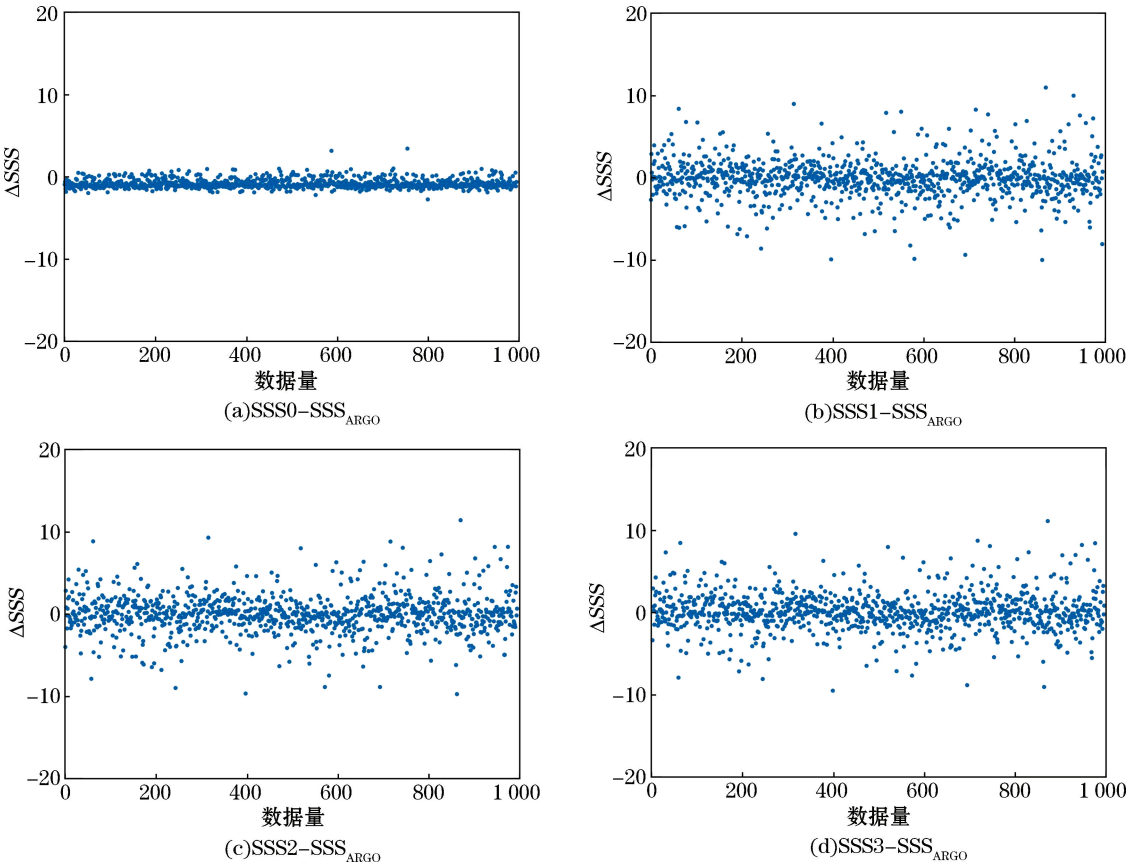


图 3 SSS0, SSS1, SSS2, SSS3 与 SSS_{ARGO} 的误差对比
Fig.3 Comparison of errors between SSS0, SSS1, SSS2, SSS3 and SSS_{ARGO}

可得其空间分布结果(图 4),选取太平洋中部($20^{\circ} \times 20^{\circ}$)海域;统计 SSS0, SSS3 与 SSS_{ARGO} 的误差(图 5)可知, SSS0 与 SSS_{ARGO} 的误差值多集中在 0 附近,而 SSS3 与 SSS_{ARGO} 的误差值分布较广,该结果表明 SSS0 与 SSS_{ARGO} 误差最小。

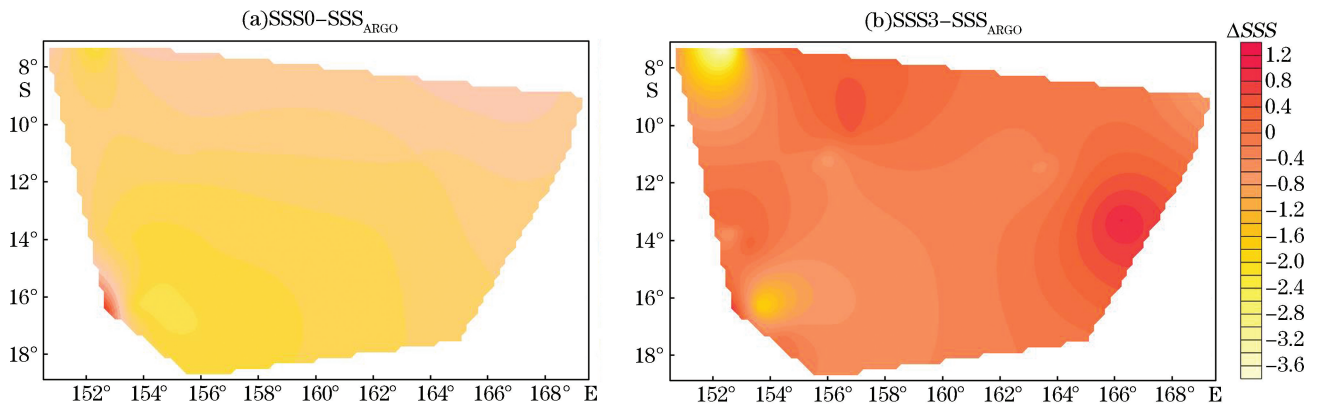


图 4 SSS0, SSS3 与 SSS_{ARGO} 的误差空间分布

Fig.4 The spatial distribution of the error between SSS0, SSS3 and SSS_{ARGO}

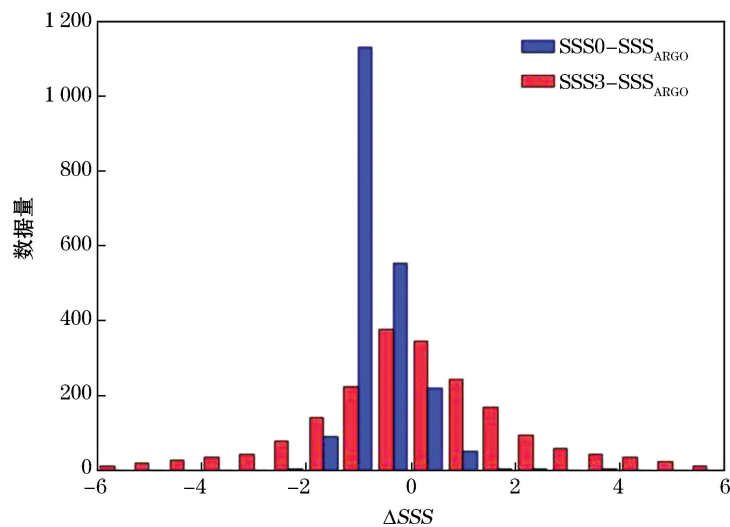


图 5 SSS0, SSS3 与 SSS_{ARGO} 的误差分布

Fig.5 The distribution of the error between SSS0, SSS3 and SSS_{ARGO}

4 结 语

海表盐度在研究海洋循环、气候变化和全球水循环中都起着不可取代的作用,由于海表盐度观测受诸多因素的干扰且数据处理过程较为繁琐,难以得到十分精确的数据结果,因此提高海表盐度的预测精度尤为关键。本文通过 BP 神经网络的方法对海表盐度进行建模,得到了一个全新的海表面盐度预测模型。本模型与 SMOS 的 3 个粗糙度模型盐度产品相比较,精度更高。

在未来的研究中,我们将继续尝试机器学习的其他方法来建立预测模型,争取建立一个精度更高的海表盐度预测模型,为我国 2022 年即将发射的海表盐度卫星做贡献。

参考文献(References):

- [1] LAGERLOEF G, COLOMB F R, VINE D L, et al. The Aquarius/SAC-D mission: designed to meet the salinity remote-sensing challenge[J]. *Oceanography*, 2008, 21(1): 68-81.
- [2] NYADIRO E S, SUBRAHMANYAM B. Spatial and temporal variability of central Indian Ocean salinity fronts observed by SMOS[J]. *Remote Sensing of Environment*, 2016, 180: 146-153.
- [3] YANG T T, CHEN Z B, HE Y J. A new method to retrieve salinity profiles from sea surface salinity observed by SMOS satellite[J]. *Acta Oceanologica Sinica*, 2015, 34(9): 85-93.
- [4] DINNAT E P, VINR D M L. Impact of sun glint on salinity remote sensing: an example with the Aquarius radiometer[J]. *IEEE Transactions on Geoscience & Remote Sensing*, 2008, 46(10): 3137-3150.
- [5] VINE D M L, LAGERLOEF G S E, COLOMB F R, et al. Aquarius: an instrument to monitor sea surface salinity from space[J]. *IEEE Transactions on Geoscience & Remote Sensing*, 2007, 45(7): 2040-2050.
- [6] LI H L, HAN Z, ZHANG Y Z, et al. Study on bright temperature gain of SMOS Satellite caused by sea surface roughness[J]. *Remote sensing information*, 2016, 31(5): 103-107. 李化良, 韩震, 张宜振. 海面粗糙度引起的 SMOS 卫星亮温增益研究[J]. *遥感信息*, 2016, 31(5): 103-107.
- [7] KOLODZIEJCZYK N, BOUTIN J, VERGELY J-L, et al. Mitigation of systematic errors in SMOS sea surface salinity[J]. *Remote Sensing of Environment*, 2016, 180: 164-177.
- [8] BOUTIN J, VERGELY J L, MARCHAND S, et al. New SMOS sea surface salinity with reduced systematic errors and improved variability[J]. *Remote Sensing of Environment*, 2018, 214: 115-134.
- [9] JIN X C, ZHU Q K, PAN D H, et al. Effects of sea surface temperature on SMOS salinity remote sensing inversion accuracy[J]. *Journal of Remote Sensing*, 2017, 21(6): 939-947. 金旭晨, 朱乾坤, 潘德炉, 等. 海表温度对 SMOS 盐度遥感反演精度的影响[J]. *遥感学报*, 2017, 21(6): 939-947.
- [10] BOUTIN J, MARTIN N, YIN X B, et al. First assessment of SMOS data over open ocean: Part II—sea surface salinity[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2012, 50(5): 1662-1675.
- [11] YIN X B, BOUTIN J, MARTIN N, et al. Errors in SMOS sea surface salinity and their dependency on a priori wind speed[J]. *Remote Sensing of Environment*, 2014, 146: 159-171.
- [12] ZHAO H, LI C J, LI H P, et al. Retrieve sea surface salinity using principal component regression model based on SMOS satellite data[J]. *Periodical of Ocean University of China*, 2016, 15(3): 399-406.
- [13] LI C J, ZHAO Q H, ZHAO H. BP neural network salinity inversion model based on SMOS satellite data[J]. *Periodical of Ocean University of China*, 2018, 48(1): 125-134. 李长军, 赵清晖, 赵红. 基于 SMOS 卫星数据的 BP 神经网络盐度反演模型[J]. *中国海洋大学学报(自然科学版)*, 2018, 48(1): 125-134.
- [14] BOUTIN J, MARTIN N, REVERDIN G, et al. Sea surface freshening inferred from SMOS and ARGO salinity: impact of rain[J]. *Ocean Science*, 2013, 9: 183-192.

High-Precision Prediction Model for Sea Surface Salinity

WANG Ying-chao¹, LIU Qing-qing², LI Hong-ping¹, ZHAO Hong³

(1. *Department of Marine Technology, Ocean University of China, Qingdao 266100, China;*

2. Business College, Qingdao University, Qingdao 266100, China;

3. School of Mathematical Sciences, Ocean University of China, Qingdao 266100, China)

Abstract: To build a high-precision ocean surface salinity prediction model, the back propagation (BP) neural network method is utilized to establish a sea surface salinity prediction model based on soil moisture and ocean salinity (SMOS) satellite level 1C brightness and temperature data and auxiliary data. The array for real-time geostrophic oceanography (ARGO) buoy observations are used as the measured value of sea surface salinity to test the accuracy of the new model's prediction results, and the verification set is used to verify the accuracy of the model. The results show that the sea surface salinity predicted by the new model (referred to as SSS0) is more accurate than the three roughness model salinity products of soil moisture ocean salinity (SMOS) satellites (referred to as SSS1, SSS2, and SSS3). The accuracy of the root mean square errors of SSS0, SSS1, SSS2, SSS3 compared to SSS_{ARGO} are 0.847 3, 2.041 7, 2.028 8 and 2.080 5, and the absolute average errors are 0.755 3, 1.422 6, 1.421 6 and 1.456 6. Both of the root mean square error and absolute average error of SSS0 are significantly smaller than SSS1, SSS2, and SSS3. Therefore, it shows that the sea surface salinity prediction model established in this paper has higher accuracy, and it provides a novel way for generating the sea surface salinity inversion algorithm.

Key words: sea surface salinity; the back propagation (BP) neural network; the soil moisture and ocean salinity (SMOS) satellite; the array for real-time geostrophic oceanography (ARGO) buoy

Received: March 29, 2019