

# 一种精细化的海洋浮标数据质量控制方法

任焕萍<sup>1,2</sup>, 张 斌<sup>1,2</sup>, 谭哲韬<sup>3,4</sup>, 李富超<sup>1,2</sup>

(1. 中国科学院海洋研究所, 山东 青岛 266071; 2. 中国科学院海洋大科学研究中心, 山东 青岛 266071; 3. 中国科学院大气物理研究所, 北京 100029; 4. 中国科学院大学, 北京 100049)

**摘要:** 针对海洋浮标数据获取特点及常见数据错情, 本文提出了一种精细化的海洋浮标温盐数据质量控制方法。通过范围检验、尖峰检验、莱因达准则等六个步骤实现浮标数据质量控制, 其创新性在于有效集成了多种质量检验法, 并且研究了一种融合传统尖峰检验与莱因达准则的质量检验新方法; 并以中国近海观测研究网络黄、东海站浮标温盐观测数据为例, 验证了本方法的有效性和适用性。该方法可广泛用于海洋浮标温盐观测数据的质量控制, 识别出长时间序列数据的异常值。经过本方法质控后的浮标观测数据对于海洋科学研究、海洋气象预报、海洋灾害预警以及渔业发展等具有重要意义。

**关键词:** 海洋浮标; 温盐; 质量控制; 莱因达准则

中图分类号: P717

文献标识码: A

文章编号: 1000-3096(2021)10-0093-11

DOI: 10.11759/hyxx20210126001

海洋浮标是一种较早开发并长期延续使用的观测设备<sup>[1]</sup>, 可以对海洋环境进行自动、连续、长期的同步监测<sup>[2]</sup>, 浮标观测数据对于海洋科学研究、海洋气象预报、海洋灾害预警以及渔业发展等具有重要意义<sup>[3]</sup>。合格的浮标观测数据是海洋科学的“生命线”。但由于种种原因, 一些原始观测数据或多或少存在精度不高、质量欠佳等问题。海洋浮标及传感器长期受恶劣的海洋环境影响, 包括海洋生物附着、海水腐蚀、盐雾侵袭、风浪袭击等, 这些因素都会影响传感器的监测结果, 造成数据异常。因此, 数据在实际使用前, 需要进行严谨而科学的质量控制(quality control, QC)<sup>[4]</sup>, 识别和剔除数据序列中的异常值, 提高数据的准确性和可用性<sup>[5]</sup>。高质量的海洋观测数据, 可以帮助人类应对全球气候变化, 提高防灾抗灾能力, 预测台风和厄尔尼诺等极端天气、海洋事件等。

海洋浮标温度、盐度资料的质量控制方法主要包括日期检验、位置检验、格式检验、范围检验、气候特征检验、尖峰检验、梯度检验、连续性检验等<sup>[6]</sup>。针对温盐观测数据, 采用范围检验可以检测出超出目标区域阈值范围的错误数据。采用统计性的质量控制方法如莱因达准则等可以检出统计分布中的离群值<sup>[7]</sup>。但是, 观测数据中还存在一种与相邻时间观测数据差异较大、但又在整体数据阈值范围内的尖峰值, 该类错误采用传统的尖峰检验法不能有效标记<sup>[8]</sup>。

不同类型的观测数据, 其异常值都有不同外观特征及性质, 很难通过普适性的异常值检测方法进行识别<sup>[9]</sup>, 因此在实际应用中需要根据不同场景具体分析、给出具体的质控方法。近年来, 在海洋温盐数据质量控制方面已有很多研究成果, 但大多是针对航次观测、Argo 漂流浮标等剖面观测手段的质量控制理论方法探讨, 而针对具有定点、长时间序列的海洋浮标实测数据的精细化质量控制方案尚有欠缺<sup>[10]</sup>。

## 1 数据质量控制方法

本文针对长时间序列的海洋浮标观测的海水温度(水温)数据和海水盐度(盐度)数据, 根据观测要素的特性以及数据本身特征, 综合运用范围检测法、尖峰检验法、莱因达准则( $3\sigma$ 法则)以及最大误差控制等方法, 提出了一种精细化的浮标数据质量控制方法, 并使用实际观测数据验证和测试了该方法的有效性和适用性。该方法以莱因达准则为核心, 通过 6 个步骤流程完成数据质量控制。整体流程步骤见图 1, 每个步骤完成后, 对检出的异常数据做不同的质控标记。

收稿日期: 2021-01-26 修回日期: 2021-04-16

基金项目: 国家重点研发计划资助(2017YFA0603200); 中国科学院十四五网络安全和信息化专项(WX145XQ07-08)

[Foundation: The National Key Research and Development Program of China, No. 2017YFA0603200; The 14th Five-year Network Security and Informatization Plan of Chinese Academy of Sciences, No. WX145XQ07-08]

作者简介: 任焕萍(1967—), 女, 山西省万荣县人, 高级工程师, 主要从事海洋科学大数据研究, 电话: 0532-82896760, E-mail: hpren@qdio.ac.cn

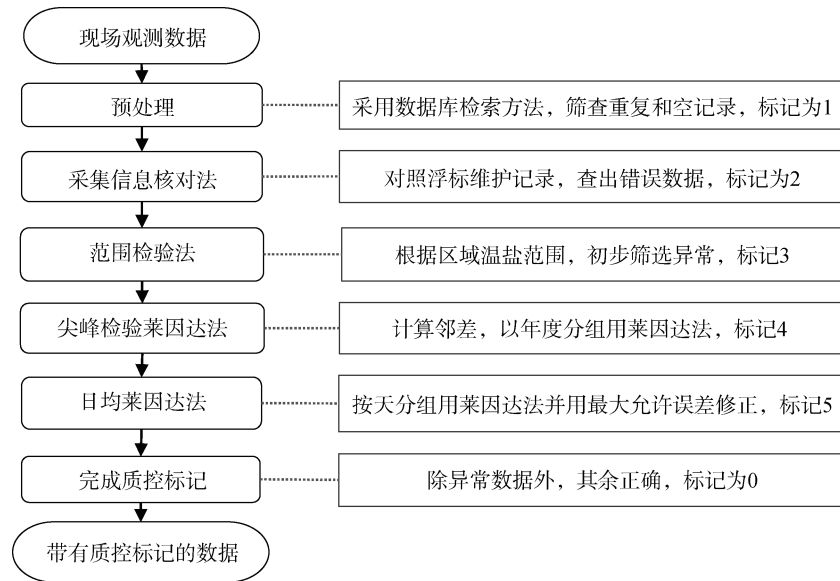


图 1 针对浮标温盐数据的精细化质量控制方案流程图  
Fig. 1 Schematic of the quality control process specific for a mooring buoy

1) 数据预处理, 主要目的是查找重复数据和空记录, 对重复记录和数值为空的记录做出标记。空记录一般是因为仪器设备故障或数据传输问题等原因, 没有采集到数据。重复记录是同一浮标在同一个时间点有 2 条或多条相同的数据。造成数据重复的主要原因是浮标早期获取设备数据时, 采用多种数据传输方式, 但在数据库存储、使用时未排除重复。

2) 采集信息核对法, 主要对照浮标的维护信息记录, 核对检查相应时间段内的温盐数据。与以往大部分仅基于数据本身的质量控制方法不同, 这里充分考虑了采集过程中观测仪器设备的检修维护等信息情况。近海浮标位于海上, 环境条件恶劣, 需要对浮标数据进行及时巡查, 对浮标搭载的传感器设备定期巡检和保养维护, 同时发现问题后需要及时对浮标进行维修, 问题严重时需要拖上岸进行大修。浮标巡检维护期间的缺失数据和错误数据, 都需要检出。对照浮标维护记录进行分析, 如果某一段时间有设备维护、安装调试、数据缺失等维护记录, 则该段时间数据不正确, 质控标记为异常数据。

例如: 对照 3 号浮标维护记录, 发现 2018 年 10 月 11 日—17 日数据巡检记录为异常, 10 月 18 日—21 日对 3 号浮标进行了维护, 完成了系统的整体安装与布放, 21 日重新正式使用, 所以将 10 月 11 日到 21 日数据均标记为异常。

3) 范围检验法, 主要目的是根据观测要素在目标区域的范围, 针对原始观测数据进行初步筛查, 对超

出区域范围外的数据, 做异常质控标记。通过历史数据、其他观测数据的分析, 确定黄东海区域海水水温范围为 0~40 °C, 黄东海区域盐度范围为 10~40<sup>[11]</sup>。如果检测出数据处于区域范围之外, 则标记为异常数据。

4) 尖峰检验莱因达法。在浮标观测数据中, 存在一种突变尖峰数据, 与相邻前后数据对比, 变化十分明显, 这种数据会在短时间内陡然增加或减小。

传统的尖峰检验法, 是对观测要素进行尖峰检验, 如果出现较大的突变, 则为异常。假设当前观测数据为  $X_i$ , 其前后邻居数据分别为  $X_{i-1}$  和  $X_{i+1}$ , 则检验值 =  $ABS[X_i - (X_{i-1} + X_{i+1})/2] - ABS[(X_{i-1} - X_{i+1})/2]$ , 如检验值超出标准范围, 则判定其为异常值。按照国家海洋信息中心发布的“海洋水文气象资料质量控制方法”规定<sup>[6]</sup>, 对于海洋观测站, 当压强 < 500 dbar 时, 如果水温检验值 > 6 °C, 则为尖峰异常; 如果盐度检验值 > 0.9, 则为尖峰异常。采用传统的尖峰检验法, 能检出检验值比较大的部分尖峰值, 由于标准范围较大, 导致很大一部分的尖峰值不能有效检出。

针对这个问题, 本文创新性地提出“尖峰检验莱因达法”, 将传统方法中标准范围为固定值的方式, 改变为通过莱因达准则计算的 3 倍标准差作为标准范围。具体方法如下: 针对时间序列观测数据, 将其中每一个数与其前一邻居相比, 计算邻差(第一个数邻差为 0), 形成邻差序列值; 邻差序列值基本满足正态分布(见图 2), 可以使用莱因达准则, 识别其中的异常值。

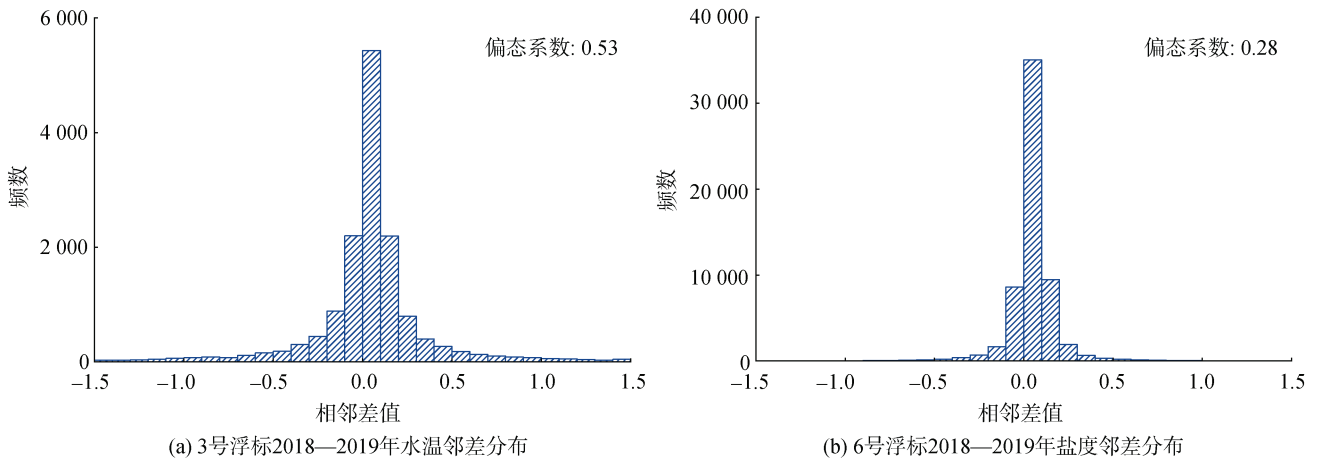


图 2 温、盐数据邻差序列值的频率分布图

Fig. 2 Frequency distribution of seawater temperature and salinity difference at two adjacent times

尖峰检验莱因达法的计算公式，假设有一组观测数据  $X_1, X_2, \dots, X_n$ ,  $i > 1$  时,  $X_i$  的前一邻居数据为  $X_{i-1}$ ,  $V_i = |X_i - X_{i-1}|$ ,  $V_i$  为邻差,  $V_1$  设为 0, 邻差序列值为  $V_1, V_2, \dots, V_n$ 。当  $V_i$  和  $V_{i+1}$  的残差均大于  $3\sigma$  时, 则  $V_i$  为尖峰变化,  $X_i$  数据为异常。  $V_i$  满足:  $|V_i - V| > 3\sigma$ , 并且  $|V_{i+1} - V| > 3\sigma$ ;  $n$  为相邻差值数据组的个数;  $V$  为均值,  $V = (\sum V_i) / n$ ;  $\sigma$  为标准差,  $\sigma = \sqrt{\sum (V_i - V)^2 / (n-1)}$ 。

相比于传统的“莱因达准则”，采用“尖峰检验莱因达法”更能明显提升异常数据的质控比例。6号浮标 2009—2012 年度水温数据多监测出 360 条异常数据，异常占比 0.34%，整体占比从 9.46% 提升到了 9.80% (见图 3)。6号浮标 2018—2019 年度水温数据多检出 477 条异常，异常占比 0.73%，整体占比从 2.97% 提升到 3.70%。盐度数据 3 号浮标 2010—2012 年度多检出 235 条异常，异常占比 1.06%，整体占比从 5.50% 提升到 6.56%。

然而，这种方法的局限性在于假定尖峰异常只有一个数据，前后邻居数据均为正常值。针对这种局限性，采用下一步的日均莱因达法做进一步处理。

5) 日均莱因达法。目的是找到超出日变率范围的异常数据。根据莱因达准则，以天为单位，分别计算每日水温和盐度的均值及标准差，查找残差(与均值差的绝对值)大于  $3\sigma$  的数据，并做异常标记。通过该方法，剔除每日数据中偏离较大的数据。与尖峰检验莱因达法相比，日均莱因达法可以进一步发现连续 2 个数据异常的情况。

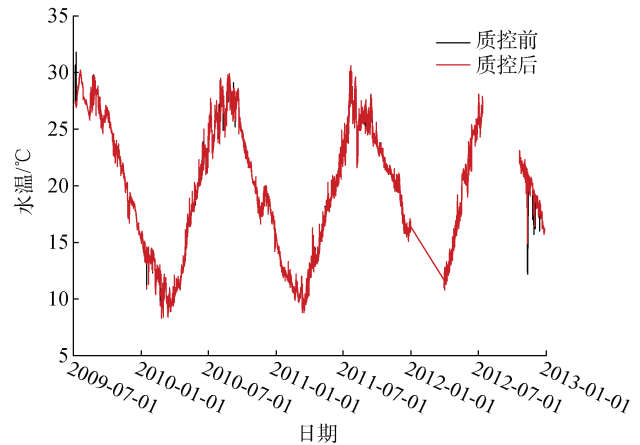


图 3 采用尖峰莱因达法质量控制前后对比

Fig. 3 Sample of the temperature time series before and after using the PauTa peak QC algorithm

此外，在浮标观测数据中，还会出现在海况较为稳定的情况下，某一天温度和盐度变化很小，这种情况下根据莱因达准则计算的 3 倍标准差会非常小，导致一些变化很小的正常值会被误判为异常。为解决这个问题，我们采用“最大允许误差”的方法，对莱因达法进行改进。

GB 17378.2-2007《海洋监测规范 第 2 部分：数据处理与分析质量控制》<sup>[12]</sup>，最大允许误差是指技术规范(如标准、检定规程)所规定的允许的误差极限值，是判定是否合格的一个规定要求，作为测量仪器的特性，规定最大允许误差和通过检定、校准去确定示值误差，在实用上具有十分现实的意义。国标 HY/T 142—2011《大型海洋环境监测浮标》<sup>[13]</sup>，其中“5.5 测量参数、范围和最大允许

误差”规定，海洋表层水温最大允许误差 0.5 摄氏度，电导率(表层盐度)最大允许误差 0.1 mS/cm。海水的电导率与盐度的换算系数 K 为 0.7<sup>[14]</sup>，可以计算出盐度的最大允许误差为 0.07。

据此，按照日均莱因达法判为异常的数据中，如果其残差小于最大允许误差，则为误判，通过本方法纠正为正确值。例如：6 号浮标 2019 年 11 月 1 日水温变化曲线(图 4)，该日平均水温 23.01，标准差  $\sigma = 0.07$ 。其中 14 点、14 点 10 分的值均为 23.3，残差(23.3 - 23.01 = 0.29) > 3 $\sigma$ ，根据日均莱因达法判断这 2 个数据为异常，但从整体曲线图和序列关系来看数据是正确的，由于残差 0.29 < 0.5(水温最大允许误差)，可以判断该数据为正确，避免了误判。

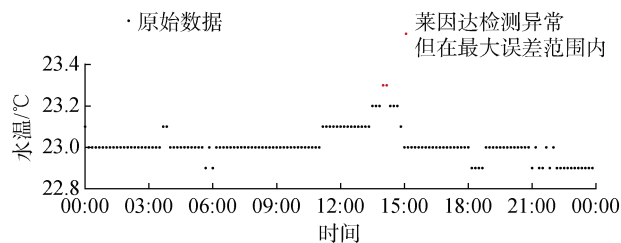


图 4 第 6 号浮标 2019 年 11 月 1 日水温变化曲线  
Fig. 4 Seawater temperature time series of Buoy 6 on November 1, 2019

6) 完成质控标记。采用上述质量控制方法，将已检测出的异常数据做了标记，这里进一步将剩余的数据标记为 0，作为正确数据。数据质控标记符号说明见表 1。

表 1 数据质控标记符号说明

Tab. 1 Summary of the quality flags

质控标记	含义	检测方法	说明
0	正确		
1	空值	预处理	数值为空的记录
2	异常	采集信息核对法	对照浮标维护记录，为异常
3	异常	范围检验法	超过区域温度/盐度范围，为异常
4	异常	尖峰莱因达法	邻差之残差超过 3 倍标准差，为异常
5	异常	日均莱因达法	数据之残差超过 3 倍标准差，为异常

## 2 方法验证与分析

为深入测试和验证上述的质量控制方案，本文选取黄海和东海共 6 个浮标温盐观测数据作为样本(图 5)进行验证，并对结果数据进行分析。

### 2.1 实验数据

本文选取中国科学院海洋研究所在黄海 1、2、3、4、5 号浮标和东海 6 号浮标在不同时间段的温度、盐度数据<sup>[15]</sup>，浮标的温盐传感器位于水下 1 m 处。具体包括：

1) 黄海站 1 号浮标：浮标类型为 3 m 浮标，观测站位置：(122°45.02'E, 38°45.58'N)，布放位置水深 52 m，观测时间自 2010 年 4 月 29 日至今，数据采样间隔 30 min。温盐观测数据 2010—2012 年总计 17 267 条记录。

2) 黄海站 2 号浮标：浮标类型为 2 m 浮标，观测站位置：(123°00.48'E, 39°04.00'N)，布放位置水深 37.3 m，观测时间自 2010 年 4 月 29 日至今，数

据采样间隔 1 h。温盐观测数据 2010—2012 年总计 11 072 条记录。

3) 黄海站 3 号浮标：浮标类型为 2 m 浮标，观测站位置：(122°35.50'E, 39°03.98'N)，布放位置水深 39.7 m，观测时间自 2010 年 3 月 27 日至今，2010—2011 数据采样间隔 1 h，2012 年之后采样间隔 10 min。温盐观测数据 2010—2012 年总计 22 102 条记录，2018—2019 年总计 15 755 条记录。

4) 黄海站 4 号浮标：浮标类型为 2 m 浮标，观测站位置：(123°06.01'E, 39°16.00'N)，布放位置水深 30 m，观测时间自 2010 年 4 月 29 日至今，数据采样间隔 1 h。温盐观测数据 2010—2012 年总计 9 634 条记录。

5) 黄海站 5 号浮标：浮标类型为 2 m 浮标，观测站位置：(122°44.98'E, 38°57.49'N)，布放位置水深 46 m，观测时间自 2010 年 4 月 29 日至今，数据采样间隔 1 h。温盐观测数据 2010—2012 年总计 9 886 条记录。

6) 东海站 6 号浮标：浮标类型为 10 m 浮标，观



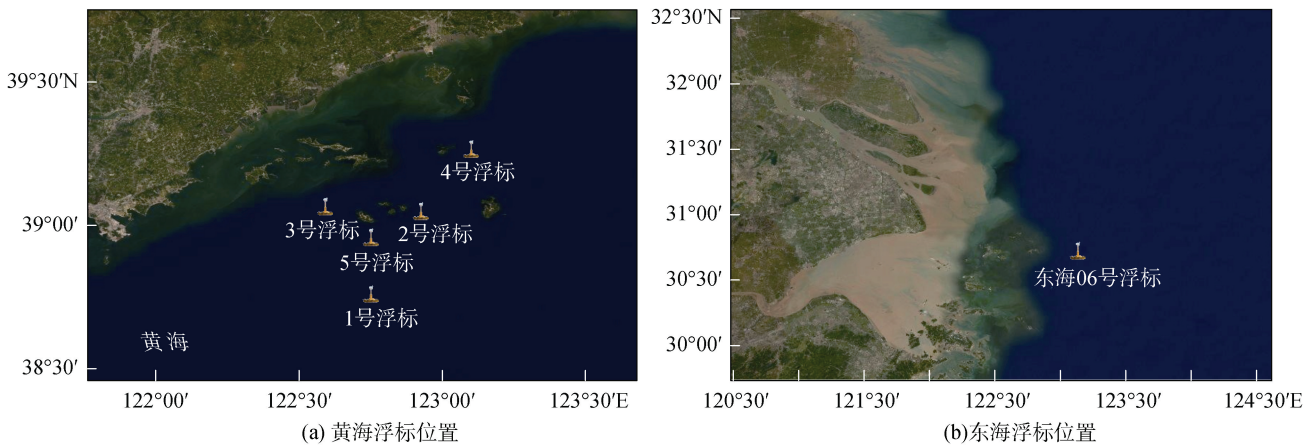


图 5 用于数据测试的黄海和东海浮标位置

Fig. 5 Location of the buoys

测站位置: (123°08.04'E, 30°43.04'N), 布放位置水深 61 m, 观测时间自 2009 年 7 月 6 日至今, 数据采样间隔 30 min。2009—2012 年总计 104 916 条记录, 2018—2020 年 5 月总计 65 519 条记录。

### 2.2 方法验证

按照质控标记对数据进行分类统计, 2018—2019 年度统计结果见下表 2, 2010—2012 年度统计结果见下表 3。

分析可见, 对于 3 号浮标, 标记为 1、2、3、4、5 的温度异常数据分别占比 0.61%、4.14%、0.00%、0.36%、0.22%; 盐度异常占比为 3.50%、1.26%、0.00%、0.55%、0.28%。总体来看, 标记了 5.33% 的

温度异常数据、5.59% 的盐度异常数据。对于 6 号浮标, 标记为 1、2、3、4、5 的温度异常数据占比分别为 1.23%、1.60%、0.00%、0.73%、0.14%, 盐度异常占比分别为 5.91%、1.17%、0.01%、0.73%、0.28%。总体来看, 标记了 3.70% 的温度异常数据、8.10% 的盐度异常数据。

其中, 采用本文提出的“尖峰检验莱因达法”, 明显提升了质量控制效果, 见表 2、表 3 中质控标记为 4 的异常数据量。2018—2019 年水温数据质量控制结果中, 3 号浮标检出 56 条异常, 6 号浮标检出 477 条异常。盐度数据结果中, 3 号浮标检出 86 条异常, 6 号浮标检出 480 条异常。

表 2 2018—2019 年度数据质量统计结果

Tab. 2 Statistical results of quality control in 2018–2019

浮标号	总数据量	质控标记	温度			盐度		
			数据量	异常占比%	异常占比合计%	数据量	异常占比%	异常占比合计%
3	15 755	0	14 915			14 875		
		1	96	0.61		551	3.50	
		2	653	4.14	5.33	199	1.26	5.59
		3	0	0.00		0.00	0.00	
		4	56	0.36		86	0.55	
		5	35	0.22		44	0.28	
6	65 519	0	63 089			60 205		
		1	809	1.23		3 876	5.92	
		2	1 051	1.6	3.7	767	1.17	8.11
		3	0	0.00		9	0.01	
		4	477	0.73		480	0.73	
		5	93	0.14		182	0.28	

表 3 2010—2012 年度数据质量统计结果

Tab. 3 Statistical results of quality control in 2010—2012

浮标号	总数据量	质控标记	温度			盐度		
			数据量	异常占比/%	异常占比合计/%	数据量	异常占比/%	异常占比合计/%
1	17 267	0	16 701			16 722		
		1	31	0.18		11	0.06	
		2	40	0.23	3.28	488	2.83	3.16
		3	438	2.54		20	0.12	
		4	48	0.28		22	0.13	
		5	9	0.05		4	0.02	
2	11 072	0	10 927			10 250		
		1	1	0.01		36	0.33	
		2	11	0.10	1.31	156	1.41	7.42
		3	78	0.70		489	4.42	
		4	49	0.44		122	1.10	
		5	6	0.05		19	0.17	
3	22 102	0	20 833			20 652		
		1	100	0.45		1 049	4.75	
		2	0	0.00	5.74	0	0.00	6.56
		3	1 120	5.07		123	0.56	
		4	32	0.14		235	1.06	
		5	17	0.08		43	0.19	
4	9 634	0	9 323			9 010		
		1	0	0.00		61	0.63	
		2	11	0.11	3.23	282	2.93	6.48
		3	244	2.53		126	1.31	
		4	41	0.43		127	1.32	
		5	15	0.16		28	0.29	
5	9 886	0	9 681			9 531		
		1	0	0.00		64	0.65	
		2	0	0.00	2.07	252	2.55	3.59
		3	167	1.69		2	0.02	
		4	29	0.29		26	0.26	
		5	9	0.09		11	0.11	
6	104 916	0	94 629			79 988		
		1	3	0.00		2 821	2.69	
		2	3	0.00	9.80	10 351	9.87	23.76
		3	9 550	9.10		10 123	9.65	
		4	569	0.54		1 540	1.47	
		5	162	0.15		93	0.09	

对数据质控前后结果进行对比,验证了质控方法的有效性,下面分别从三个方面说明。

1) 3、4、5、6号浮标水温时间序列对比(图6),质

控后,很多黑色异常点被标识,整体来看,质控后显示出明显的波动规律性。图中曲线中断的时间段没有数据,原因是浮标出现故障,数据缺失。3号浮标水温数据

起始时间为2010年4月29日,其中2011年11月5日到2012年9月26日数据缺失。6号浮标2012年1月1日到3月30日数据缺失。密集的黑色区域,属于浮标维护期间的异常数据,例如6号浮标在2012年7月14日到10月20日期间进行了大修,数据为异常。还有一类突然增大或减小的黑色点,属于异常突变数据。

护期间的异常数据,例如6号浮标在2012年7月14日到10月20日期间进行了大修,数据为异常。还有一类突然增大或减小的黑色点,属于异常突变数据。

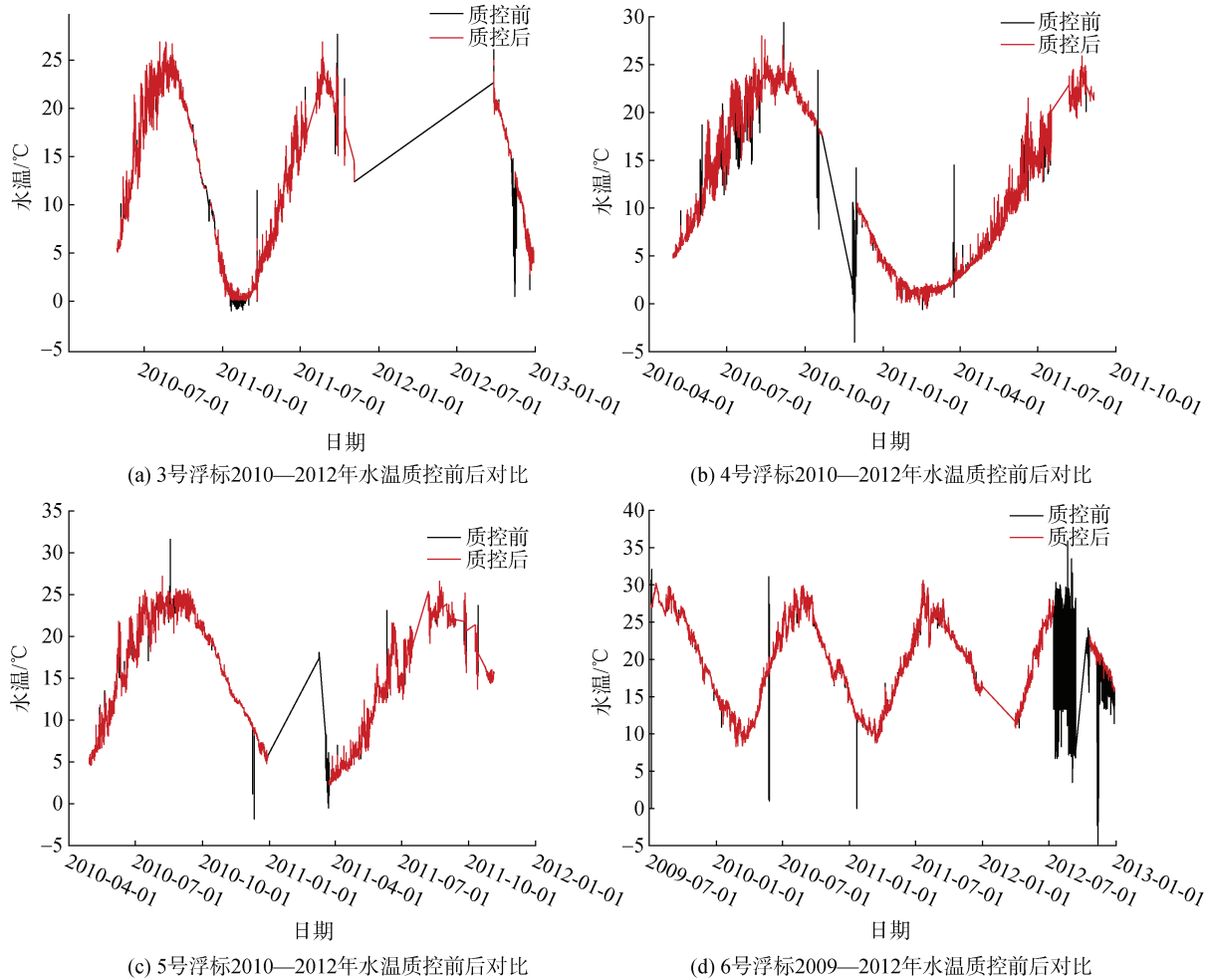


图6 质控前(黑色)后(红色)的温度时间序列对比图  
Fig. 6 Temperature time series before and after quality control

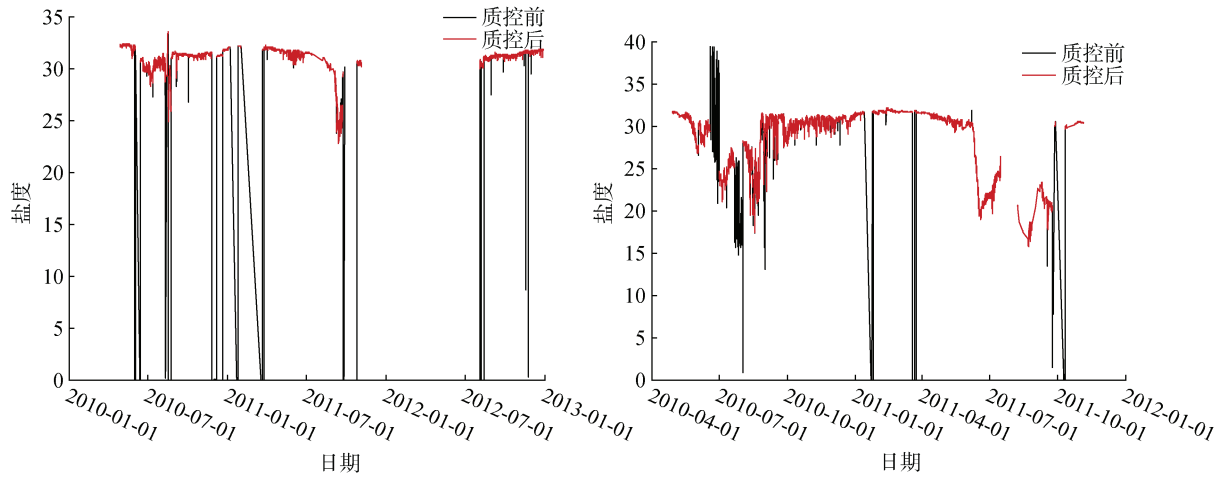
2) 1、2、3、6号浮标质控前后盐度时间序列对比(图7),可以看出,质控前盐度振幅变化很大,质控后振动幅度明显缩小。例如:2010—2012年,3号浮标振幅为0~38.9,质控后振幅为16.7~32.4。2018—2020年5月,6号浮标振幅为0~39.9,质控后振幅减小为18.9~34.9。

3) 质控后数据质量得到明显提升,异常点已经被去掉,典型示例见图8。其中a图是6号浮标2010年5月质控前后温度时间序列数据,5月25日到26日,有数据突然大幅下降和上升(黑色点),经过质控后这些突变点被检出,排除了异常。b图为6号浮标2011年11月盐度数据质控前后对比,可以明显看出,其中很多突然下降的异常数据被识别出来。

### 2.3 数据分析

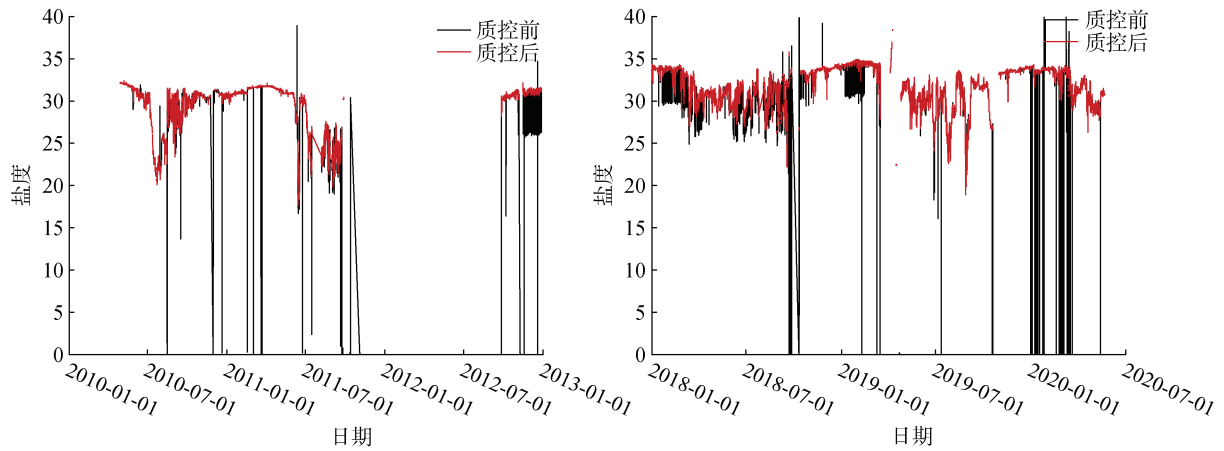
对质控后数据进行分析可以看出,黄东海温盐具有显著的季节变化特征(图6、图7),并且水温易受台风极端事件影响。

海表水温冬低夏高,季节循环变化明显<sup>[16]</sup>。每年2月份温度达到最低点,之后逐渐上升,到夏季8月份达到最高点,之后又逐渐下降,直至第二年的2月份,循环往复。例如2019年3号浮标最低温度在2月19日为1.1℃,最高温度在8月9日达26.7℃。2018年6号浮标最低温度出现在冬季2月12日为9.2℃,最高温度在8月9日达30.6℃。黄海、东海海水温度具有季节变化特征,这与前人的观测结果一致<sup>[17]</sup>,进一步验证了本文质量控制方法的可靠性和合理性。



(a) 1号浮标2010—2012年盐度质控前后对比

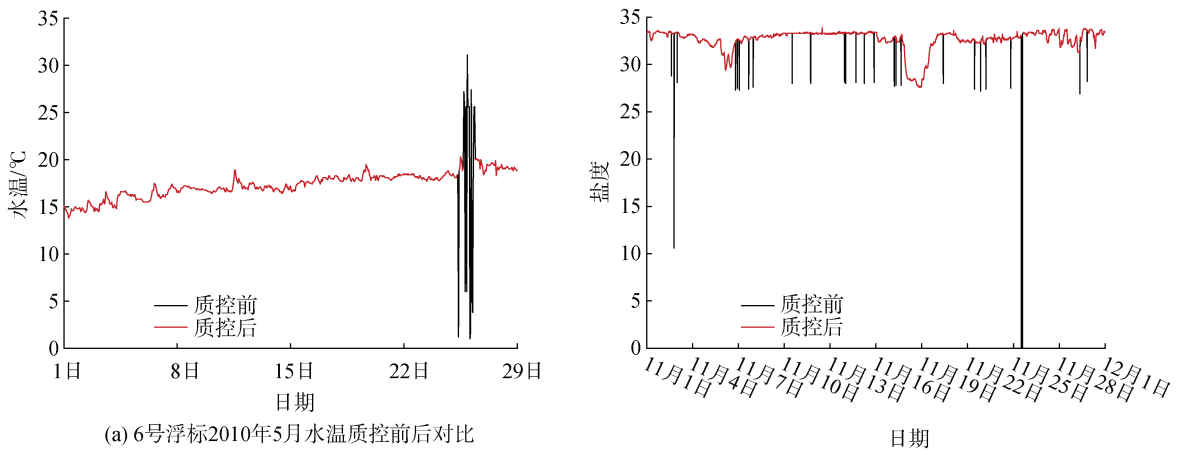
(b) 2号浮标2010—2012年盐度质控前后对比



(c) 3号浮标2010—2012年盐度质控前后对比

(d) 6号浮标2018—2020年盐度质控前后对比

图 7 质控前(黑色)后(红色)的盐度时间序列对比图  
Fig. 7 Salinity time series before and after quality control



(a) 6号浮标2010年5月水温质控前后对比

(b) 6号浮标2011年11月盐度质控前后对比

图 8 质控前后异常点明显被去掉示例  
Fig. 8 Samples of non-normal data are deleted after quality control

注: 采样时间为 0 时



海表盐度夏季偏低, 冬季平稳。同水温一样, 由于处于亚热带区域, 影响盐度变化的区域降雨、径流以及结冰、融冰等因子均具季节性, 使得黄东海表层盐度变化亦具有季节性变化特征。海表层盐度的季节性变化不如水温那么规则, 但总的变化趋势与水温的季节性变化相反<sup>[18]</sup>。每年夏季各地雨水多、雨量较大, 长江水位涨高, 长江口径流量增大, 影响到 6 号浮标的盐度下降明显。冬季雨水少, 盐度值较高, 并且比较平稳。

东海 6 号浮标 2019 年 8—9 月水温多次出现突然下降(图 9), 3 组红色标注的区间, 恰好是 4 次台风经过东海之时, 分析可见海洋水温的突变明显受到极端事件台风的影响。图中纵坐标数据为温度异常 (temperature anomaly), 表示在已剔除日变化和季节循环基础之上的温度异常。大于 0 表示暖异常, 即比平均态更暖; 小于 0 表示冷异常。可以看出三个台风经过东海时, 都导致了冷异常。

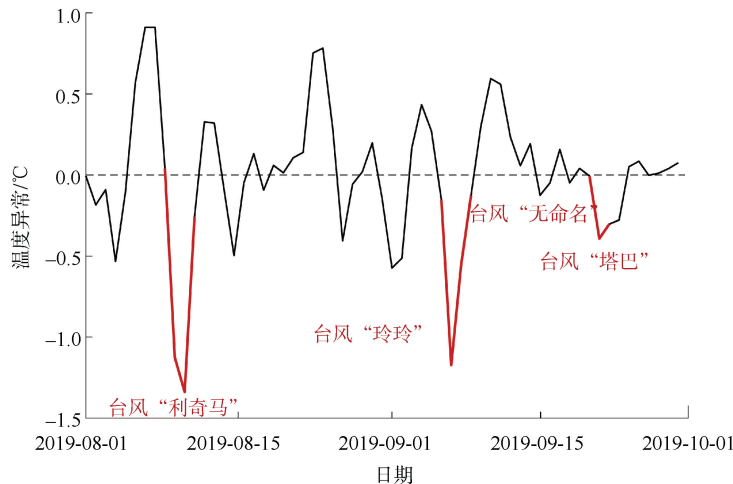


图 9 2019 年 8—9 月 6 号浮标水温异常曲线

Fig. 9 Seawater temperature anomaly of Buoy 6 from August 2019 to September 2019

第一组红色区 8 月 9、10、11 日台风 1909 号“利奇马”经过浙江、江苏到达中国黄海, “利奇马”是 2019 年最强的一次台风, 最大风速风力 62 m/s (18 级)<sup>[19]</sup>。这 3 天海水温度突然下降, 导致约 $-1.4\text{ }^{\circ}\text{C}$  的冷异常, 是这 2 个月最强的一次冷异常。

第二组红色区为 9 月 6—7 日台风 1913 号“玲玲”经过东海, 最大风速 55 m/s (16 级)。这 2 天海水温度出现突降, 导致约 $-1.2\text{ }^{\circ}\text{C}$  的冷异常, 是这 2 个月仅次于“利奇马”台风的第二强的一次冷异常。

第三组红色区 9 月 22—23 日台风 1917 号“塔巴”经过东海, 最大风速 33 m/s (12 级), 其间浮标水温再次下降, 导致约 $-0.4\text{ }^{\circ}\text{C}$  的冷异常。由于“塔巴”强度较弱, 离浮标位置也较远, 所以冷异常也相应较小。

### 3 结论和展望

本文研发了一种包括采集信息核对法、范围检验法、尖峰检验莱因达法、日均莱因达法在内的新型的、精细化的海洋浮标温盐观测数据 6 步质量控制方法和流程, 该方法具有如下优势:

1) 方法结合了浮标数据巡检记录、观测维护记录等辅助信息, 实现了对数据采集情况的考虑。这优于以往大部分仅基于数据本身的质控方法。

2) 针对浮标数据中广泛存在的异常峰值问题, 提出了一种创新的、将尖峰检验与莱因达准则相融合的尖峰检验莱因达法, 有效剔除了异常峰值。

3) 针对连续异常峰值, 提出了日均莱因达法进行质控, 有效检测出了同一日数据中的连续异常数据。

本文也针对质控后的数据进行了分析, 再次印证了黄东海地区海表温度和盐度变化具有明显的季节循环和易受极端事件影响的特征。包括中国科学院海洋研究所在内的国内机构在中国近海布放了 很多浮标, 针对不同年度、不同区域浮标、不同海洋要素, 积累了大量长时间序列的海洋观测数据。该精细化质量控制方法可有效支持其他浮标观测数据的质量控制, 为科学研究、海洋预报、海洋渔业生产等提供更准确的数据服务。

致谢: 感谢中国科学院近海海洋观测研究网络黄海站、东海站的数据支撑。

## 参考文献:

- [1] 戴洪磊, 牟乃夏, 王春玉, 等. 我国海洋浮标发展现状及趋势[J]. 气象水文海洋仪器, 2014, 6(2): 118-125. DAI Honglei, MOU Naixia, WANG Chunyu, et al. Development status and trend of ocean buoy in China[J]. Meteorological, Hydrological and Marine Instruments, 2014, 6(2): 118-125.
- [2] 卢勇夺, 王朝阳, 王豹, 等. 我国海洋锚系浮标数据异常值检测方法研究——以 QF110 和 QF306 为例[J]. 海洋预报, 2019, 36(6): 37-43. LU Yongduo, WANG Zhaoyang, WANG Bao, et al. Research on outlier detection method for marine anchor buoys in China, using QF110 and QF306 as an example[J]. Marine Forecasts, 2019, 36(6): 37-43.
- [3] 黄龙生, 张书杰. 海洋水文资料质量控制的新方法——盐度—密度统计方法[J]. 北京工业大学学报, 1993, 19(3): 92-97. HUANG Longsheng, ZHANG Shujie. A New quality control method for ocean hydrological data[J]. Journal Beijing University of Technology, 1993, 19(3): 92-97.
- [4] QIAN Chengcheng, LIU Aichao, HUANG Rui, et al. Quality control of marine big data—a case study of real-time observation station data in Qingdao[J]. Journal of Oceanology and Limnology, 2019, 37(6): 195-205.
- [5] 邢巍巍. 浅谈海域水质监测的质量保证和质量控制[J]. 环境与可持续发展. 2015, 40(6): 162-163. XING Weiwei. Discuss about the quality assurance and quality control of marine environment monitoring[J]. Environment and Sustainable Development, 2015, 40(6): 162-163.
- [6] 陈春满, 杨锦坤, 董明媚, 等. 国家海洋信息中心技术性文件 TA2014-001: 海洋水文气象资料质量控制方法[R]. 天津: 国家海洋信息中心, 2014. CHEN Chunman, YANG Jinkun, DONG Mingmei, et al. Technical Guidance Document of National Marine Data Information Center TA2014-001: Quality control method for ocean hydrological meteorological Data[R]. Tianjin: National Marine Data Information Center, 2014.
- [7] 于婷, 刘玉龙, 杨锦坤, 等. 实时和延时海洋观测数据质量评估方法研究[J]. 海洋通报, 2013, 32(6): 610-614, 625. YU Ting, LIU Yulong, YANG Jinkun, et al. Study on the quality control and quality assurance for the oceanographic data of real-time and delayed mode[J]. Marine Science Bulletin, 2013, 32(6): 610-614, 625.
- [8] KUIK A J, VAN VLEDDER G PH, HOLTHUIJSEN L H, et al. A method for the routine analysis of pitch-and-roll buoy wave data[J]. Journal of Physical Oceanography, 1988, 18(7): 1020-1034.
- [9] 史静涛, 周智海. 海洋站数据质量控制技术探讨[J]. 海洋技术, 2011, 30(1): 114-117. SHI Jingtao, ZHOU Zhihai. Study on technology of ocean station data quality control[J]. Journal of Ocean Technology, 2011, 30(1): 114-117.
- [10] 刘首华, 陈满春, 董明媚, 等. 一种实用海洋浮标数据异常值质控方法[J]. 海洋通报, 2016, 35(3): 264-270. LIU Shouhua, CHEN Manchun, DONG Mingmei, et al. A quality control method for the outlier detection of buoy observations[J]. Marine Science Bulletin, 2016, 35(3): 264-270.
- [11] 冉珊珊, 时宇, 杨一帆, 等. 东海冬季海水溶解氧、盐度的分布特征分析[J]. 海洋科学前沿, 2017, 4(4): 118-126. RAN Shanshan, SHI Yu, YANG Yifan, et al. Distribution characteristics of dissolved oxygen and salinity in sea water in the East China Sea in winter[J]. Advances in Marine Sciences, 2017, 4(4): 118-126.
- [12] 国家质量监督检验检疫总局. 海洋监测规范(第 2 部分: 数据处理与分析质量控制): GB 17378.2-2007[S]. 北京: 中国标准出版社, 2007. General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China. The specification for marine monitoring (part 2: Data processing and quality control of analysis): GB 17378.2-2007[S]. Beijing: China Quality and Standards Publishing Press, 2007.
- [13] 国家海洋局. 大型海洋环境监测浮标: HY/T 142-2011[S]. 北京: 国家海洋局, 2011. State Oceanic Administration. Large mooring buoy system for ocean observing: HY/T 142-2011[S]. Beijing: State Oceanic Administration. 2011.
- [14] 汪一航. 一种快速计算标准海水 Rt 值的实用方法[J]. 海洋通报, 1999, 18(2): 76-79. WANG Yihang. A practical method for rapidly calculating the Rt value of standard seawater[J]. 海洋通报, 1999, 18(2): 76-79.
- [15] 刘长华, 王旭, 贾思洋, 等. 东海浮标综合观测网络支撑海洋环境保护和防灾减灾预报[J]. 中国科学院院刊, 2019, 34(10): 1186-1194. LIU Changhua, WANG Xu, JIA Siyang, et al. Comprehensive observation network of buoys in East China Sea supporting marine environmental protection, disaster prevention, and mitigation prediction[J]. Bulletin of Chinese Academy of Science, 2019, 34(10): 1186-1194.
- [16] 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 the shore-based observation data in the coastal of Bohai Sea[J]. Ocean Development and Management, 2015, 32(12): 31-34.

- [17] 于非, 许一. 东海海表面温度长期变化趋势研究[J]. 海洋科学进展, 2003, 21(4): 477-481.  
YU Fei, XU Yi. Long term variation trend of sea surface temperature in the East China Sea[J]. *Advances in Marine Science*, 2003, 21(4): 477-481.
- [18] 苗庆生, 杨锦坤, 杨扬, 等. 东海 30 °N 断面冬季温盐分布及年际变化特征分析[J]. 中国海洋大学学报(自然科学版), 2016, 46(6): 1-7.  
MIAO Qingsheng, YANG Jinkun, YANG Yang, et al. Distribution and interannual variation of temperature and salinity in winter at 30 °N section of the East China Sea[J]. *Zhongguo Haiyang Daxue Xuebao (Ziran Kexue Ban)*, 2016, 46(6): 1-7.
- [19] 向纯怡, 赵海坤, 刘青元, 等. 1909 号台风“利奇马”登陆后强降水分布特征[J]. 气象科学, 2020, 40(3): 294-302.  
XIANG Chunyi, ZHAO Haikun, LIU Qingyuan, et al. The distribution of torrential rain after landfalling of Typhoon “Lekima”(1909)[J]. *Journal of the Meteorological Sciences*, 2020, 40(3): 294-302.

## A new quality control scheme for marine buoy temperature and salinity data

REN Huan-ping<sup>1, 2</sup>, ZHANG Bin<sup>1, 2</sup>, TAN Zhe-tao<sup>3, 4</sup>, LI Fu-chao<sup>1, 2</sup>

(1. Institute of Oceanology, the Chinese Academy of Sciences (CAS), Qingdao 266071, China; 2. Center for Ocean Mega-science, CAS, Qingdao 266071, China; 3. Institute of Atmospheric Physics, CAS, Beijing 100029, China; 4. University of Chinese Academy of Sciences, Beijing 100049, China)

**Received:** Jan. 26, 2021

**Key words:** marine buoy; seawater temperature and salinity; quality control; PauTa criterion

**Abstract:** Based on the error diversity of marine buoy observational data, this study presents a new quality control (QC) system for marine buoy temperature and salinity data. This QC scheme includes six checks with the PauTa criterion as its core. An innovation of this scheme is a new QC algorithm that combines the traditional Spike check and PauTa criterion check. Its robustness has been verified by the seawater temperature and salinity data of buoys in the Yellow Sea and East China Sea. Several spurious measurements or outliers in long time series can be effectively identified. We propose that this new scheme can be used in the QC of buoy observation data in the future. Furthermore, the quality-controlled data could be used for marine scientific research, marine meteorological prediction, marine disaster forecast, and fishery development.

(本文编辑: 赵卫红)