

Quality control of time-series seawater temperature and wave data adapted to the regional conditions of the Baltic Sea

by

Michał Iwaniak, Tamara Zalewska,
Patryk Sapiega*

DOI: <https://doi.org/10.26881/oahs-2025.1.06>

Category: **Original research papers**

Received: **August 29, 2024**

Accepted: **January 31, 2025**

*Institute of Meteorology and Water Management - National Research
Institute, Gdynia, 42 Waszyngtona St., Poland*

Abstract

Marking outliers using available methods for identifying such observations should be a standard practice in the database management process. The research aimed to adapt universal data quality control tools and tests to their applicability in the southern Baltic Sea by setting new limit values, enabling the detection of erroneous or suspicious data, which can be subjected to expert verification at a later stage. This verification stage may include analysing current conditions and processes and determining the values measured at a given time and space. Our research has proven that using global algorithms requires adapting the limit values based on regional conditions. The global quality control tests, such as the spike test, Dixon's $4(\sigma)$, Q-Dixon, Hampel, quartile, and gradient tests for oceanographic data, were examined, and their application in the research area was verified. As part of the task, a few tests were conducted in various areas of the Baltic Sea, and they were modified to adapt them to the research results of the southern Baltic Sea. Depending on the methodology adopted, verification tests result in the selection of suspicious observations, enabling their expert assessment and the final qualification of measurements that may be considered outliers.

Key words: quality control, long-series data, seawater temperature, wave, Baltic Sea

* Corresponding author: patryk.sapiega@imgw.pl

1. Introduction

Quality control of oceanographic measurement results is the primary activity on data sets to obtain a quality database. Although actual measurements constitute the most reliable data source and measurement devices are often equipped with internal data quality control algorithms, they are not free from potentially wrong values, that is, outliers.

Outliers in a data set may result from natural phenomena or human errors made during analyses, malfunctioning equipment, or methodological errors. Eliminating outliers is a condition for obtaining a reliable assessment of the environment. Identifying unusual results may allow you to discover the causes of errors and eliminate them in the future (Budka et al., 2013).

One of the reasons for questionable measurement results in a measurement series is the so-called gross errors, sometimes referred to as excessive errors. They are due to a single impact acting temporarily, ephemerally, and occurring only in certain measurements (Twardowski & Traple, 2006).

Appropriate algorithms enable the verification of the correctness of recording sets of hydrodynamic and physicochemical parameter results. The identification of outliers is important both in relation to large data sets used to analyse trends or correlation relationships over many years and in relation to smaller data sets that are particularly important in annual assessments of water status. Many researchers have conducted research on the quality control of seawater temperature (Boyer and Levitus, 1994; Castelhão, 2015; Cummings, 2011; Good et al., 2013, 2023; Ingleby and Huddleston, 2007; Peterson et al., 1998) and wave parameters (Doong et al., 2007; Min et al., 2017; Morang, 1990; Salcedo Parra et al., 2008; Xie et al., 2023). Currently, many measuring devices are equipped with an internal data quality control system (NORTEK, 2020; RADAC, 2020). However, only very extreme values are flagged, for example, very high or very low, which could indicate a severe failure of the device. This solution is not effective in detecting outliers. There are several data quality control tests for the Baltic Sea (Copernicus Team, 2017, 2018, 2020; FINO, 2020; IOC, 1993; IOOS, 2019), but they have drawbacks because of their global nature, which often makes it impossible to detect an incorrect measurement value in local conditions.

The aim of the research was to adapt universal data quality control tools and tests to their applicability in the southern Baltic Sea by setting new limit values, enabling the detection of erroneous or suspicious data, which can be subjected to expert verification

at a later stage. This verification stage may include analysing current conditions and processes and determining the values measured at a given time and space. The necessity to adapt global tests to regional ones is indicated by many research studies dealing with the quality control of data with both sea water temperature (Good et al., 2013; Kennedy, 2014; Lellouche et al., 2013) and wave parameters (Bitner-Gregersen and de Valk, 2008; Team C.M.I.S., 2020), pointing out the imprecision of global tests in water areas with different bathymetric and climatic conditions and hydrodynamics compared to the entire sea or ocean where the analysed water body is located. Determining new limit values applicable to available tools is crucial for building reliable databases that are the basis for analysing changes taking place in the southern Baltic area and are used to verify numerical models.

The study assumed the adaptation of universal (global) data quality control procedures recommended by European scientific organisations and institutions (SeaDataNet, Copernicus, ICES (International Council for the Exploration of the Sea), HELCOM) to the conditions of the southern Baltic Sea. The global quality control, such as the spike test, Dixon's $4(\sigma)$, Q-Dixon, Hampel, quartile, and gradient tests for oceanographic data (water temperature and wave parameters), were examined. Based on data from the coastal and open sea water monitoring networks and expedition measurements, calculations were made on test sets, considering the control of limit values, temporal consistency, space-time consistency, and internal (logical) consistency for wave and water temperature parameters. The selected global tests were tested on the analysed data set using the recommended limit values of these tests, and then new regional values for these tests were determined based on test statistics.

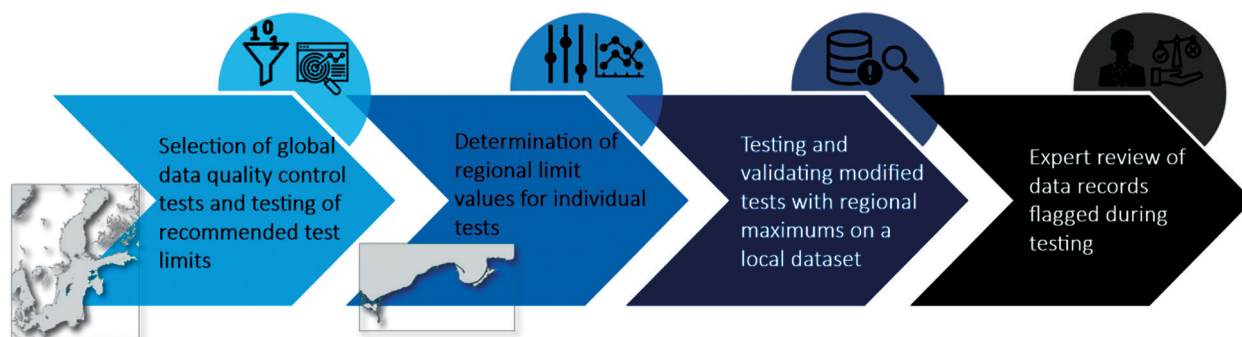
2. Methodology and research area

2.1. Procedure for determining test limits and verification (Fig. 1)

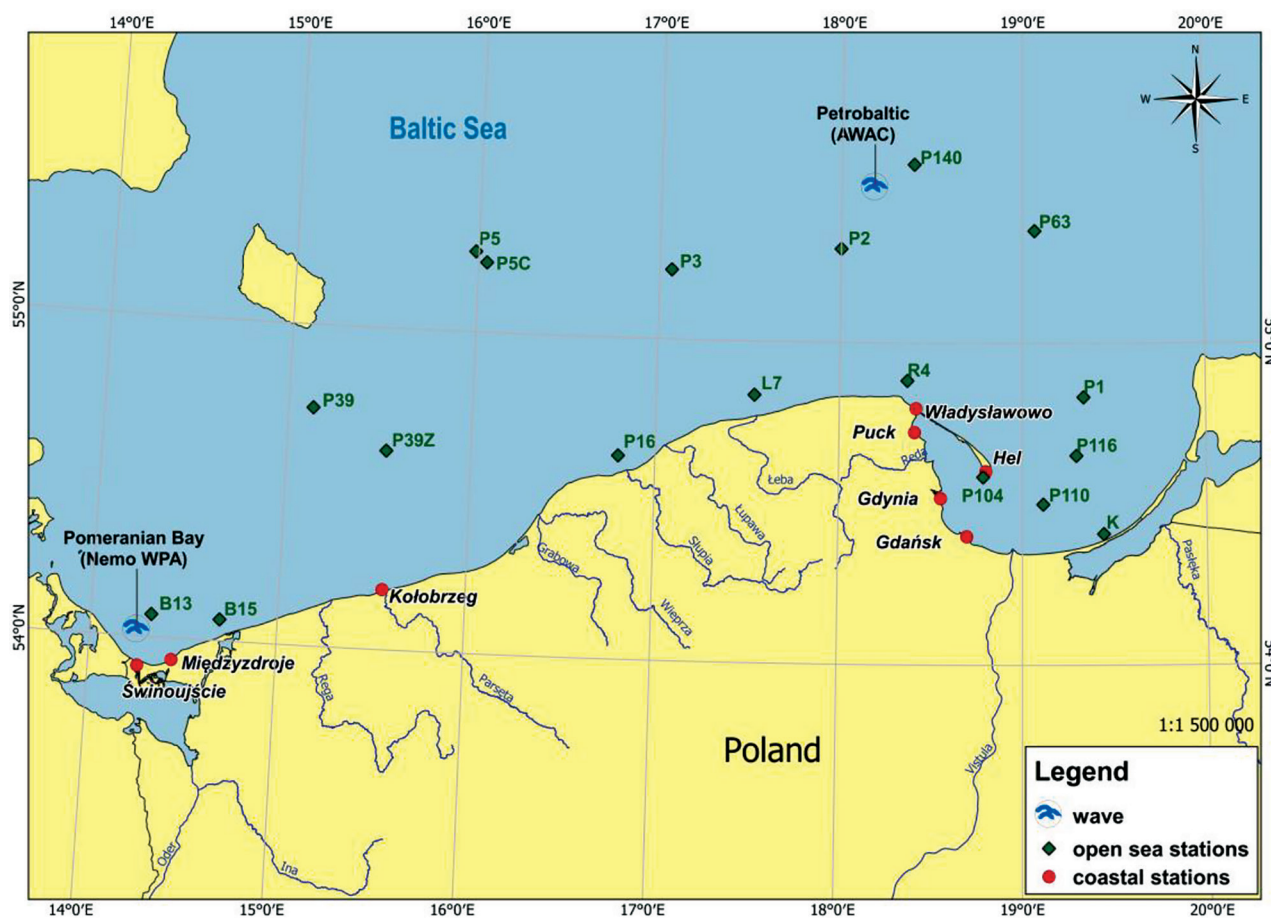
Figure 1 presents the procedure for determining test limits, consisting of the following stages:

1. Selecting data quality control tests (a.o. spike test, gradient test) and testing selected global tests with recommended thresholds on a local dataset.
2. Determining regional values based on calculation statistics—new limit value tests were



**Figure 1**

Methodological procedure for modifying global tests.

**Figure 2**

Location of coastal and sea stations.

set at the 99th percentile of data series on sea water temperature and wave parameters.

3. Testing and validating tests with new set limit values with regional maximums on a local dataset.
4. Verifying the correctness and effectiveness of the tests combined with regional limit values.

2.2. Measurement data and research area

The research includes temperature and wave measurements performed at stations located in the coastal zone and the open sea (Fig. 2). It also includes an analysis of available data quality control tests for identifying outlier measurements described

in the scientific literature and methodological guidelines recommended by the Baltic Sea Marine Environment Protection Commission (Svendsen, 2019) and the pan-European marine data management infrastructure, that is, SeaDataNet (GOSUD, 2003; GTSP, 2010).

The described methods were applied to the measurement results of wave parameters: significant wave height (H_s), mean wave period (T_m), and peak wave period (T_p) examined from 2018 to 2021; temperature results in open sea area examined from 1959 to 2019; and coastal areas studied from 1946 to 2019 (Table 1).

Temperature measurements were carried out by the Institute of Meteorology and Water Management—National Research Institute. The database includes nearly 200,000 results (time resolution—1 day) for coastal stations in 1946–2019 and 40,000 results (time resolution—6 times per year) for open sea stations in 1959–2019. Measurement methods have changed over such a long period; therefore, it is not possible to provide a unified measurement specification. It should be noted that these measurements have been performed with valid standards. The distribution of temperature is presented as histograms for coastal and open sea stations, respectively (Fig. 3).

The histogram (Fig. 3) shows the distribution of temperature measurements at all coastal and open sea stations. In the case of coastal stations, the majority of results are in the range from 0 to 3, while the least results are in the range above 24°C. The average temperature in the coastal zone is 9.5°C, with a standard deviation of 6.6. The median temperature is 8°C. In the case of open sea stations, the majority of results are in the range from 3 to 6, while the least results are in the range above 23°C. The average temperature in the coastal zone is 7.6°C, with a standard deviation of 5.2. The median temperature is 5.7°C.

The research included measurement results ranging from tens to almost 200,000 for the mentioned parameters. Verifying such large data sets, without appropriate algorithms to automate this process, is difficult or even impossible within an acceptable time frame. The tests performed and the results indicating outliers constitute an indicator for the substantive user (data operator), who finally assesses the quality of the selected data records and properly marks or eliminates them from the data set. As part of this task, the effectiveness of specific tests was also assessed, and actions were taken to verify and adapt data quality control methodologies to the Polish area of the southern Baltic Sea (Fig. 2).

Table 1

Area and range of analyses

Area of tests	Research period	Parameter	Number of stations	Number of analysed results
Coastal stations	1946–2019	Water temperature	8	190,967
Open sea stations	1959–2019	Water temperature	18	40,651
	2018–2021	Wave	2	66,792

Wave data from the Petrobaltic platform come from the AWAC (Acoustic Wave and Current Profiler) 400 kHz device from Nortek, an ADCP (Acoustic Doppler Current Profiler)-type device. The wave measurement range is limited to 15 m in height, and the accuracy of recording significant wave height is <1% of the measurement value (approx. 1 cm). ADCP device is characterised by 1.5 Hz sampling of the surface elevation and a cell size of 1.0–8.0 m with several cells of 200. The second device, NEMO WPA (Waves Processing Array Nemo), located in the Pomeranian Bay, also uses ADCP technology. Its significant wave height recording range is 0–30, with an accuracy of 1 cm.

The histogram (Fig. 4) shows the distribution of significant wave height values for both analysed stations over the entire period. Most measurements are in the range of 0.3–0.6 m. The size of the classes gradually decreases as the wave height increases. Both values close to zero and those above 1.6 m constitute <9000 measurements. The average value of the significant wave height is 1.02 m, with a standard deviation of 0.75 m, and the median is 0.85 m.

2.3. Spike test

The spike test is described in the SeaDataNet Data Quality Control Procedures (Version 2.0, May 2010) (GOSUD, 2003). The test compares specific values with the results obtained in an earlier period and with the result obtained in a later period. It applies to wave measurements, temperature, and salinity results at one level. The formula for the spike test is as follows:

$$\text{Test value} = \left| V_2 - \frac{V_3 + V_1}{2} \right| - \left| \frac{V_3 - V_1}{2} \right| \quad (1)$$

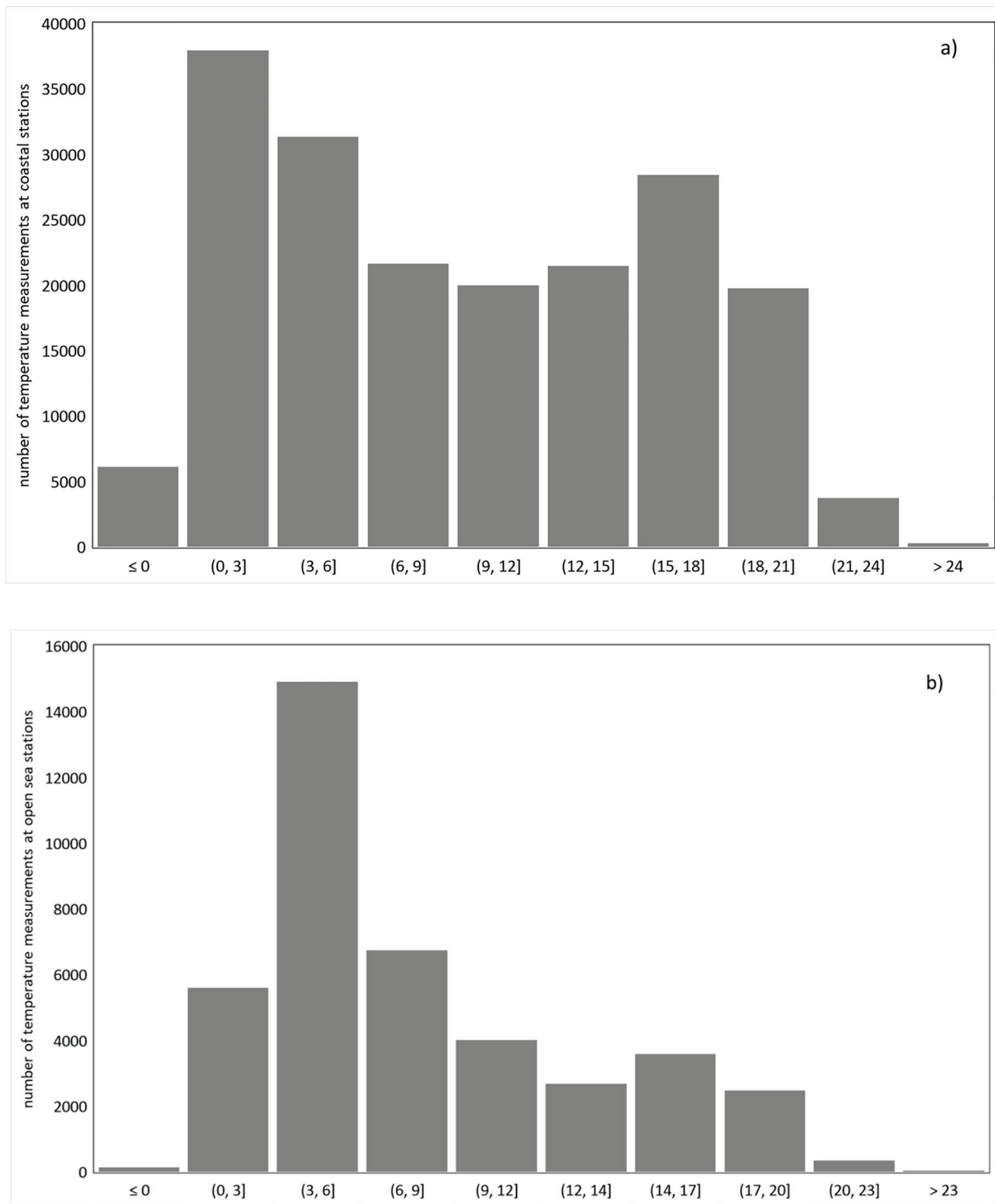
where

V_2 —analysed measurement

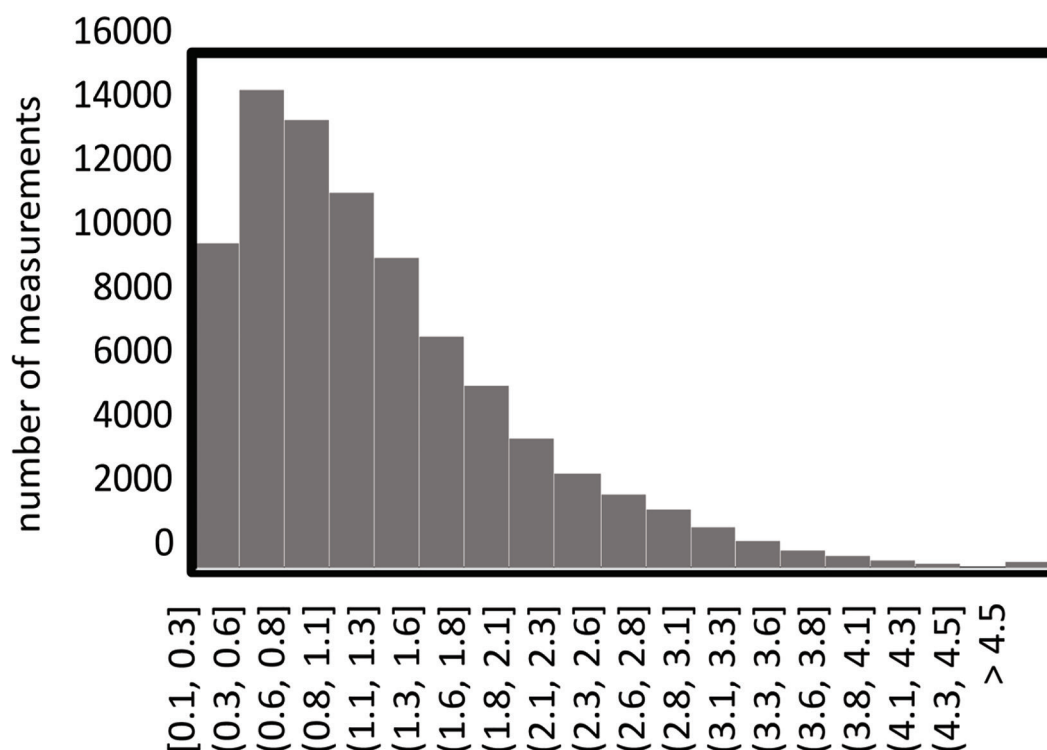
V_1 —previous measurement

V_3 —another measurement



**Figure 3**

Histograms of temperature [°C] at coastal (A) and open sea stations (B) in the southern Baltic Sea.

**Figure 4**

Significant wave height [m] distribution for both stations.

For water temperature, the V2 value is flagged when the test value exceeds 6°C.

2.4. Dixon's 4σ

In the HELCOM guidelines for PLC (Pollution Load Compilation) - Guidelines for waterborne pollution inputs to the Baltic Sea, in Chapter 11.3 Outliers, Dixon's 4σ is indicated as one of the universal tests recommended for identifying outliers. This test is universal in its assumption and relatively simple in construction. Outliers are values that fall outside the interval defined by the mean ±4 times the standard deviation. The formula for Dixon's 4 sigma is as follows:

$$(\mu - 4\sigma, \mu + 4\sigma) \quad (2)$$

where:

μ—mean from the analysed period

σ—standard deviation from the analysed period

2.5. Q-Dixon's test

The Q-Dixon test is used to check a data set for the presence of gross errors. The main limitation of

this test is the size of the set, which can contain 3–10 results. The literature (Namieśnik et al., 2007) also presents a modified Q-Dixon test that can be applied to sets containing up to 40 results.

To perform this test, the measurement results should be arranged in ascending order. Then, the values of the Q_1 and Q_n parameters should be calculated, and the obtained results should be compared with the critical value Q_{kr} for a given α level. If any of the calculated parameters Q_1 and Q_n is greater than the critical value Q_{kr} , the result based on which x_1 or x_n was calculated should be rejected as a gross error. The values of the Q_1 and Q_n parameters, depending on the number of analysed results, are calculated using the following formulas:

$$n = 3 - 10 \quad R = x_n - x_1 \quad Q_1 = \frac{x_2 - x_1}{R} \quad Q_n = \frac{x_n - x_{n-1}}{R} \quad (3)$$

$$n = 8 - 12 \quad Q_1 = \frac{x_2 - x_1}{x_{n-1} - x_1} \quad Q_n = \frac{x_n - x_{n-1}}{x_n - x_2} \quad (4)$$

$$n > 12 \quad Q_1 = \frac{x_3 - x_1}{x_{n-2} - x_1} \quad Q_n = \frac{x_n - x_{n-2}}{x_n - x_3} \quad (5)$$



Using formula (3), and applying values to variables: for the significance level $\alpha = 0.1$, $\alpha = 0.05$, $\alpha = 0.01$ and the set of values $n \leq 10$.

In the case of formulas (4) and (5), the Q_{kr} value should be read from the right table, modified for $n \leq 40$. The mentioned tables are available in the literature (Namieśnik et al., 2007).

2.6. Hampel test

The literature (Namieśnik et al., 2007) also discusses the Hampel test, which is used to identify outlier measurements. The popularity of this test is due to its simplicity, and there is no need to refer to tables of critical values. The procedure for this test is as follows:

- arrange the values (x_i) in ascending order,
- calculate the median value (Me) from the ranked values (x_i),
- calculate the value of deviations r_i from the median value for each result according to the formula $r_i = (x_i - \text{Me})$
- calculate $|r_i|$,
- arrange the values $|r_i|$ in ascending order,
- calculate the value of the median of $\text{Me}_{|r_i|}$ deviations,
- check the presence of outliers based on the criterion: if the condition $|r_i| \geq 4.5\text{Me}_{|r_i|}$, the result x_i is considered an outlier.

2.7. Quartile test

Another example is the quartile test. In this method (like Hampel test), there is no need to use statistical tables (Chromiński & Tkacz, 2010). Identifying outliers using the quartile test requires:

- calculating quartile 1 (Q_1) based on the test set,
- calculating quartile 3 (Q_3) based on the test set,
- calculating the quartile range using the formula $H = Q_3 - Q_1$,
- checking the values: those that are lower than $Q_1 - 1.5H$ and higher than $Q_3 + 1.5H$ are considered suspicious (potentially outliers),
- checking the values: those that are lower than $Q_1 - 3H$ and higher than $Q_3 + 3H$ are considered outliers.

2.8. Gradient test

The gradient test, also described in the [in:] SeaDataNet Data Quality Control Procedures guidelines (Version 2.0, May 2010) (GOSUD, 2003; GTSP, 2010), enables the comparison of monitoring

measurement results performed in the water column. The gradient test is described by the formula:

$$\text{Test value} = \left| V_2 - \frac{V_3 + V_1}{2} \right| \quad (6)$$

where

V_2 —analysed measurement from level n

V_1 —measurement from level $n-1$

V_3 —measurement from level $n+1$

This test evaluates the difference between the measured temperature value at a specific level and the mean temperature of the higher and lower levels. The V_2 value is flagged when the test value exceeds 9°C .

3. Results

3.1. Spike test

3.1.1. Wave parameters

A spike test, which determines the relationship between adjacent measurements, is a good verification indicator for flagging suspicious records from the database. The Copernicus Marine Environment monitoring service (Copernicus Team, 2020) estimated and presented the spike test results, which are valid for the entire Baltic Sea area (Table 2). In the case of wave parameters, the test formula remains unchanged, and the test result depends on the analysed parameter.

Considering the data resolution and limitations of the reanalysis data, the authors attempted to identify new spike test result limits or confirm those proposed by Copernicus. A result for the maximum wave height parameter was also introduced, which was not included in Copernicus.

The data come from the open sea zones (Tables 3 and 4) and the shallow water zones (Tables 5 and 6). In both locations, the period was divided into a

Table 2

Summary of the results of the spike test formula according to Copernicus for the entire Baltic Sea area

test value = $\left V_2 - \frac{(V_3 + V_1)/2}{2} - \frac{(V_3 - V_1)/2}{2} \right $	Spike test result according to Copernicus for the entire Baltic Sea area:
Wave parameter:	
Hs	3
Tm	4
Tp	10

Table 3

Proposed and determined new limit values for the non-storm season in the open waters of the Baltic Sea, based on measurement data from the Petrobaltic point

	Proposed value (Copernicus)	99th percentile of result	Proposed value	Number of records	Percentage of measurements flagged (%)
Hs	3	3.54	3.5 m	1	0.005
Tm	4	4.69	5.0 s	685	3.335
Tp	10	19.29	20.0 s	71	0.346

Table 4

Proposed and determined new limit values for the storm season in the open waters of the Baltic Sea, based on measurement data from the Petrobaltic point

	Proposed value (Copernicus)	99th percentile of result	Proposed value	Number of records	Percentage of measurements flagged (%)
Hs	3	5.49	5.5 m	37	0.129
Tm	4	9.97	10.0 s	12	0.042
Tp	10	17.37	18.0 s	4	0.014

Table 5

Proposed and determined new limit values for the non-storm season in the open waters of the Baltic Sea, based on measurement data from the Pomeranian Bay point

	Proposed value (Copernicus)	99th percentile of result	Proposed value	Number of records	Percentage of measurements flagged (%)
Hs	3	0.52	0.55 m	64	0.787
Tp	10	3.6	4.0 s	235	2.889

Table 6

Proposed and determined new limit values for the storm season in the open waters of the Baltic Sea, based on measurement data from the Pomeranian Bay point

	Proposed value (Copernicus)	99th percentile of result	Proposed value	Number of records	Percentage of measurements flagged (%)
Hs	3	0.74	0.8 m	27	0.286
Tp	10	4.35	4.5 s	174	1.841

non-storm season (April–August) and a stormy season (IX–III), allowing us to specify the test result values in seasonal terms.

In the non-storm season, the test results were close to those proposed by Copernicus. Based on the data from the Petrobaltic point, the test result value for significant wave height remains unchanged compared to the one proposed by Copernicus, that is, 3 m. The mean wave period increased from 4 s to 5 s, and the peak wave period doubled from 10 s to 20 s. In the storm season, the resulting values were 5.5 m for significant wave height, 10 s for the mean wave period, and 17.4 s for the peak wave period.

The point in the Pomeranian Bay is not classified as open water of the Baltic Sea, which means it

is impossible to refer to the values proposed by Copernicus. In addition, the measuring device measures parameters such as significant wave height and peak wave period, and the following limit values have been proposed: Hs—0.55 m and Tp—4 s in the stormless season and 0.8 m and 4.5 s in the stormy season.

3.1.2. Temperature measurements

The set of temperature measurements tested using the spike test formula (1) includes 190,967 results measured at eight coastal stations from 1946 to 2019. The test limit value for temperature results is 6°C.



At the same time, it should be noted that this tested formula skips the first and last measurements from the analysed time interval due to the lack of possibility of comparison to neighbouring values. Therefore, 190,951 values were tested out of all results. Only one outlier measurement was identified using the test limit indicated in the SeaDataNet guidelines (6°C).

It is strongly recommended that measurements be carried out at 1-day intervals. In the case of longer time intervals, differences between measurements may result from seasonal pattern temperature variations rather than from incorrect values.

Considering such a large number of water temperature measurements at coastal stations from 1946 to 2019, it may be surprising that only a single outlier was identified using the spike test. On the one hand, this may indicate a very good and careful system of collecting, verifying, and recording results. On the other hand, it is necessary to consider whether adopting the globally indicated limit values of this test for regional results is appropriate.

The global limit value for the spike test in the guidelines is 6°C.

After calculating the test statistics for the temperature results according to the formula (1), the 99th percentile was calculated as 1.5. The test results indicate that 99% of the differences between adjacent temperature measurements are $\leq 1.5^\circ\text{C}$. This result was mathematically rounded to 2, and this value was adopted as the test limit value for temperature measurements in the southern Baltic Sea (Table 7). Simultaneously, the indicated limit value is used to flag suspicious values whose correctness should be subjected to expert assessment.

The second test conducted on 190,951 temperature results indicated 703 outlier measurements.

3.2. Dixon's 4 σ

3.2.1. Wave parameters

As a result of the test based on the mean and four times the standard deviation, 154 records were flagged at the Petrobaltic point, constituting a total of 0.32% of all records in the analysed data. The largest number, that is, 101 records, was flagged for the significant wave height parameter in the stormless season and 41 in the stormy season. Both the mean and peak wave periods in the storm season were characterised by a small number of flagged records (2–5), while in the non-storm season, they were four and one, respectively, representing only 0.012% (Table 8).

In the Pomeranian Bay, the mean and four times the standard deviation test showed a higher percentage of flagged measurements compared to the data set from the Petrobaltic point. In the storm season, the values for H_s and T_p were 0.71% and 0.80%, respectively, and in the stormless season 0.21% and 1.02%, respectively (Table 9).

3.2.2. Temperature measurements

The set of temperature measurements, tested using Dixon's 4 σ formula (2), includes 190,967 results measured at eight coastal stations from 1946 to 2019. Due to seasonal temperature, the range of meteorological frequencies of the year is extended. For each station, the mean and standard deviation were calculated for the results measured in one of the four meteorological seasons. As a result of the test, eight outlier measurements were identified (Table 10).

The observations allow us to conclude that a positive feature of this test is its universal nature,

Table 7

Proposed spike test limits for temperature results at southern Baltic coast stations

	Limit value for the test	99th percentile score	Proposed limit value	Number of records flagged	Percent of records flagged (%)
Temperature measurements at coastal stations	6	1.5	2	703	0.37

Table 8

Number of flagged measurements in the Dixon's 4 σ result at the Petrobaltic point

Sezon sztormowy ($n = 28667$)					Sezon bezsztormowy ($n = 20540$)				
	Sr-4Q	Sr + 4Q	Number of records flagged	Percentage of records flagged (%)		Sr-4Q	Sr + 4Q	Number of records flagged	Percentage of records flagged (%)
H_s	-1.88	4.67	41	0.14	H_s	-1.48	3.04	101	0.49
T_m	0.52	7.32	2	0.01	T_m	0.49	6.1	4	0.02
T_p	0.24	11.24	5	0.02	T_p	-0.2	9.76	1	0.00

Table 9

Number of flagged measurements in the Dixon's 4σ result at the Pomeranian Bay point

Sezon sztormowy ($n = 9452$)					Sezon bezsztormowy ($n = 8133$)				
	Sr-4Q	Sr + 4Q	Number of records flagged	Percentage of records flagged (%)		Sr-4Q	Sr + 4Q	Number of records flagged	Percentage of records flagged (%)
Hs	-2.19	4.13	67	0.71	Hs	-1.72	3.6	17	0.21
Tp	-2.07	9.89	76	0.80	Tp	-1.37	9.21	83	1.02

Table 10

The results of outlier identification at coastal stations—Dixon's 4σ

Name of the stations	Measurement date	Season	The temperature measurement result	Dixon's 4σ		
				Control range		Assessment result
				$\mu - 4\sigma$	$\mu + 4\sigma$	
Gdynia	01.06.1962	Summer	5.9	6.01	29.14	Attention
Gdynia	12.06.1978	Summer	5.5	6.01	29.14	Attention
Hel	01.06.1980	Summer	4.6	5.86	27.76	Attention
Międzyzdroje	10.06.1955	Summer	8.6	9.13	26.70	Attention
Puck	01.06.1976	Summer	9	9.09	29.42	Attention
Puck	17.06.1982	Summer	8.9	9.09	29.42	Attention
Puck	23.06.1982	Summer	8.6	9.09	29.42	Attention
Świnoujście	04.12.1960	Winter	8.4	-4.54	8.02	Attention

which means that it can be properly applied to any parameter.

This test is also used to identify gross errors. Therefore, low temperatures in summer and high temperatures in winter are particularly noteworthy in the above list. However, there are no measurements from the spring and autumn seasons.

The set of temperature measurements at open sea stations includes 40,651 results measured from 1959 to 2019. At both coastal and offshore stations, the seasonality of meteorological seasons was taken into account when calculating average values and standard deviation. . With the adopted assumptions, the test identified 23 outlier measurements among temperature measurements at open sea stations (Table 11).

In the case of open sea stations, a much greater temporal and spatial diversity of outliers can be observed. Low temperatures during the summer season continue to dominate and concern mainly measurements made <50 m.

3.3. Q-Dixon's test

3.3.1. Wave parameters

The Dixon test was distinguished by the high detection of significant wave height measurements

compared to the mean wave period. At the Petrobaltic point, 142 significant wave height records and 6 records each for the mean and peak wave periods were identified. Using the Dixon test at the Pomeranian Bay point, it was possible to flag 143 phallic height records and 143 mean wave period records.

3.3.2. Temperature measurements

The test was performed on 190,967 temperature results at coastal stations. Taking into account the limitation of the test resulting from the maximum number of tested results (40), each month in which measurements were made over the years 1946–2019 was analysed separately. The results were categorised into ranges: from 3 to 7 measurements per month, from 7 to 12, and above 12 measurements to use the right Q-Dixon test formula to calculate the critical values of Q_1 and Q_n and to assign the right critical value of Q_{kr} . This test showed the presence of 282 outliers in the analysed data set.

3.4. Hampel test

3.4.1. Wave parameters

The Hampel test at the Pomeranian Bay point detected 143 outliers for significant wave height and the



Table 11

The results of outlier identification at open sea stations—Dixon's 4σ

Name of the stations	Measurement date	Season	Depth	The temperature measurement result	Dixon's 4σ		
					Control range		Assessment result
					$\mu - 4\sigma$	$\mu + 4\sigma$	
L7	28.05.2007	Spring	2.5	15.5	-5.9	14.9	Attention
L7	28.05.2007	Spring	5	15.6	-5.9	14.9	Attention
P1	07.08.1961	Summer	50	13.4	-3.9	12.5	Attention
P1	17.10.1990	Autumn	70	12.8	-2.1	11.5	Attention
P110	29.07.1984	Summer	68	12.5	-2.3	11.0	Attention
P110	10.09.1985	Autumn	68	15.4	-3.9	15.2	Attention
P110	24.07.1990	Summer	60	16.6	-6.2	14.8	Attention
P110	09.07.1993	Summer	68	14.3	-2.3	11.0	Attention
P110	12.08.2005	Summer	60	14.9	-6.2	14.8	Attention
P110	07.09.2006	Autumn	67	16.6	-3.9	15.2	Attention
P116	27.09.1978	Autumn	70	13.1	-2.0	11.9	Attention
P116	29.07.1984	Summer	50	13.4	-4.2	12.9	Attention
P116	09.07.1993	Summer	50	13.7	-4.2	12.9	Attention
P116	09.07.1993	Summer	60	13.4	-3.1	10.5	Attention
P116	09.07.1993	Summer	70	12.2	-1.1	9.0	Attention
P116	07.09.2006	Autumn	60	17.1	-5.9	16.6	Attention
P140	06.08.2007	Summer	40	9.9	-0.5	9.5	Attention
P140	31.05.2016	Spring	2.5	13.8	-5.7	13.6	Attention
P140	31.05.2016	Spring	5	13.8	-5.7	13.6	Attention
P140	31.05.2016	Spring	10	13.8	-5.7	13.6	Attention
R4	28.05.2007	Spring	1	17.0	-7.0	16.5	Attention
R4	28.05.2007	Spring	2.5	17.0	-6.4	15.3	Attention
R4	28.05.2007	Spring	5	16.9	-6.4	15.3	Attention

same number for the mean wave period, constituting 0.81% of the total data set. At the Petrobaltic point, representing deep water conditions, there were many fewer flagged measurements compared to shallow water conditions, that is, the Pomeranian Bay, and they amounted to 68 measurements each for significant wave height and mean wave period.

3.4.2. Temperature measurements

The Hampel test was conducted for all temperature measurements taken at coastal stations. As in the case of the Q-Dixon test, temperature measurements were analysed separately for each month. Among 190,967 temperature results, this test identified 7,649 outlier measurements. It is characterised by the highest sensitivity to the presence of outlier measurements in the analysed data sets.

3.5. Quartile test

3.5.1. Wave parameters

By dint of the quartile test, it was possible to flag the largest number of records in the shallows of the Pomeranian Bay, that is, 259 measurement records for significant wave height and the same number for the mean wave period. In deep water conditions, this test showed 121 records for both significant wave height and mean wave period.

3.5.2. Temperature measurements

The quartile test was also applied to all temperature measurements at coastal stations, analysing each month separately. This test showed the presence of 931 outlier measurements.

3.6. Gradient test

3.6.1. Temperature measurements

The test was applied to 40,651 temperature results taken between 1959 and 2019 at 18 open sea stations. At the test limit value (9°C), one outlier measurement was identified (station P5, 1965-04-06, level 20). At the same time, it should be noted that this test, in its formula, skips the measurement made at the first and last levels due to the lack of possibility of comparison to the values at adjacent levels. Therefore, it was possible to test 31,590 of the 40,651 total results.

Also, in the case of this test, measures were taken to adapt it to the conditions of the southern Baltic Sea. In the first step, the formula of that test was changed in this way: the measurement from the first level was compared only with the measurement made at the lower level, while the measurement made at the last level was compared with the measurement made at the higher level. In the case of measurements made at intermediate levels, the previous formula was used. The changed test formula is as follows:

$$\begin{aligned} \text{Measurement at the first level} \quad \text{Test value} &= |V_1 - V_2| \\ \text{Measurement from the} \\ \text{second level—previous formula} \quad \text{Test value} &= \left| V_2 - \frac{V_3 + V_1}{2} \right| \\ \text{Measurement at the last level} \quad \text{Test value} &= |V_3 - V_2| \end{aligned} \quad (7)$$

where

V_2 —measurement from level n

V_1 —measurement from level $n-1$

V_3 —measurement from level $n+1$

After calculating the test statistics for the temperature results according to formula (7), the 99th percentile was calculated as 4.7. The test results indicate that 99% of the differences between adjacent temperature measurements are $\leq 4.7^\circ\text{C}$. This result was mathematically rounded to 5, and this value was adopted as the test limit value for temperature measurements performed at southern Baltic open sea stations (Table 12). At the same time, the indicated limit value is used to flag suspicious values whose correctness should be subjected to expert assessment.

The second test, with the limit lowered to 5°C , conducted on 40,651 temperature results indicated 304 outlier measurements.

3.7. Conclusion

3.7.1. Wave parameters

The conducted tests (spike, Dixon, Hampel, and quartile) for significant wave height and wave period in two different locations, representing deep water (Petrobaltic) and shallow water (Pomeranian Bay) conditions, enabled the assessment of the applicability of the analysed tests in the southern Baltic Sea (Fig. 5).

A mutual comparison of the convergence of detection and flagging of identical measurement records by all tests for significant wave height (Table 13) and mean wave period (Table 14) allowed for the estimation of the applicability and reliability of the analysed tests. The highest number of overlapping flagged wave height records was the Spike Test and the Hampel Test, which detected only three less measurement records. The Dixon test in both locations (Petrobaltic, Pomeranian Bay) flagged 226 records of significant wave height measurements, of which as many as 117 records were flagged by the quartile test

- The Hampel test showed 211 outlying measurements of significant wave height, of which 207 out of 211 also include a quartile test.
- The quartile test, which flagged a total of 380 records, showed consistency with other tests with the number of the same flagged measurements: 118, 117, and 207 for the spike test, Dixon test, and Hampel test, respectively (Table 13).

The spike test for the mean wave period flagged 484 records, with 377 of the same measurements were flagged by the quartile test. The Dixon test, like the Quartile test, contains the most flagged records. The Hampel test flagged 211 records, of which the quartile test flagged 209 of the same measurements. Notably, 377 of the same records out of 380 were flagged by the quartile test and the spike test (Table 14).

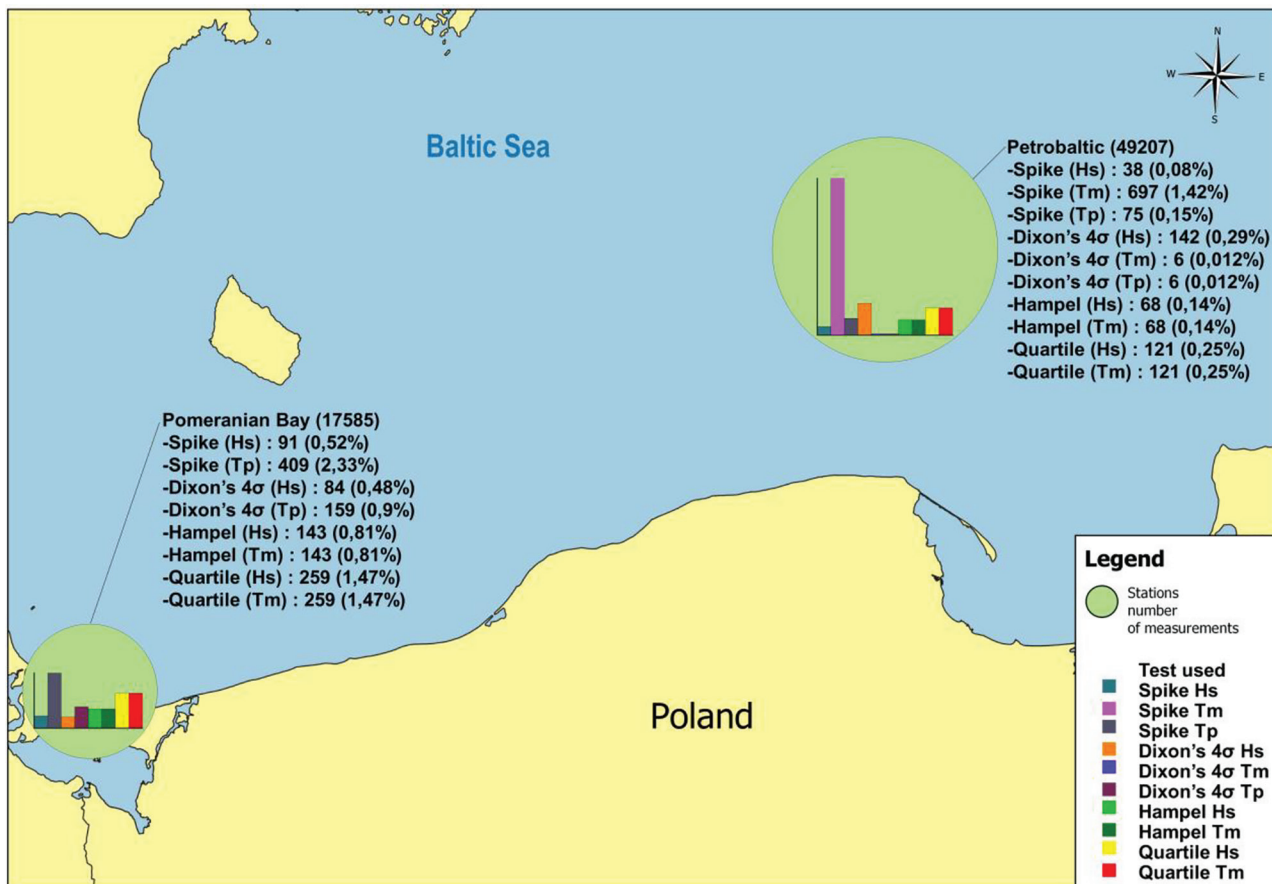
The example period, December 2018–January 2019, shows (Fig. 6) the identification of records flagged by one or more tests. It should be emphasised

Table 12

Proposed gradient test limits for temperature results at southern Baltic open sea stations

	Limit value for the test	99th percentile score	Proposed limit value	Number of records flagged	Percent of records flagged (%)
Temperature measurements at open sea stations	9	4.7	5	304	0.75



**Figure 5**

Number of flagged measurement records using the analysed tests at both locations.

Table 13

The total number of outlier measurements for the tests performed for the significant wave height parameter

Test	Spike	Dixon's 4σ	Hampel	Quartile
Spike	129	38	126	118
Dixon's 4σ	38	226	29	117
Hampel	126	29	211	207
Quartile	118	117	207	380

that flagging values are not rejected or removed values, but they only narrow the data set and designate data records that will be further subjected to expert evaluation. All extremely low values, where the difference from the previously recorded measurement exceeds 1 m, and in some cases even 5 m, were flagged by four tests: spike, Hampel, Dixon, and quartile. Between January 2 and 4, 2019, some records were flagged by three tests (spike, Hampel, quartile), two tests (spike and quartile), and one test (spike).

Table 14

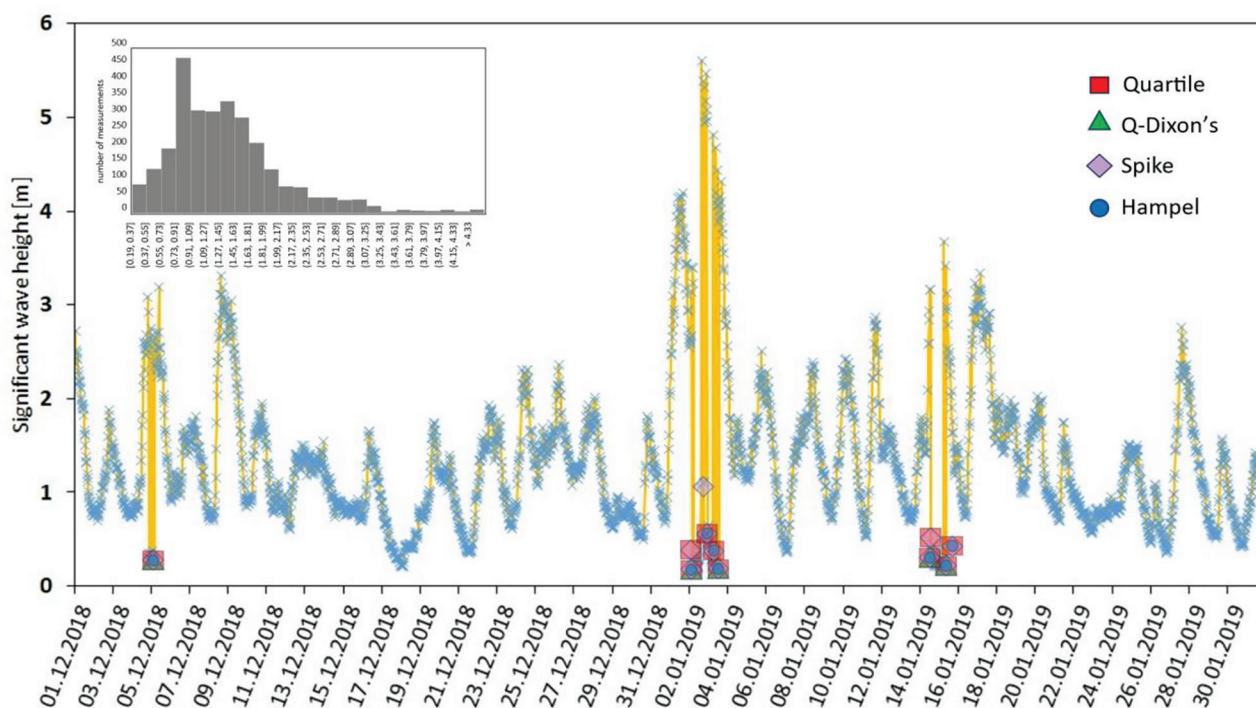
The total number of outlier measurements for the tests performed for the mean wave period parameter

Test	Spike	Dixon's 4σ	Hampel	Quartile
Spike	484	47	198	377
Dixon's 4σ	47	165	151	162
Hampel	198	151	211	209
Quartile	377	162	209	380

3.7.2. Temperature measurements

Figure 6 shows the station's location in the coastal zone. The bar chart indicates the type of test and the number of outlier measurements, while the size of the circle shows the number of measurements made at a given station in the analysed period from 1946 to 2019.

Analysing the above figure, it can be concluded that most outlier measurements were found based on the Hampel test. Based on this test, the number of outlier measurements identified ranges from 695

**Figure 6**

Significant wave height measurements with marking of flagged measurements from December 2018 to January 2019 for the Petrobaltic point.

at the Gdańsk station to 1,277 at the Władysławowo station. The least outlier measurements were found based on the Dixon's 4σ test, only eight. This test's results coincided with the Hampel test results only in three cases and with the results of the spike test in two cases. Among the stations presented in Fig. 7, the Kołobrzeg station is noteworthy, located on the central coast, where the largest number of outlier measurements indicated using the spike test were observed with the limit lowered to 2°C . This shows a wide variation in the temperature results measured here. The final data verification stage is an expert assessment of results indicated as outlier measurements. The phenomenon of upwelling should be considered among the factors that may influence the results of temperature measurements measured in the central coast region.

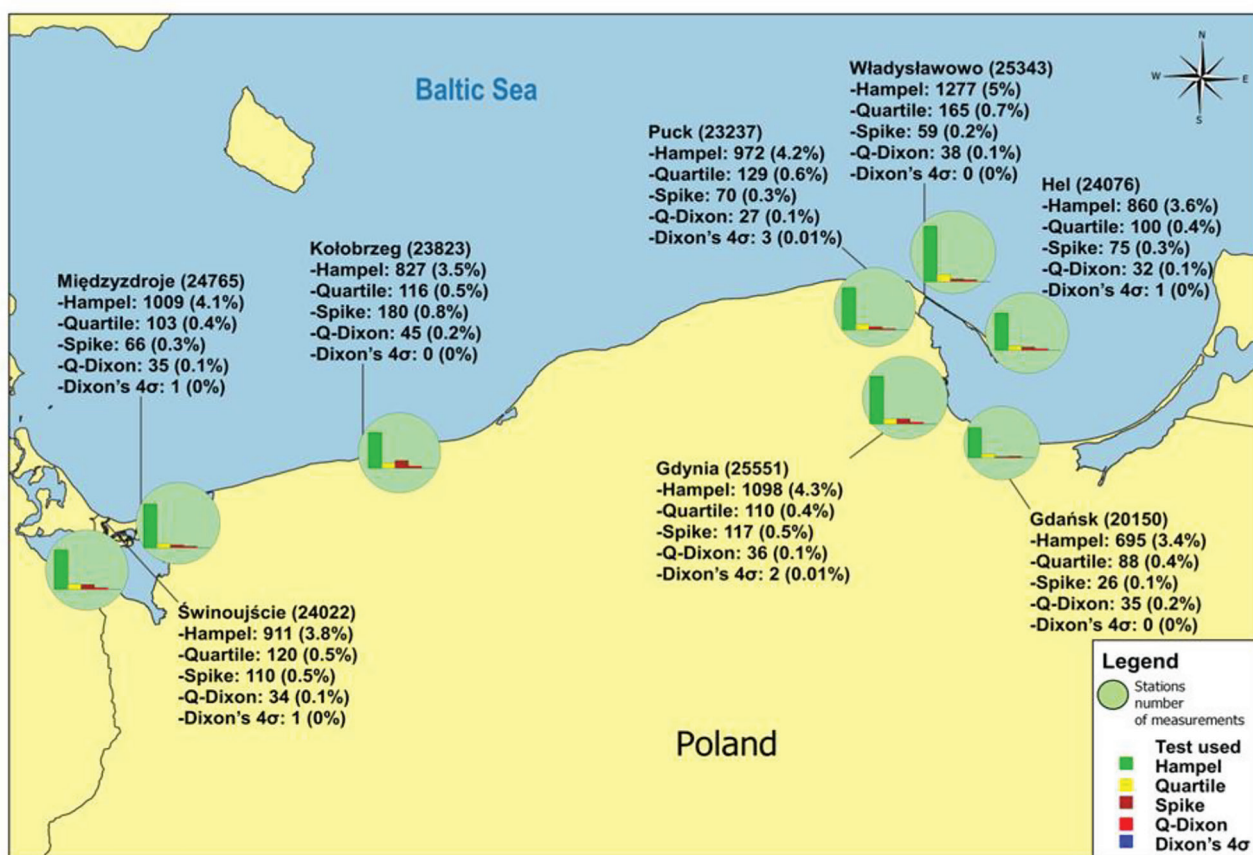
Upwelling is the phenomenon of rising ocean or deep-sea waters, usually from a depth below the thermocline to the surface due to the influence of strong along-shore winds. These winds, according to Ekman's theory, cause surface water to flow away from the shore and replace it with water from lower layers (Leppäranta and Myrberg, 2009). On the Polish coast, the areas most exposed to the occurrence of upwelling are its central part, particularly around Kołobrzeg, Łeba, and the Hel Peninsula.

The temperature difference between the outflow of bottom water masses reaches up to 10°C . This phenomenon is especially visible in summer and occurs in various forms. However, the dominant ones are the so-called filaments (Gurova et al., 2013), which create longitudinal ribbons/tongues with a concentric structure, where the lowest water temperature values characterise the centre of the tongue due to the strongest frontal uplift of deep-sea water masses. This phenomenon is often sudden, causing significant daily fluctuations in water temperature. To assess the impact of upwelling on the water temperature, the measurement value from 1 August 2019 for the Kołobrzeg coastal station (13.3°C), which was flagged with the spike test, the quartile test, and the Hampel test, was analysed in detail. This measurement's difference compared to the previous day's value is 3.3°C .

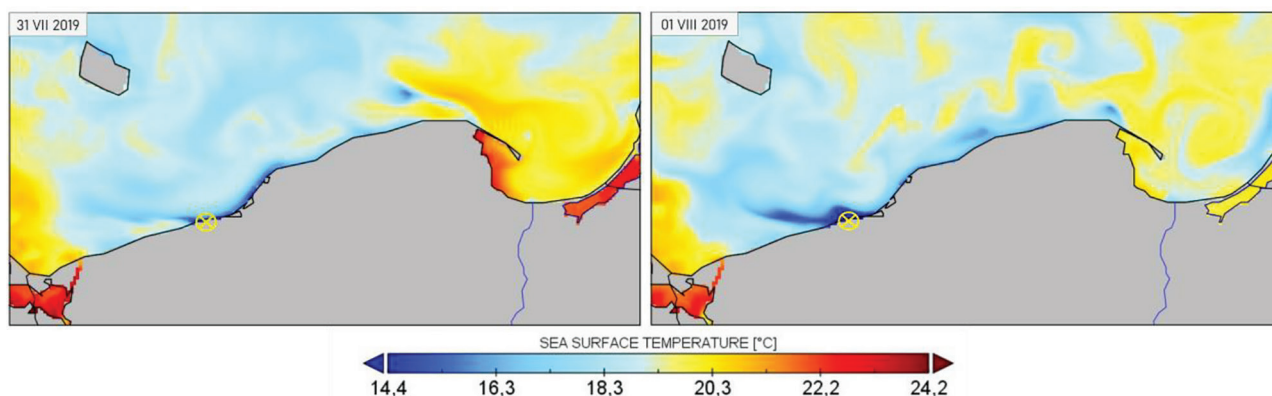
The spatial distribution of water temperature from 31 July to 1 August 2019, based on reanalysis data (Fig. 8), confirms the phenomenon of upwelling in the analysed period.

The temperature of the upwelling tongue is approximately 14°C – 15°C , which is equal to the temperature of the bottom water (Fig. 9). This indicates the upwelling of deep-sea water masses into higher layers.



**Figure 7**

Location of coastal stations taking into account the results of the data verification tests used.

**Figure 8**

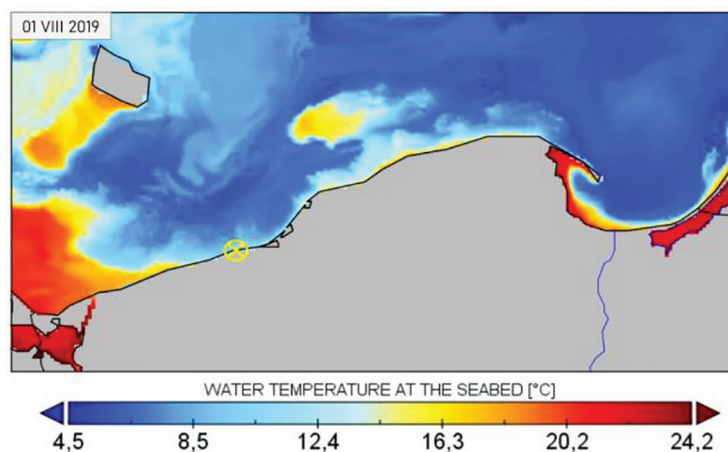
Surface water temperature from 31 July to 1 August 2019 in the southern Baltic Sea, based on (CMEMS, 2023).

Such a detailed analysis, therefore, gives grounds to question the mentioned temperature measurement result as an outlier measurement (Table 15).

In the next step, the impact of outliers on the basic statistical values of the analysed set of temperature measurement results at coastal stations (190,967),

such as the mean median and standard deviation, was analysed (Table 16).

With such a large data set, identifying and eliminating a relatively small number of outlier measurements from the set do not significantly impact the results of statistical analyses. However, the results

**Figure 9**

Surface water temperature from 31 July to 1 August 2019 in the southern Baltic Sea, based on (CMEMS, 2023).

Table 15

The number of outlier measurements common to the tests

Type of test	Spike	Dixon's 4σ	Q-Dixon	Hampel	Quartile
Spike	703	2	51	137	32
Dixon's 4σ	2	8	0	3	0
Q-Dixon	51	0	282	246	118
Hampel	137	3	246	7649	922
Quartile	32	0	118	922	931

of the Hampel test have a noticeable impact on the statistical values.

In the next step, the impact of outlier measurements was analysed for a small data set limited to one month. For this purpose, changes in statistical values in temperature measurements made at the Kołobrzeg station in August 2006 were analysed (Table 17). This is one of many examples where the spike, Hampel, Q-Dixon, and quartile tests indicated the same values as outliers.

By analysing the selected range of measurements for a single station, a clear impact of outlier measurements on the size of statistical analyses can already be observed. In that case, the identification and elimination of outliers have a significant impact, for example, on the standard deviation result. Eliminating a few outliers from the set of analysed values contributes to lowering the standard deviation value, making the set of results more homogeneous, which has a significant impact on monthly, annual, and long-term assessments and analyses.

The chart (Fig. 10) shows the identification of records flagged by one or more tests at the Kołobrzeg station in 2006. It should be emphasised that flagging

values are not rejected or removed values, but only narrow the data set and designate data records that will be further subjected to expert evaluation. Outliers were identified using the spike, Hampel, Dixon, and quartile tests. It can be seen that in August, there is a measurement flagged by four tests, while in September and November, by two tests. In the other months, we can see measurements detected by one test.

The percentage of flagged data varies depending on station location (Fig. 11). The bar diagrams indicate the type of test and the number of outlier measurements. At the same time, the size of the circle corresponds to the number of measurements made at a given station in the analysed period from 1959 to 2019.

In more than half of the analysed open sea stations, outlier measurements were found due to the gradient test using the test limit indicated in the guidelines (9°C). However, the percentage of these measurements is marginal compared to the total number of measurements performed at these stations. It ranges from 1 measurement at stations B13, R4, K, and L7 located in the transition zone to 2 measurements at stations P63, P2, P3, P110, P5, and P1 located mainly in the transitional and deep-water zones. The situation changes when the gradient test limit is lowered from 9°C to 5°C.

As a result of this procedure, the occurrence of outlier measurements is observed in almost all stations in the Polish marine areas of the southern Baltic Sea, except station B15. The largest percentage of outlier measurements related to the number of measurements taken is observed at stations located in the deep-water zone, such as P5, P3, P116, P140, and P63. In the case of open sea stations, it should be that in the summer, the Baltic Sea heats up quite quickly,



Table 16

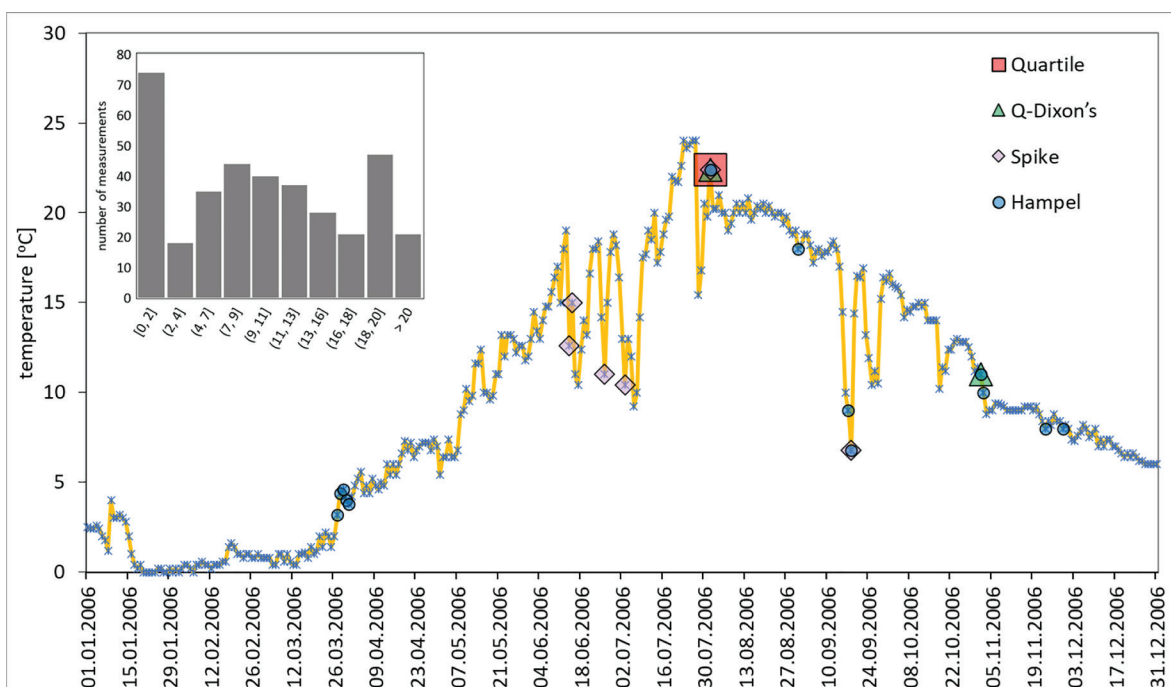
The impact of identifying outlier measurements on the results of statistical analyses in relation to temperature measurements at coastal stations in the years 1946–2019

Parameter	Test	Number of observations			Statistical values		
		All	Outliers	Used to calculate statistical values	Mean	Median	Standard deviation
Temperature of water in coastal stations	Zero trial	190,967	0	190967	9.47	8.80	6.64
	Spike		703	190264	9.46	8.80	6.64
	Dixon's 4σ		8	190959	9.47	8.80	6.64
	Q-Dixon		282	190685	9.48	8.80	6.64
	Hampel		7649	183318	9.56	9.00	6.61
	Quartile		931	190036	9.48	8.80	6.64

Table 17

The impact of identifying outlier measurements on the results of statistical analyses in relation to temperature measurements at the Kołobrzeg station in August 2006

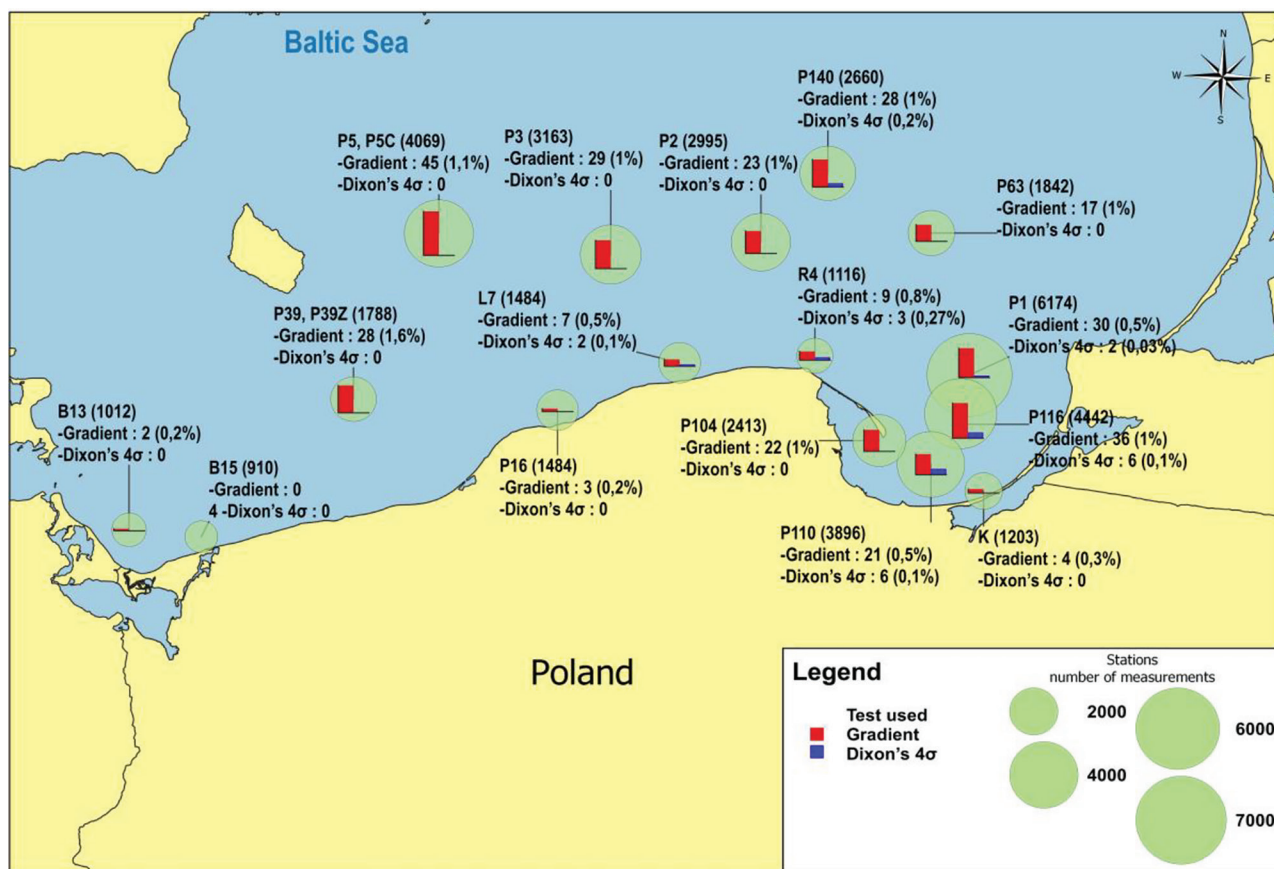
Parameter	Test	Number of observations			Statistical values		
		All	Outliers	Used to calculate statistical values	Mean	Median	Standard deviation
Temperature of water in coastal station Kołobrzeg in August 2006	Zero trial	31	0	31	19.96	20	0.76
	Spike, Hampel, Q-Dixon and Quartile		2	29	19.95	20	0.53

**Figure 10**

Temperature measurements with flagged measurements at the Kołobrzeg station in 2006.

but the heat is transferred into the water area slowly. As a result, a horizontal thermal stratification is created, where warm waters remain above the cold ones, and

a thermocline forms between these waters, a layer where temperature decreases rapidly with depth. This phenomenon may affect the verification test results at

**Figure 11**

Location of open sea stations taking into account the results of the data verification tests used.

Table 18

The impact of identifying outlier measurements on the results of statistical analyses in relation to temperature measurements at open sea stations from 1959 to 2019

Parameter	Test	Number of observations			Statistical values		
		All	Outliers	Used to calculate statistical values	Mean	Median	Standard deviation
Temperature of water in open sea stations	Zero trial		0	40.651	7.57	5.65	5.20
	Gradient	40.651	304	40.347	7.55	5.64	5.18

a globally set limit value. Also, in the case of open sea stations, the impact of outlier measurements identified by the gradient test on the result of calculating statistical values was analysed (Table 18).

As in the case of coastal stations, eliminating outlier measurements has a small but noticeable impact on the results of statistical analyses.

4. Discussion

The presence of outlier observations in the set of analysed values may have a significant impact on the

analysis result. Therefore, a good solution is to mark, remove, or, if possible, verify such observations. The appearance of outlier observations in the data set may result from natural factors (the above-mentioned upwelling phenomenon, thermocline). However, it often happens that outlier observations result from a malfunctioning measuring device or may be caused by the human factor due to incorrectly entering information into the database.

Global tests often do not meet the needs of regional conditions, which is reflected in the limits assigned to them. Due to the significant meridional extension of the Baltic Sea, it covers various climate



zones, that is, from subpolar to temperate climate, which means that different, specific conditions characterise different water areas. Moreover, the Baltic Sea is characterised by varying degrees of water mixing with the North Sea, which determines the distribution of salinity and water temperature. Regardless of the cause, outliers should be identified and flagged or removed to ensure comparable, standardised, and scientifically documented databases that form the basis for reliable assessments and analyses. Marking outliers using available methods for identifying such observations should be an element of good practice in the database management process. The method indicated in the publication for determining the limit value of the spike and gradient tests for wave and temperature measurements in stations in the Polish zone of the southern Baltic Sea, based on the 99th percentile, is only one possible course of action. Considering the physical and geographical regionalisation of the Baltic Sea, it is advisable to investigate and indicate test limit values for the coastal, transitional, and deep-water zones, respectively. Identifying factors that provide the basis for a deeper analysis to determine the limits of tests adapted to regional conditions makes it necessary to move away from their universal character towards a direction that can be described as single function, adapted to a specific region or even a specific measurement station. This approach will have a significant impact on the precision of the test in detecting suspicious values. Even when using such a universal test as Dixon's 4σ , seasonal variability of a parameter such as water temperature should be taken into account. Depending on the adopted methodology, the use of verification tests results in the selection of suspicious observations in the analysed data sets and limits them to such a number that is susceptible to human perception and enables their expert assessment and final qualification of measurements that may be considered outliers. The described data verification methods are a technical procedure; therefore, the final evaluation of test results and measurements indicated as outliers requires expert knowledge. If outlier observations are found, they must be marked and replaced using mathematical data interpolation methods or removed from the database.

References

- Baltic Sea Physics Reanalysis. E.U. Copernicus Marine Service Information (CMEMS). Marine Data Store (MDS). <https://doi.org/10.48670/moi-00014> [accessed December 14, 2023].
- Bitner-Gregersen, E. M., & de Valk, C. (2008). Quality control issues in estimating wave climate from hindcast and satellite data. *International Conference on Offshore Mechanics and Arctic Engineering*, June 15–20, 2008. Estoril, Portugal (Vol. 48197, pp. 819–827) <https://doi.org/10.1115/OMAE2008-57865>.
- Boyer, T. P., & Levitus, S. (1994). *Quality control and processing of historical oceanographic temperature, salinity, and oxygen data* (Vol. 81). US Department of Commerce, National Oceanic and Atmospheric Administration, National Environmental Satellite, Data, and Information Service.
- Budka, A., Kayzer, D., Pietruczuk, K., & Szoszkiewicz, K. (2013). Zastosowanie wybranych procedur do wykrywania obserwacji nietypowych w ocenie jakości rzek. *Infrastruktura i Ekologia Terenów Wiejskich*, (3/II).
- Castelão, G. P. (2015). A flexible system for automatic quality control of oceanographic data. <https://doi.org/10.1016/j.cageo.2021.104803>
- Chromiński, K., & Tkacz, M. (2010). Comparison of outlier detection methods in biomedical data. *Journal of Medical Informatics & Technologies*, 16, 89–94.
- Copernicus Team. (2017). Copernicus in situ TAC, real time quality control for WAVES. Copernicus in situ TAC.
- Copernicus Marine in Situ tac Data Management Team (2018). Copernicus Marine In Situ TAC - physical parameters list. Copernicus Marine In Situ TAC. <https://doi.org/10.13155/53381>
- Copernicus Team. (2020). Copernicus in situ TAC, real time quality control for WAVES [accessed: 28.08.2024]. CMEMS-INS-WAVES-RTQC. DATAWELL. 2020. *Datawell Manuals* [Online] [accessed: 28.08.2024]. <http://www.datawell.nl/Support/Documentation/Manuals.aspx> [accessed: 28.08.2024].
- Cummings, J. A. (2011). Ocean data quality control. *Operational oceanography in the 21st Century*, 91–121. (Chapter), Springer Nature, ISBN : 978-94-007-0331-5
- Doong, D. J., Chen, S. H., Kao, C. C., Lee, B. C., & Yeh, S. P. (2007). Data quality check procedures of an operational coastal ocean monitoring network. *Ocean Engineering*, 34(2), 234–246. <https://doi.org/10.1016/j.oceaneng.2006.01.011>
- FINO. (2020). The FINO project website [Online]. <https://www.fino-offshore.de/de/>. [accessed: 28.08.2024].
- Good, S. A., Martin, M. J., & Rayner, N. A. (2013). EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *Journal of Geophysical Research: Oceans*, 118(12), 6704–6716. <https://doi.org/10.1002/2013JC009067>
- Good, S., Mills, B., Boyer, T., Bringas, F., Castelão, G., Cowley, R., Goni, G., Gouretski, V., & Domingues, C. M. (2023). Benchmarking of automatic quality control checks for ocean temperature profiles and recommendations for optimal sets. *Frontiers in Marine Science*, 9, 1075510. <https://doi.org/10.3389/fmars.2022.1075510>

- GOSUD real time QC procedures, version 1.0. (2003). http://www.coriolis.eu.org/cdc/quality_control.htm
- GTSP real-time quality control manual, First Revised Edition. UNESCO-IOC 2010. (IOC manuals and guides No. 22, Revised Edition.) (IOC/2010/MG/22Rev.) by the United Nations Educational, Scientific and Cultural Organization 7, Place de Fontenoy, 75352, Paris 07 SP UNESCO 2010
- Gurova, E., Lehmann, A., & Ivanov, A. (2013). Upwelling dynamics in the Baltic Sea studied by a combined SAR/ infrared satellite data and circulation model analysis. *Oceanologia*, 55(3), 687–707. <https://doi.org/10.5697/oc.55-3.687>
- Ingleby, B., & Huddleston, M. (2007). Quality control of ocean temperature and salinity profiles – Historical and real-time data. *Journal of Marine Systems*, 65(1–4), 158–175. <https://doi.org/10.1016/j.jmarsys.2005.11.019>
- IOC. (1993). Manual of quality control procedures for validation of oceanographic data. *Manuals and guides* (Vol. 26). <https://unesdoc.unesco.org/ark:/48223/pf0000138825.locale=fr>
- IOOS. (2019). Manual for real-time quality control of in-situ surface wave data. In QARTOD (Ed. M. Bushnell), *A guide to quality control and quality assurance of in-situ surface wave observations*. Version 2.1
- Kennedy, J. J. (2014). A review of uncertainty in in situ measurements and data sets of sea surface temperature. *Reviews of Geophysics*, 52(1), 1–32. <https://doi.org/10.1002/2013RG000434>
- Lellouche, J. M., Le Galloudec, O., Drévilon, M., Régnier, C., Greiner, E., Garric, G., Ferry, N., Desportes, C., Testut, C.-E., Bricaud, C., Bourdallé-Badie, R., Tranchant, B., Benkiran, M., Drillet, Y., Daudin, A., & De Nicola, C. (2013). Evaluation of global monitoring and forecasting systems at Mercator Océan. *Ocean Science*, 9(1), 57–81. <https://doi.org/10.5194/os-9-57-2013>
- Leppäranta, M., & Myrberg, K. (red.). (2009). *Physical oceanography of the Baltic Sea*. Springer.
- Min, Y., Jeong, J., Shim, J. S., & Do, K. (2017). Quality enhancement of wave data observed by Radar at the Socheongcho Ocean Research Station. *Journal of Coastal Disaster Prevention*, 4(4), 189–196. <https://doi.org/10.20481/kscdp.2017.4.4.189>
- Morang, A. (1990). *Quality Control and Management of Oceanographic Wave-gage Data*. US Army Engineer Waterways Experiment Station.
- Namieśnik, J., Konieczka, P., & Zygmunt, B. (2007). Ocena i kontrola jakości wyników pomiarów analitycznych. *WNT, Warszawa*, 225–299.
- NORTEK. (2020). Manuals & quick guides [Online]. <https://www.nortekgroup.com/manuals-quick-guides> [[accessed: 28.08.2024].].
- Peterson, T. C., Vose, R., Schmoyer, R., & Razuvaev, V. (1998). Global Historical Climatology Network (GHCN) quality control of monthly temperature data. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 18(11), 1169–1179.
- RADAC. (2020). Documentation [Online]. Available: <https://radac.nl/documentation/> [accessed: 28.08.2024].].
- Salcedo Parra, O. J., Puente, F. J., & Ortiz, R. V. (2008) Oceanographic Data Quality Control By Means Of A Computational Tool. Vol. 13 No. 1/2008: January – June <https://doi.org/10.14483/23448393.2091>.
- SeaDataNet. Data quality control procedures. Version 2.0. May 2010. https://www.seadatanet.org/content/download/596/file/SeaDataNet_QC_procedures_V2_%28May_2010%29.pdf
- Svendsen, L. M., Barnich, J., Boutrup, S., Enne, P., Gustafsson, B., Frank-Kamenetsky, D., Haapaniemi, J., Knuuttila, S., Koch, D., Kokorite, I., Larsen, S. E., Oblomkova, N., Plunge, S., Raïke, A., Sokolov, A., Sonesten, L., & Ullrich, A. (2019). *HELCOM Guidelines for the annual and periodical compilation and reporting of waterborne pollution inputs to the Baltic Sea (PLC-Water)*. HELCOM. <https://helcom.fi/wp-content/uploads/2019/08/PLC-Water-Guidelines-2019.pdf>
- Team C.M.I.S. (2020). Copernicus in situ TAC, real time quality control for WAVES, Ref. CMEMS-INS-WAVES-RTQC, doi: 10.13155/46607
- Twardowski, K., & Trzpiele, J. (2006). Uwagi dotyczące wątpliwych wyników pomiarów. *Wiertnictwo, Nafta, Gaz*, 23, 699–714.
- Xie, J., Jiang, H., Song, W., & Yang, J. (2023). A novel quality control method of time-series ocean wave observation data combining deep-learning prediction and statistical analysis. *Journal of Sea Research*, 195, 102439. <https://doi.org/10.1016/j.seares.2023.102439>

