DOCUMENT CONTROL SHEET

Earth System Science Organization (ESSO)

Ministry of Earth Sciences (MoES)

Indian National Centre for Ocean Information Services (INCOIS)

ESSO Document Number: ESSO-INCOIS-ODM-TR-02(2022)

Title of the report: INCOIS-Real time Automatic Weather Station(IRAWS) dataset – Enhanced Quality control procedure

Author(s) [Last name, First name]:
Kameshwari Nunna, TVS Udaya Bhaskar, B.V. Satyanarayana
Originating unit
Ocean Data Management Division (ODM)

Type of Document:
Technical Report (TR)
Number of pages and figures: 18, 8
Number of references: 3
Keywords:
Marine-MET observations, IRAWS, Quality Control, time series
Security classification:
Open
Distribution:
Open
Date of publication:
March, 2022

Abstract (200 words) :

The INCOIS-Real time Automatic Weather Station(IRAWS) setups are installed over 30 ships. These setups communicate all the meteorological observations, namely, air temperature, air pressure, relative humidity, wind speed and direction, longwave radiation, shortwave radiation, rainfall, and sea surface temperature in real-time at hourly intervals and stored in the log at 10 minutes and 30 minutes interval. This report describes the enhanced quality control procedure that is carried out upon all the IRAWS datasets in delayed mode. The entire QC procedure is divided into primary QC and special QC procedures. Each QC step in the above procedures is vividly described and the effect that each QC step brings to the dataset is clearly explained. For this, a sample of 3 months of data at 10 minutes intervals observed onboard a ship is considered. The same data is used to explain each QC step and its effect throughout the report for better understanding. A comparison study is made between the same sample data that has been checked by the old QC procedure and the current QC procedure in order to highlight the improvements obtained by using the current QC procedure.

TECHNICAL REPORT **Report No.: ESSO-INCOIS-ODM-TR-02(2022)**

INCOIS-Real time Automatic Weather Station(IRAWS) dataset – Enhanced Quality control procedure

by

Kameshwari N, TVS Udaya Bhaskar, B.V. Satyanarayana

Indian National Centre for Ocean Information Services (INCOIS)

(Earth System Science Organization (ESSO),Ministry of Earth Sciences (MoES))

HYDERABAD, INDIA

www.incois.gov.in

March, 2022

Contents

List of figures

List of Tables

Abstract

The INCOIS-Real time Automatic Weather Station(IRAWS) setups are installed over 30 ships. These setups communicate all the meteorological observations, namely, air temperature, air pressure, relative humidity, wind speed and direction, longwave radiation, shortwave radiation, rainfall, and sea surface temperature in real-time at hourly intervals and stored in the log at 10 minutes and 30 minutes interval. This report describes the enhanced quality control procedure that is carried out upon all the IRAWS datasets in delayed mode. The entire QC procedure is divided into primary QC and special QC procedures. Each QC step in the above procedures is vividly described and the effect that each QC step brings to the dataset is clearly explained. For this, a sample of 3 months of data at 10 minutes intervals observed onboard a ship is considered. The same data is used to explain each QC step and its effect throughout the report for better understanding. A comparison study is made between the same sample data that has been checked by the old QC procedure and the current QC procedure in order to highlight the improvements obtained by using the current QC procedure.

IRAWS dataset - Enhanced quality control procedure

1.Introduction

INCOIS has started INCOIS-Remote Automatic Weather Stations(IRAWS), under the Ocean Observations and Information Services (OOIS) program of the Ministry of Earth Sciences (MoES), government of India in the year 2009. Currently, INCOIS is operating Automatic Weather Station installations onboard 30 ships. Table 1 of technical report by Kameshwari *et al* (2019)(ESSO-INCOIS-ODG-TR-03(2019)) provides the metadata of these AWS installations onboard all the ships and the Figure 1 shows the spatial distribution of the dataset. These systems transmit real-time data to INCOIS through INSAT satellites (Harikumar *et al*, 2013). The obtained real-time data is plugged into validation and data assimilation in NWP models used in ocean state forecasting(OSF) system at INCOIS, which is one of their key operational services. The surface winds so obtained from IRAWS setups are combined with other available ocean surface winds namely, ECMWF and NCMRWF which are used to force the wave models. The primary forecast in OSF system is ocean surface waves provided by the wave models and the most important input to wave models is surface winds. The current study focuses on the quality control of the data obtained from IRAWS before being used for any application.

Table 2 of Kameshwari *et al* (2019)(ESSO-INCOIS-ODG-TR-03(2019)) gives the details of various parameters observed onboard IRAWS setups. In the current report, Section 2 describes the common quality control procedure of all the parameters measured onboard the ship. Section 3 describes additional QC steps carried out separately for each parameter.

The primary aim of this technical report is to explain the quality control procedure applied upon the data observed from IRAWS setups in delayed mode. For the same, a sample of data observed onboard CRV Sagarpurvi during 1/1/2021 to 23/3/2021 at 10 minutes interval is considered and the time series of various parameters observed is shown in Figure 1. The plots shown represent the raw data obtained from the ship log. The current technical report can be considered as an extension of the QC procedures described in the report 'ESSO-INCOIS-ODG-TR-03(2019)' by Kameshwari *et al* (2019).

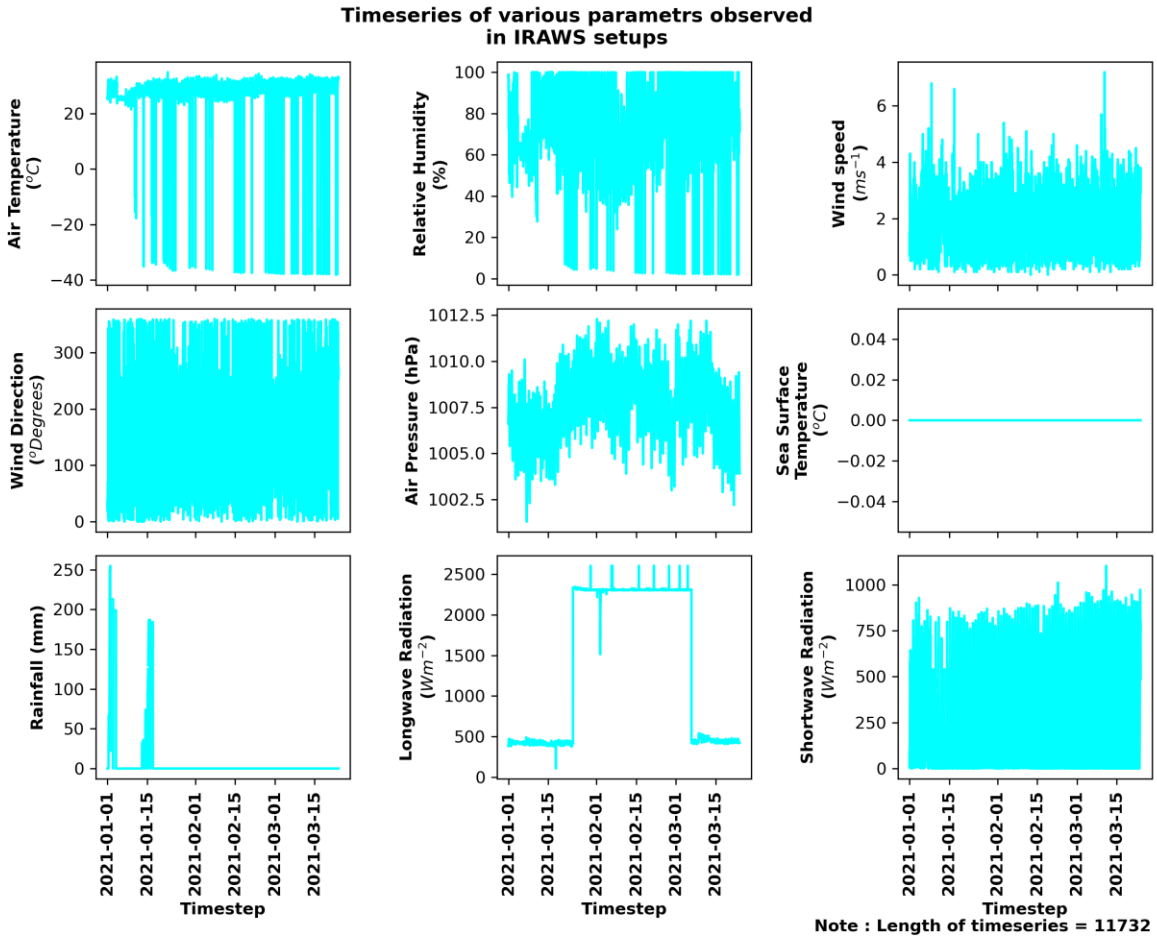

Figure 1 Time series of all parameters observed at IRAWS setup onboard CRV Sagarpurvi

2. Primary Quality Control :

Quality control(QC) procedures vary with the type of data, not alone with the parameter but also they vary with the duration of the data observed, they vary with the frequency of the observations, they vary with the type of platform used to measure the observations,etc. The current dataset is time series observed onboard moving platform with varying frequency, 10 minutes, 30 minutes and 1 hour observations. The data found erroneous in the QC procedure are flagged with a QC flag : '8' and are not removed/replaced, whereas the duplicated records are removed. Some QC tests provide a QC flag '4' which represents 'Data value is suspected-Use with precaution'. All the data values that pass all the QC tests are given a QC flag '1'. Currently,

only QC flags, ‘1’, ‘4’, ‘8’ are used in case of IRAWS dataset. Additional QC flags may be assigned with future enhancements in QC procedure.

The QC procedure starts with duplicate check followed by range check, climatological standard deviation check,etc. The duplicate check is performed first as the duplicates are removed and in all other checks, data is always retained and is only given QC flags. Since, the dataset is a time series, the QC procedure includes checks testing the data with respect to time stamps at which they are observed like spike test, stuck value test, etc. The following points describe each QC check in detail and explains how the data is flagged. Please note that this section covers the common QC procedure for a time series dataset.

2.1 Duplicate Check :

Duplicates can be classified into two categories :

1. Soft duplicates : In this case, the entire record of data transmitted is repeated i.e. both organizational data and the parameter observed is repeated. These kind of duplicates are very common and mostly occur during the initiation of the sensor after installing/restarting the setup. In some cases, transmission of records multiple times, is done on purpose to ensure that the data is received at the receiving site.

2. Hard duplicates : In this case, there exists data records, where, all the parameters measured are repeated but there are only slight differences in the organizational data. These records occur at a very less frequency. The chances are high during manual intervention at the observing site.

Note : Organizational data : Date describing the time and location of observation

Figure 2 shows the above time series shown in Figure 1 with duplicates removed. Please note the reduction in number of observations.

2.2 Range Check :

Both organizational data and parameters observed are checked to be with in the realistic ranges. Those values lying outside the realistic range are flagged with a QC flag ‘8’. The ranges defined are appropriate to the Indian Ocean basin and usually inlcude the limits that could occur

during extreme events. The table 4 of technical report by Kameshwari *et al* (2019)(ESSO-INCOIS-ODG-TR-03(2019)) gives the upper and lower bounds of various parameters. Except for few changes where the lower bound for SLP is changed to 970 hPa, the upper bound for SWR is changed to 1400 Wm^{-2}. Figure 3 shows the observations of each parameter that failed the range test. Another point to be noted is, the air pressure measured at the AWS station is recomputed to sea level using hypsometric formula.

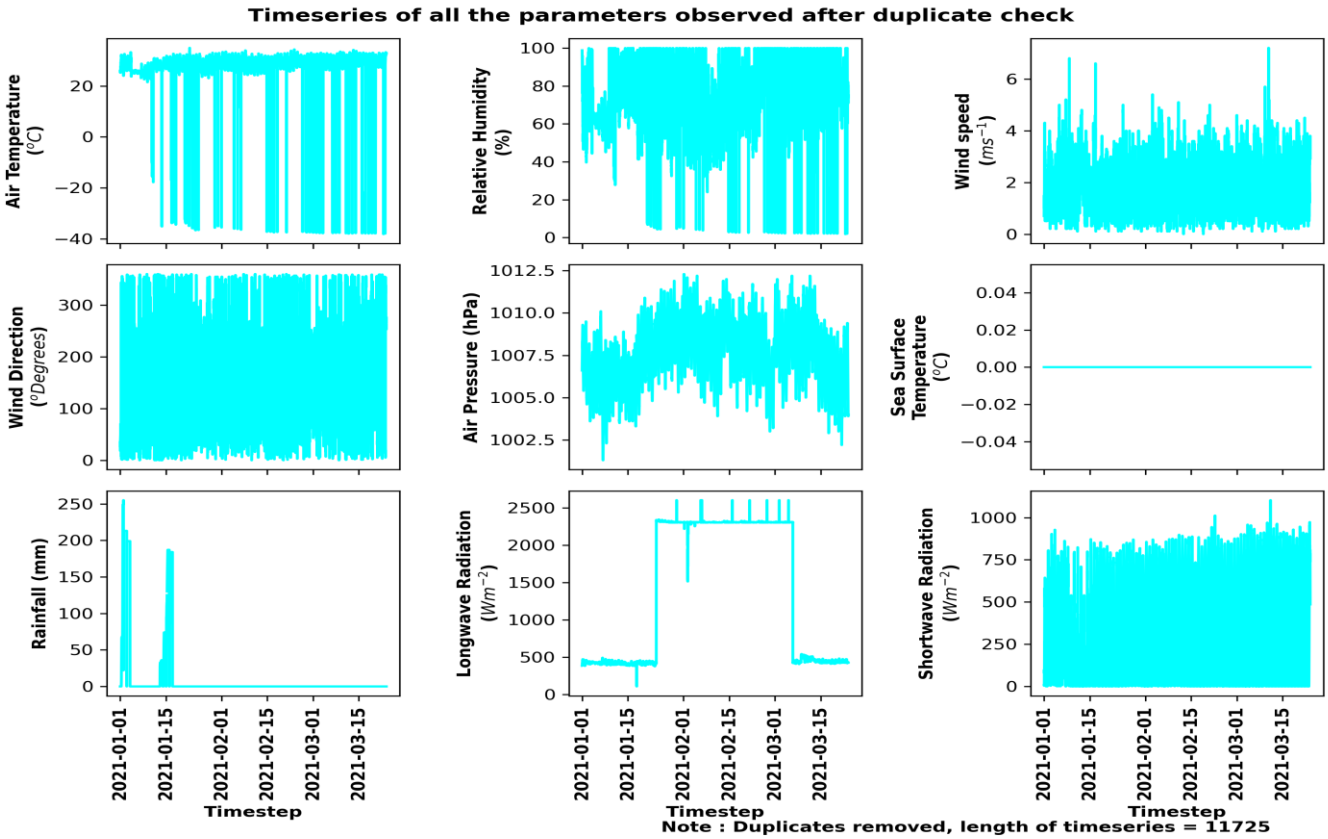

Figure 2 Time series of all parameters with duplicates removed

2.3 Climatological standard deviation check

This is a univariate check, where a parameter value is compared with climatological value corresponding to that location and month. The climatology considered is ICOADS R3.0 on a 1^o x 1^o grid and consists of data from 1960-2021. Both climatological mean and standard

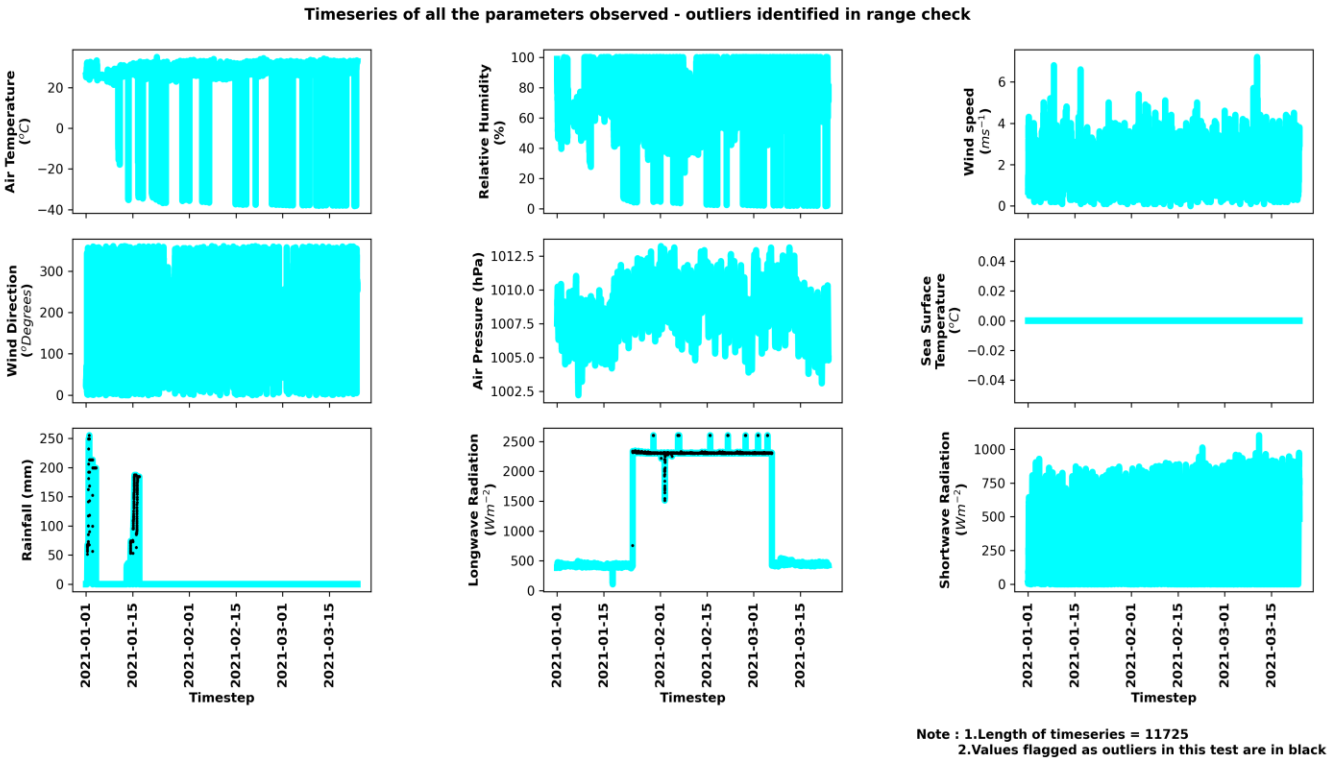

Figure 3 : Time series of all parameters with observations failed in respective range test marked in black color

deviation(STDEV) files of several parameters are used. In case of wind speed and wind direction, firstly, zonal and meridional wind speed are calculated and are compared to their respective climatologies. The QC procedure is as follows:

1. The monthly climatology of the month when the observation is made is first chosen.

2. The latitude-longitude limits which are 1^{o} away from the location of observation are defined. All the values existing within this grid limits are considered. And their corresponding mean and STDEV are calculated.

3. Then the climatological mean is subtracted from the data value and the deviation is calculated.

4. This deviation is determined to be within limits of ±3.5* STDEV is considered and all the values falling outside -3.5*STDEV and +3.5*STDEV are flagged with a QC flag ‘8’.

This check is applicable to the following parameters:

1. Sea surface temperature

2. Sea level pressure

3. Wind speed and wind direction (zonal and meridional wind speed)

4. Humidity

5. Air temperature

Climatological gridded fields of meteorological parameters are usually at 10 m height. Though the above parameters are observed at different heights, we are ignoring the small differences that would exist for the current dataset as stability based height correction is not feasible. However, the recomputed values at 10 m would only be slightly different from the original values, as the measurement heights are not very different from 10 m for most of the ships. Figure 3 shows the observations that failed climatological standard deviation check.

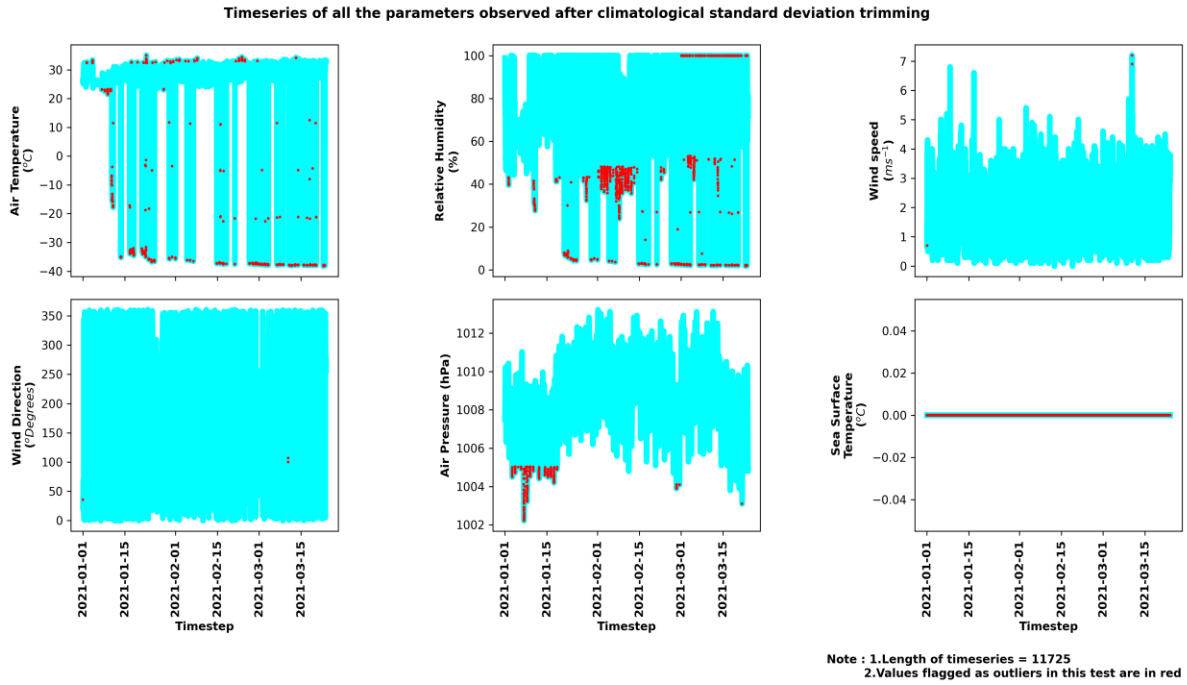

Figure 4 : Time series of all parameters with observations failed in climatological standard deviation test marked in red color

2.4 Spike test :

Spikes are sudden increase or decrease in parameter value in the time series at a particular time step and are characterized by large deviation from the rest of the values. These spikes could be well within the realistic range and will not be identified by the above QC checks. The spike test employed here is the standard deviation test and the method is similar to the one described in Kameshwari *et al* (2019). However, sometimes manual intervention is required to determine whether the particular value identified as spike is usual or not. Also, it is difficult to differentiate spikes in low temporal resolution data. Figure 5 shows the observation values identified as spikes.

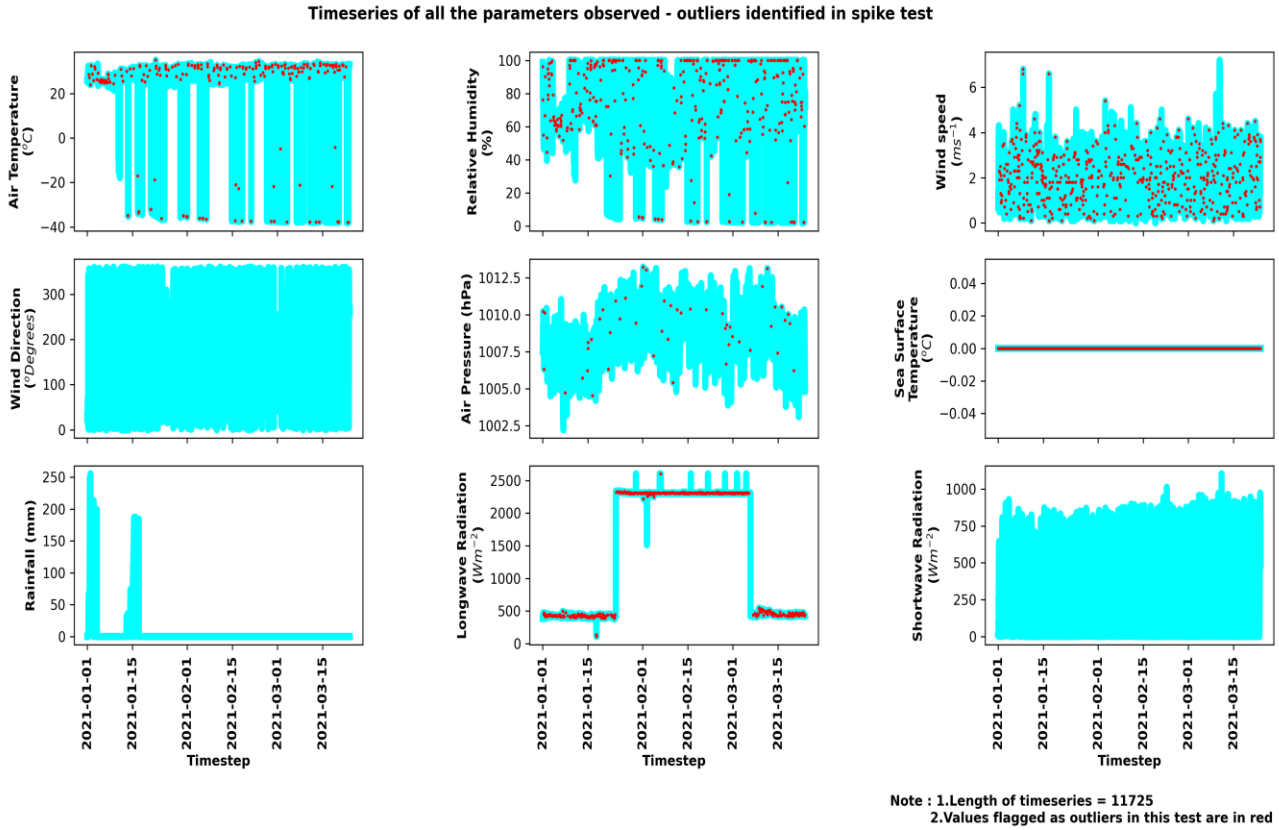

Figure 5 : Time series of all parameters with observations failed in spike test marked in red color

2.5 Stuck value test :

This test is also known as persistence check. It determines if instrument is transmitting a constant value without any change. The number of repetitions of the parameter value allowed depends on the parameter and the frequency of observation. Figure 6 shows the observation values that failed stuck value test. The below table describes the same:

Data Interval	Number of repetitions tolerated
10 minutes	6
30 minutes	4
60 minutes	3

Table 1 : Number of repeated values tolerated vary with the frequency of the observations

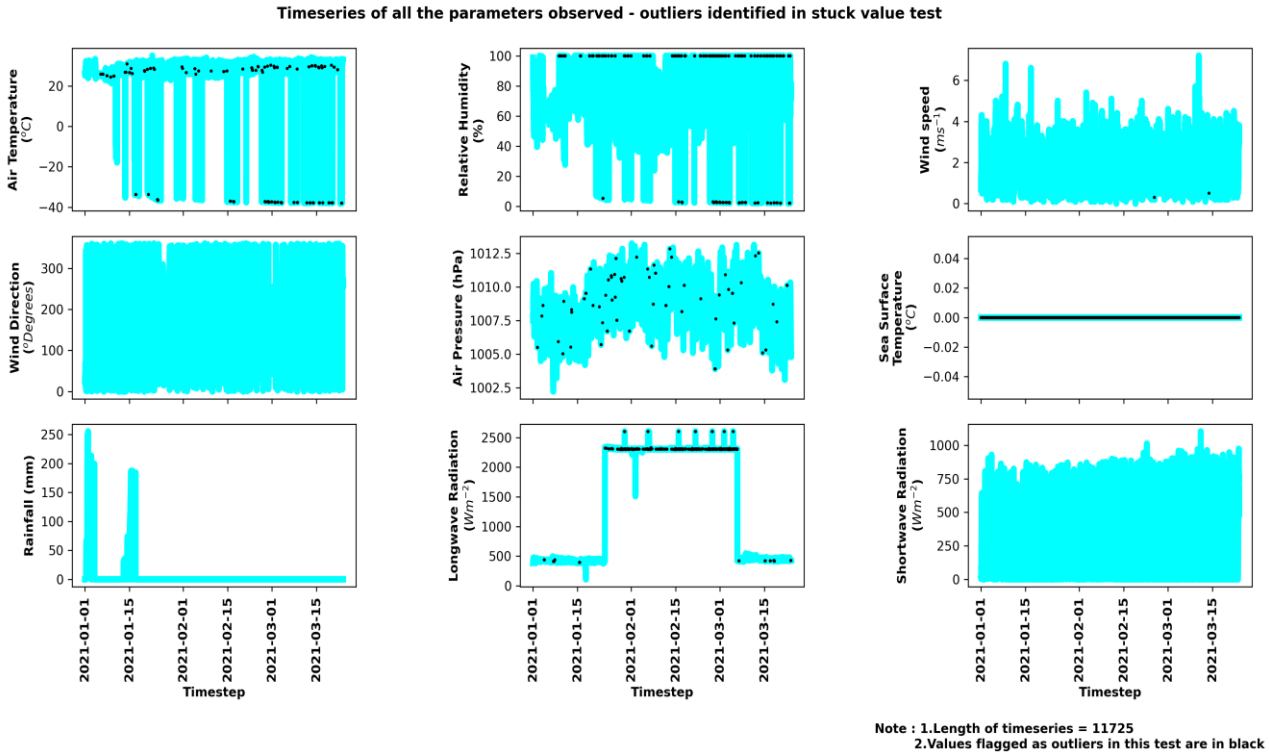

Figure 6 : Time series of all parameters with observations failed in stuck value test marked in black color

2.6 Time sequence check :

This check applicable to data observed onboard moving platforms simultaneously checks the speed of the platform and the possible subsequent location of the platform. The change in latitude and longitude in the given time difference is calculated and is checked to be within threshold values. The narrowing of longitudes with increasing latitudes is taken care of. If the position hasn't changed, then the ship is considered to be stationary. Values of all the parameters are of a record are flagged as '8' if the observation (organizational data) fails this test. Please note that stationary positions are not flagged as '8'.

3. Additional special tests:

This section describes additional tests which are particular to each parameter and hence the name 'special'. Several tests are performed upon each parameter of interest. This kind of additional exercise is required because, there will be cases, where the entire time series passes all the QC checks mentioned above but still the values do not seem to be accurate and this can be sometimes observed during manual QC check. Following are the list of special parameter based QC checks.

3.1 Internal Consistency Check:

Certain logical conditions comparing the two parameters in query are checked. If the condition fails, then all the parameters in relation with the two parameters checked are flagged '8'. Figure 7 shows the observations that are flagged in the internal consistency test.

3.1.1 Consistency between wind speed and direction:

The conditions checked in this test are, if wind direction is zero and wind speed is non zero, or if wind speed is zero and wind direction is non zero then all the parameters related to wind, i.e. wind speed, wind direction, zonal wind speed and meridional wind speed are all flagged as '8'.

3.1.2 Consistency between humidity and air temperature :

Usually if dew point temperature is measured then, air temperature could be compared with dew point temperature, however in the current dataset, humidity is being measured directly. To double ensure that the humidity measured is accurate, dew point temperature is calculated from the measured relative humidity and then compared with air temperature. If the dew point so calculated is more than air temperature, then both air temperature, humidity are flagged as '8'. Following are the formulae used to calculate dew point temperature from humidity and air temperature :

Firstly, saturation vapor pressure is calculated from air temperature. Relative humidity is fraction of vapor pressure in saturation vapor pressure multiplied by 100. Using both formulae and rearranging the terms, we arrive with the formula for dew point temperature as :

A1 = (17.5 * air_temperature) / (240.97 + air_temperature)

B1 = log(0.01*relative humidity)

Dew point temperature = ((A1+B1) * 240.97) / (17.5 - A1 – B1)

3.2 Parameter based special tests :

Some of these QC checks have been adopted from 'A Guide For Quality Control Of Surface Meteorological Data' prepared by Meteorology Group, Range Commanders Council.

1. Wind speed :

a. Wind speed measured >2 ms^{-1} and an increases or decreases in the wind speed is more than a factor of '2', such wind speed values are flagged as '4'

b. Wind speed increases from calm magnitudes to values greater than 1 ms^{-1}.

2. Wind direction :

All the consecutive records with wind speed greater than 2 ms^{-1} are grouped together.

a. The standard deviation of wind-direction values in each of the group is checked. If it is less than 1° or greater than 60° then all the wind direction in that group are flagged as '4'

b. If the change in wind direction between consecutive readings is more than 45^{o} within each group, then the respective wind direction values are flagged as '4'.

c. Considering groups of observations made within 2 hours, if the most frequently occurred wind direction value does not change between the groups, then all the wind direction values in that group are flagged as '4'.

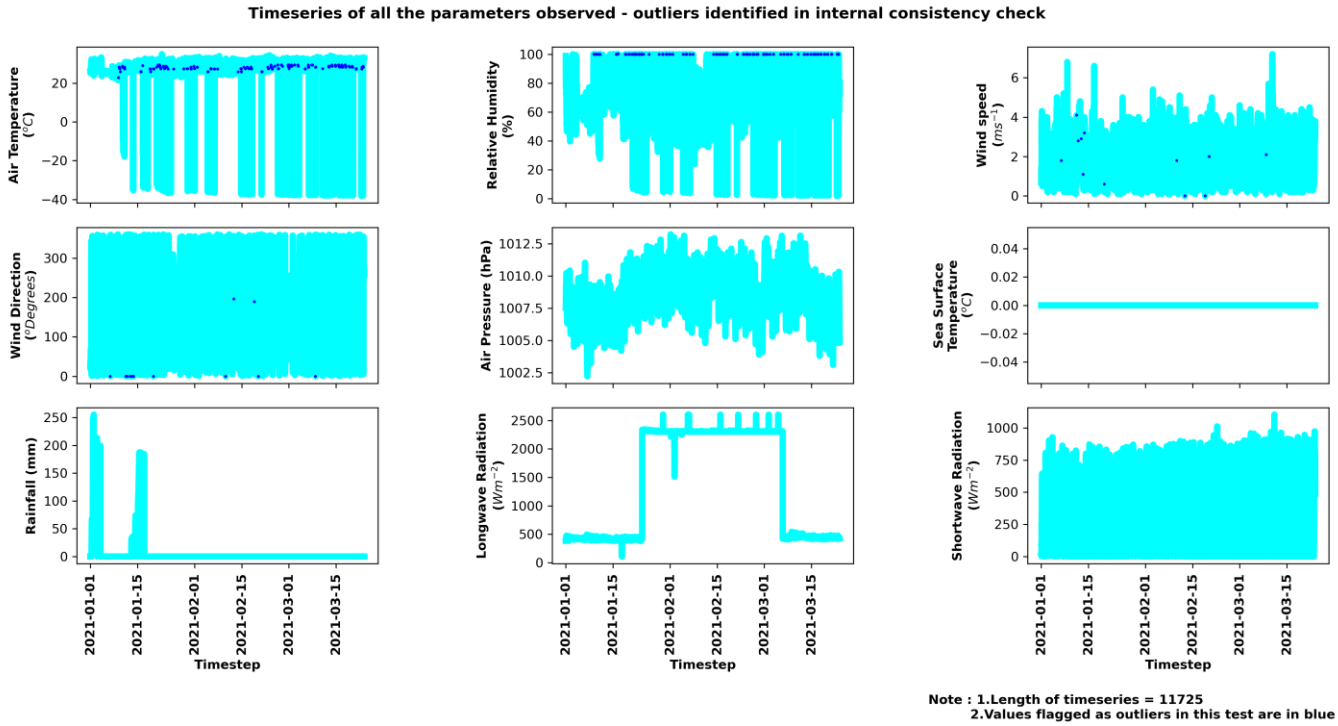

Figure 7 : Time series of all parameters with observations failed in internal consistency test marked in blue color

3. Air temperature :

a. The rate of change in temperature is greater than 10 °C per hour, then such values are flagged as '4'. For this test, both 10 minutes, 30 minutes data are first averaged to hourly data.

4. Humidity :

a. The rate of change in dew point temperature is greater than 10 °C per hour, then such values are flagged as '4'. For this test, both 10 minutes, 30 minutes data are first averaged to hourly data

b. The rate of change in humidity is greater than 30% per hour, then such values are flagged as ‘4’. For this test, both 10 minutes, 30 minutes data are first averaged to hourly data.

5. Sea level pressure :

a. The rate of change in sea level pressure is greater than 3 hPa per hour, then such values are flagged as ‘4’. For this test, both 10 minutes, 30 minutes data are first averaged to hourly data.

b. If the difference between the maximum and the minimum sea level pressure over a 24-hour period is greater than 15 hPa, then all those values are flagged as ‘4’.

6. Shortwave radiation :

a. If the shortwave radiation observed is zero during day time stamps, then such values are flagged as ‘4’. The observations are in UTC and the local time at that particular location(longitude) is calculated from UTC, longitude value. The day time stamps are considered from 06:00 hours to 19:00 hours local time.

b. If the shortwave radiation observed is greater than 100 Wm^{-2} during night time stamps, then such values are flagged as ‘8’. The night time stamps are considered from 19:00 hours to 06:00 hours local time.

Figure 8 shows the observations that are flagged in all of the parameter based special tests.

4. Results and Discussions :

To determine the betterment in the QC procedure, a simple comparitive study is done with the same dataset mentioned above. A comparison is made between the QC flags obtained for the dataset with the previous QC procedure descirbed in Kameshwari *et al* (2019)(ESSO-INCOIS-ODG-TR-03(2019)) and QC flags obtained for the dataset with the current QC procedure. Table 2 shows the number of observations flagged in the old QC procedure and the current QC procedure. It has been verified that the increase in flagged observations could be mostly attributed to the internal consistency and parameter based special checks.

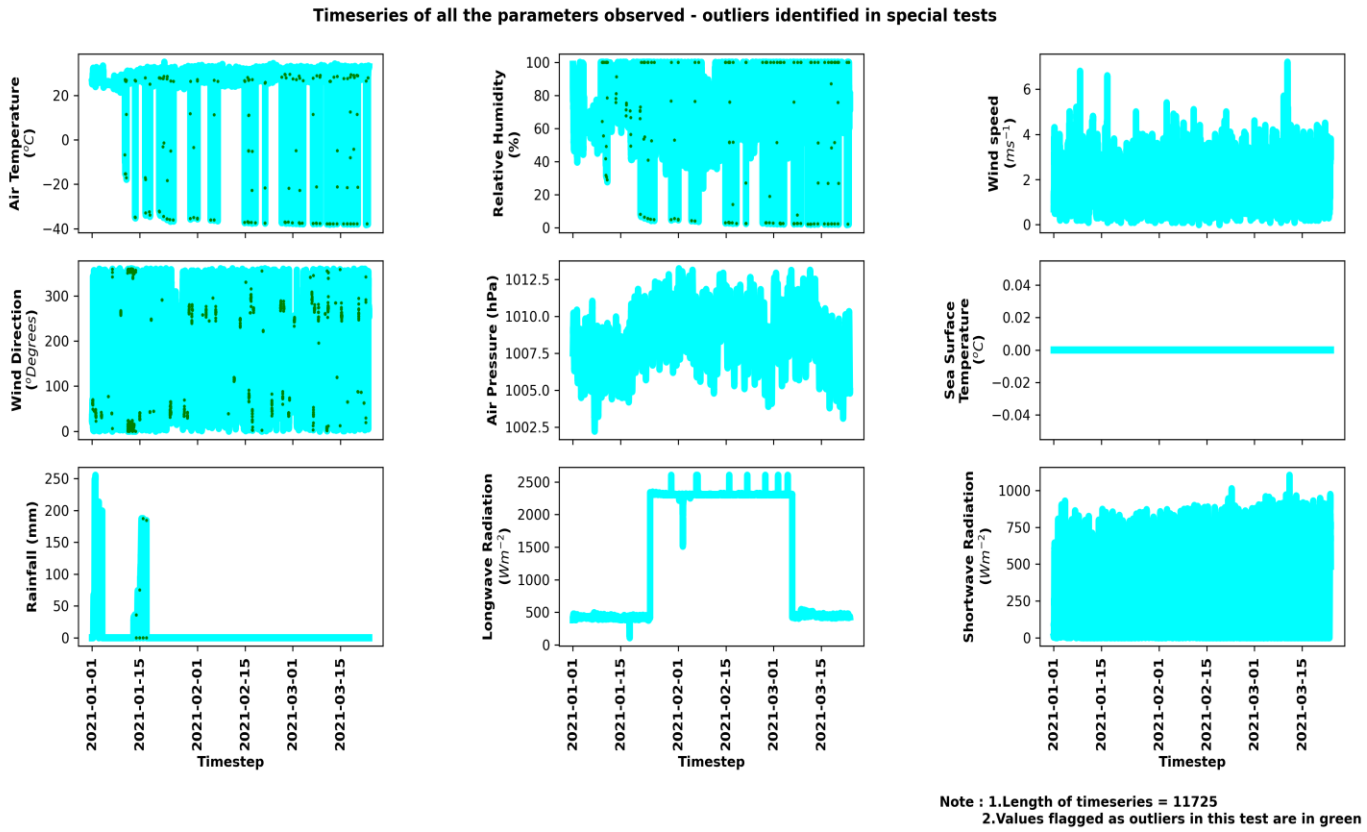

Figure 8 : Time series of all parameters with observations failed in internal consistency test marked in green color

Parameter	Number of values flagged by old QC procedure	Number of values flagged by current QC procedure
Sea level pressure	498	693
Air temperature	1565	1922
Relative humidity	3821	3874
Sae surface temperature	11729	11729

Wind speed	1011	1082
Wind direction	0	593
Rainfall	722	722
Longwave radiation	6333	6366
Shortwave radiation	0	230

Table 2 : Comparison between number of observations flagged in the old and current QC procedures

5. Conclusions and Future Scope:

INCOIS has installed IRAWS setups onboard Indian ships with the aim of obtaining real time data and use the same for validation purpose. Currently, the wind speed measured across all the IRAWS setups is being used in validating the model winds that are used to force wave models for forecasting the waves.

In the day to day operational activities of ODM, the QC percentages obtained for the various parameters observed onboard with these setups is around 90% in most of the cases except for few instances, where the sensor completely fails. However, as all the marine meteorological parameters are observed, it would be more advantageous if sea surface temperature is also measured with proper methods. This will ensure both increased scientific applicability and more accurate representation of the meteorological parameters (like height re computation to 10m).

As compared to the previous QC procedure, the current procedure described in this report is greatly enhanced and the result can be observed in the increased number of flagged data. A question may be raised as what could be the reliability of the data being flagged as there is always a possibility of good data being flagged. Furthermore, bad data being left out is always

tolerated rather than good data being flagged. Thus, to ensure that good data is not lost, the flags of such uncertain instances are given as ‘4’ instead of ‘8’ so that the end user can take a decision.

Additional QC checks specifically, for rainfall, shortwave radiation are being worked on and shall be included in the future. In case of shortwave radiation, a QC test to check the presence of sensor shadowing is to be developed. In case of rainfall, a check to the amount of rain (in mm) received in a stipulated time is also being worked on. Also, another internal consistency check between air temperature and sea surface temperature can also be carried out if the dependency on wind speed can be sorted out quantitatively for our Indian Ocean basin. These checks shall be included in the future.

The aim of enhancing the QC procedure is to increase the user reliability of the data from the user perspective and also to reduce the manual intervention from data managers perspective. Manual intervention is cumbersome when the quantity of data increases. Furthermore, in these lines, an attempt is being made to completely automate the QC procedure and to carry out the QC process using a graphical user interface.

References

1. Harikumar, R., et al. "Ship-mounted real-time surface observational system on board Indian vessels for validation and refinement of model forcing fields." *Journal of Atmospheric and Oceanic Technology* 30.3 (2013): 626-637.

2. Kameshwari, N., R. Venkat Shesu, A. Pasha, K. Jyothi, K. Suprit, E. P. Rama Rao, T. V. S. Udaya Bhaskar, and V. Jampana. "INCOIS-Real time Automatic Weather Station (IRAWS) dataset-Quality control and significance of height correction." (2019).(ESSO-INCOIS-ODG-TR-03(2019)).

3. “A Guide For Quality Control Of Surface Meteorological Data." (1994). Meteorology Group, Range Commanders Council