EMODnet Thematic Lot n°4 – Chemistry

EASME/EMFF/2018/1.3.1.8

Quality Control steps for EMODnet Chemistry Eutrophication aggregated datasets - v2021

Contents

Disclaimer

History

Authors	Date	Comments	Version
L. Buga, G. Sarbu, L. Fryberg, K. Wesslander, J. Gatti, S. Iona, M. Tsompanou, M. M. Larsen, A.K. Østrem, M. Lipizer, M.E. Molina Jack, A. Giorgetti	10/03/2021	Creation	0

How to cite this document:

L. Buga, G. Sarbu, L. Fryberg, K. Wesslander, J. Gatti, S. Iona, M. Tsompanou, M. M. Larsen, A.K. Østrem, M. Lipizer, M.E. Molina Jack, A. Giorgetti, 2021 Quality Control steps for EMODnet Chemistry Eutrophication aggregated datasets - v2021, 10/03/2021, 13 pages, 10.6092/4e85717a-a2c9-454d-ba0d-30b89f742713

Introduction

This document describes Quality Control steps procedure adopted to obtain the 2021 release of the standardised, harmonised and validated data collections concerning eutrophication (nutrients, chlorophyll and oxygen) per sea region (Mediterranean Sea, Black Sea, Arctic Region, Baltic Sea, Greater North Sea and North East Atlantic Ocean).

All data entering EMODnet Chemistry are managed according to standard protocols developed in the framework of the European consolidated SeaDataNet marine data infrastructure (https://www.seadatanet.org), and undergo a commonly agreed metadata and data validation process (Giorgetti et al., 2020).

An automatic Robot Harvester, properly configured with predefined criteria of geographical and temporal coverage and parameters, is adopted to retrieve specific data sets from distributed data centers to obtain data collections per sea region from the heterogeneous datasets originating from multiple institutions. Two types of data are available:

- vertical profiles (VP) for data that have been collected roughly at the same time and location for several consecutive vertical depths,

- time series (TS) for data collected at the same location and depth but repeated in time.

The resulting collections are **aggregated**, **harmonized** and **validated** by regional experts using ODV software and following a commonly agreed methodology, compiled in discussion with the wider international community (e.g. IOC/IODE, ICES and JCOMM, SeaDataNet Data Quality Control Procedures). As a result, quality flags are assigned to all data according to a standard scale (SeaDataNet measurand qualifier flags, L20 vocabulary, https://vocab.seadatanet.org/v_bodc_vocab_v2/search.asp?lib=L20). In case of data quality flag modifications, those are reported to data originators in order to validate the modification request and revise and correct original datasets (Fig. 1).

Parameter aggregation is carried out at Regional Sea scale by means of a dedicated vocabulary (P35 vocabulary: EMODnet Chemistry aggregated parameter names) implemented to combine various P01 terms associated to a same substance (e.g. nitrate in seawater) but measured with different protocols or expressed with different measurement units, into a unified aggregated term with a uniquely identified standard unit. In particular, **harmonization** consists in the conversion to standard units agreed within the EMODnet Chemistry scientific community. The ODV software (Schlitzer, 2020) has a built-in

aggregation procedure, making use of the P35 vocabulary and also applying a number of business rules, such as for averaging and possible unit conversions. The information related to the specific parameters (as described by the P01 vocabulary) is available as metadata, together with details about the acquisition instrument, if provided by data originators. **Data validation**, or Quality Control, is done according to the steps described in the following section (Specific Quality Control procedures per sea region).
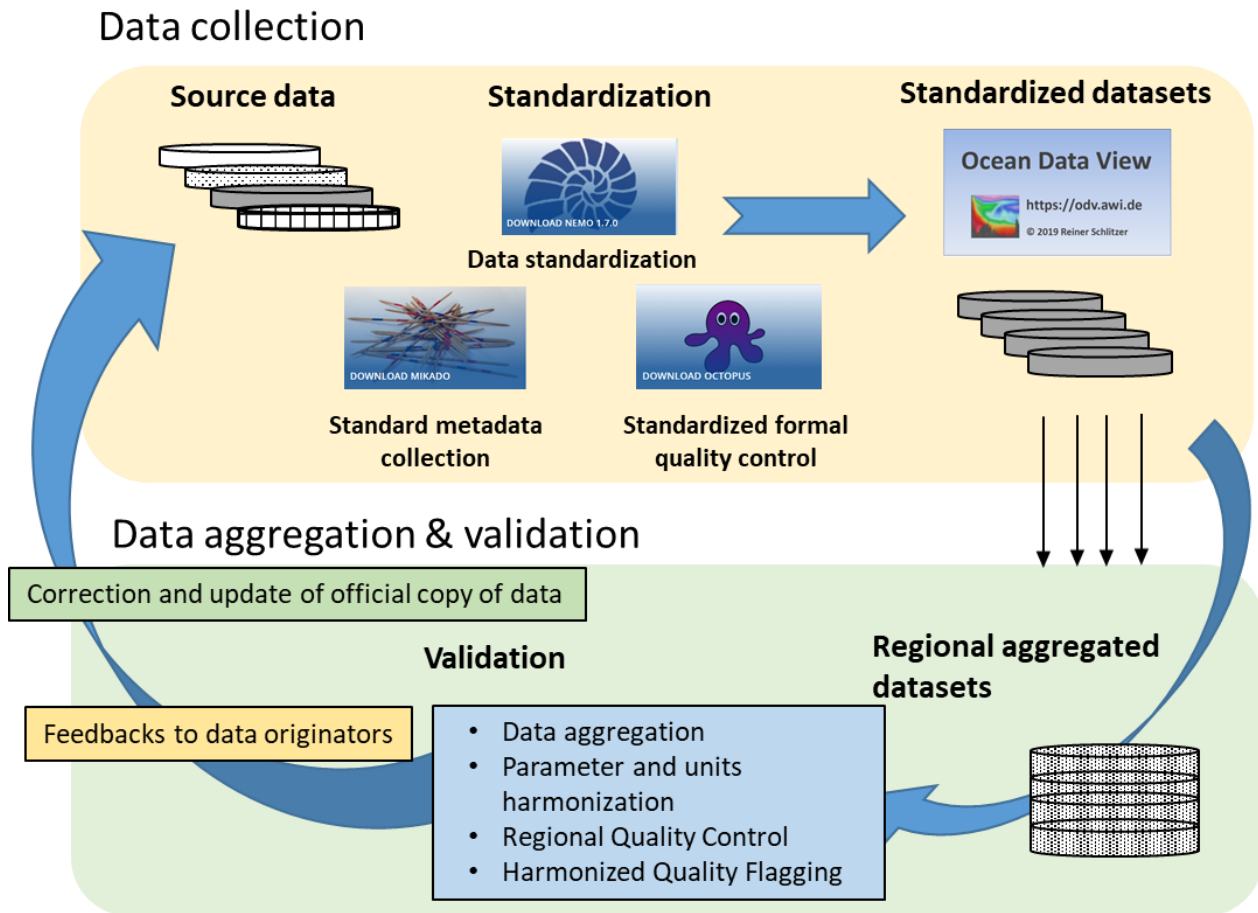

Fig. 1 The loop implemented within EMODnet Chemistry for data quality control, assures continuous data quality upgrade (from Giorgetti et al, 2020, modified).

Specific Quality Control procedures per sea region:

Mediterranean Sea

● Import ODV.txt files and create one collection for profiles and one for time series.

● Correct the ODV.txt files which are not imported because of a format error for empty lines.

● Import the corrected ODV.txt files

● Make the aggregation of the collections.

● Correct some stations with primary variable pressure in meters.

● Erase the P35 which are not contained in the collections from the first harvest.

● Exclude the stations outside the Mediterranean according to the new boundaries with the Atlantic (5° 55′ W) and send them to the regional leader of the area.

● Delete stations with Temperature and Salinity data only.

● Delete stations with no metadata.

● For negative values, change the QF from any value to QF=4.

● For default values such as 9999 or 999.999, change the QF from any value to QF=4.

● For extreme values out of "broad range", change the QC flags from QF=1 or QF=0 to QF=4. If outliers lie close to the "broad range" limits, change the QC flags from QF=1 or QF=0 to QF=3.

● For values equal to 0 change their flags from QF=0 or QF=1 to QF=6. In some cases that these values equal to 0 is obvious that are not values below detection limit the flag changed to QF=4.

● Constant profiles flagged to QF=4.

● Spikes detection and flag change from QF=0 or QF=1 to QF=3.

● "Good" profiles with QF=0 changed to QF=1.

● Keep the QF=0 for DO and DO saturation from EDMO 120, as in the previous fase of EMODnet III.

● Identify cases where N02>N03 and flag them. The check consists of the following steps:

- Calculate the derived variable "N03 – NO2" and find negative values.

o When the flag of NO3 and NO2 is QF=1, change both NO3 and NO2 flags to QF=3.

o   When the flag of NO3, NO2 is QF=6, keep both NO3 and NO2 flags to QF=6.

o   When the flag of NO3 is QF=6 and NO2 is QF=1, change the NO2 flag to QF=3.

o   When the flag of NO3 is QF=1 and NO2 is QF=6, change the NO3 flag to QF=6

●   Identify cases where NO3 plus NO2<NO3 and flag them. The check consists of the following steps:

-       Calculate derived variable N03 plus NO2 – NO3 and find negative values.

o   When the flag of NO3 plus NO2 and NO3 is QF=1, change both NO3 plus NO2 and NO3 flags to QF=3.

o   When the flag of NO3 plus NO2 is QF=6 and NO3 is QF=1, change the NO3 flags to QF=3.

●   We change the QF=0 for EDMO 120 in oxygen concentration, and oxygen saturation according to the instructions of the provider not only in the collection with new and updated data but also in the collection with the old data.

●   Calculate aggregated DIN and find the values equal to 0. Change these flags from QF=1 to QF=6.

●   Merge the collection with the old data with the collection with new and updated data.

●   Elimination of duplicates stations

Atlantic Ocean

●   Import ODV.txt files and create one collection for profiles and one for time series

●   Exclude trajectory data outside the NE Atlantic domain from the profiles collection.

●   Run the P35 aggregation using the Best-Of ODV algorithm.

●   Merge the ocean pressure and the ocean depth collections into one depth collection for profiles using Depth/Pressure conversion

●   Delete stations with Temperature and Salinity data only.

●   Delete stations with no metadata.

●   Delete duplicates stations.

●   Correct latitude and longitude of stations on land.

● Delete NRT Argo floats/profiles. They are considered as "wrong" profiles because they are profiles in Near Real Time (NRT) not adjusted and/or not validated.

● For negative values, change the QF from any value to QF=4 or QF=3 when data come from an uncalibrated instrument.

● For default values such as 9999 or 999.999, delete the data and change the QF to QF=9.

● For values equal to 0 change their flags from QF=0 or QF=1 to QF=6. Except for zero values from Russian data (EDMO 681) changed to QF=4 (historic data, impossible to know automatically if the zero values are wrong encoded missing value or LOD).

● For extreme values out of "broad range" , change the QC flags from QF=1 or QF=0 to QF=4. If outliers lie close to the "broad range" limits, change the QC flags from QF=1 or QF=0 to QF=3.

● Spikes detection and for only obvious flag change from QF=0 or QF=1 to QF=3.

● "Good" profiles with QF=0 changed to QF=1.

● Calculate aggregated DIN.

● Check for nutrients the ratios OP:TP and DIN:TN>110%. After inspection of the profiles, change the flag of the outlier parameter to QF=3 or change the flags of all the parameters associated (for DIN: Nitrate, Nitrite and Ammonium) if no obvious outlier parameter.

● Checks for oxygen supersaturation in surface open waters: the upper limit is capped at 115% saturation above 150 m, 130% above 100 m, and 150% above 10 m depth. Values above these upper limits are flagged as bad data (QF=4).

Black Sea

● Import ODV files from the 2nd harvest (new and updated) in ODV. It results in 2 collections: ocean_depth_profiles.odv and ocean_pressure_profiles.odv

● Merge the two collections and then aggregate the merged collection.

● Delete stations with Temperature and Salinity data only.

● Delete stations with no metadata.

● Delete duplicates stations.

● Delete stations on land.

● Delete stations with obvious wrong P06 (all data are far outside the range) if the stations are from the same cruise.

● For negative values (including depth), change the QF from any value to QF=4 (if not already)

● For values equal to 0 change their flags from QF=any to QF=6. In the cases when the values=0 were obvious not under detection limit, change the QF to QF=4.

● For extreme values out of "broad range", change the QC flags to QF=4 (if not already)

● Spikes detection and for, only obvious wrong data, change the QF to QF=4 or QF=3 (if not already).

● Change the QF of stuck data to QF=4.

● Change the QF for two or more identical consecutive depth values to QF=4.

● "Good" profiles with QF=0 changed to QF=1 or QF=2.

● Identify cases where N02>N03 and flag them. The check consists of the following steps:

    o   >> Calculate derived variable "NO3/NO2".

       If 0<NO3/NO2<=1, change the QF for both NO3 and NO2 to QF=3

● Identify cases where PO4>TP and flag them. The check consists of the following steps:

    o   >> Calculate derived variable "TP/PO4".

       If 0<TP/PO4<=1, change the QF for both TP and PO4 to QF=3

● Calculate aggregated DIN and find the values equal to 0. Change these flags from QF=1 to QF=6.

● Identify cases where DIN>TN and flag them. The check consists of the following steps:

    o   >> Calculate derived variable TN/ DIN

       If 0< TN/ DIN <=1 and QF=1 for DIN, change the QF of TN to QF=3

● Merge the collection with the 1st harvest aggregated collection.

● Delete new duplicated stations

● Checks for oxygen supersaturation in surface open waters: the upper limit is capped at 115% saturation above 150 m, 130% above 100 m, and 150% above 10 m depth. No values above these upper limits were found

    ➢ **Issue reports with results of all above steps, separately for each data originator (according to EDMO code) and send them by email.**

Arctic region

● Import new+updated ODV text files in ODV, creating 2 collections: ocean depth profiles and ocean pressure profiles.

- Merge the two collections to create one vertical profiles collection
- Delete profiles outside the defined region (south of 62 N)
- Aggregate the variables
- Delete variables that don't belong in the eutrophication collection (not aggregated, metals, particulate)
- Delete stations with Temperature and Salinity data only
- Check for duplicates (none found)
- Correct latitude and longitude of stations on land
- Some wrong bottom depths have been corrected
- For values outside broad range limits, flags were changed from QF = 0, 1 or 2, to QF=3 or QF=4. (Some partners have used QF=2 for bad data)
- Default values (-10, -9.98) with QF = 0 or 1 were deleted, and flags changed to QF=9
- Negative values (except negative temperatures) were flagged with QF=4
- "Good" profiles with QF=0 were changed to QF=1
- Zero values were flagged with QF=4 if outside broad range limits. In upper layers where values could be below LOD, flags are changed to QF=6
- Values below LOD were flagged with QF=6 when LOD was provided by data originator
- Visual checks – flags changed to QF=4 for the whole profile in case of
    - o   Decimal point errors – values 10 times higher or lower than normal
    - o   Bad profiles, mainly in dissolved oxygen concentration
- Check if ratio OP:TP > 110%. Flags were changed to QF=3 for both parameters if no obvious outlier parameter (very little TP data, difficult to decide which one is wrong)
- Spikes flagged with QF = 3 or 4
- Constant profiles were flagged QF=4
- For uncalibrated chl-a data, flags were changed to QF=3
- Calculated aggregated NO2+NO3 and DIN. For zero values, change flags from QF=1 to QF=6
- Import the new+updated collection to the old collection
- Check for duplicates

Baltic Sea

● Import the new and updated ODV.txt in ODV. This resulted following collections: ocean_depth_profiles.odv, ocean_minimum_depth_profiles.odv, ocean_pressure_profiles.odv and time_series.odv

● Aggregate the collections and merge all 'profiles' collections to finally create one collection for profiles and one for timeseries.

● Delete stations with Temperature and Salinity data only.

● Delete stations with no metadata.

● Delete duplicates stations.

● Delete stations on land (1 in the Baltic Sea).

● Exclude stations (4 in the Baltic Sea) that do not contain the relevant data variables (parameters for nutrients or acidification).

● Remove empty variables.

● Remove variables that shall not be used in this product (such as for example Secchi depth, contaminants).

● Negative values were flagged with QF=4 except for oxygen that was flagged with QF=6 (if not obviously wrong).

● Values=0 (zero) and with QF=0 or 1 were flagged with QF=6 if not obviously wrong, if so the flag 4 was given instead.

● For extreme values out of "broad range", change the QC flags from QF=1 or QF=0 to QF=4. If outliers lie close to the "broad range" limits, change the QC flags from QF=1 or QF=0 to QF=3.

● Spikes detection and flag change from QF=0 or QF=1 to QF=3.

● Calculate aggregated DIN and find the values equal to 0. Change these flags from QF=1 to QF=6.

● Ratio of inorganic nutrients / total nutrients. After inspection of the profiles with ratio > 1, change the flag of the outlier parameter to QF=3 or change the flags of all the parameters associated (for DIN: Nitrate, Nitrite and Ammonium) if no obvious outlier parameter.

● Data with QF=0 have been changed to QF=1 if no other flag was assigned during the quality check.

North Sea

● Import new and updated ODV.txt in ODV. This resulted in the following data collections: ocean_depth_profiles.odv, ocean_minimum_depth_profiles.odv,

![EMODnet logo — European Marine Observation and Data Network]

EASME/EMFF/2018/1.3.1.8 – EMODnet Thematic Lot 4 - Chemistry
Quality Control steps for EMODnet Chemistry Eutrophication aggregated datasets - v2021

ocean_pressure_profiles.odv and time_series.odv, sediment_depth_profiles and time_series_CJD.odv

● Aggregate the collections and merge all 'profiles' collections except sediment_depth_profiles which did not contain nutrient data, to finally create one collection for profiles and one for time series.

● Delete stations with Temperature and Salinity data only.

● Delete stations with no metadata.

● No duplicate stations.

● No stations on land

● Removed stations outside of the area, particularly transatlantic CTD tracks and stations in the Baltic Sea.

● Exported to a new dataset containing only stations with relevant data (nutrients, oxygen, chlorophyll, pH and alkalinity), keeping salinity and temperature also.

● Negative values were flagged with QF=6 and values changed to 0

● Values=0 (zero) and with QF=0 or 1 were flagged with QF=6 if not obviously wrong, if so the flag 4 was given instead.

● For extreme values out of "broad range", change the QC flags from QF=1 or QF=0 to QF=4. If outliers lie close to the "broad range" limits, change the QC flags from QF=1 or QF=0 to QF=3.

● Calculate DIN based on aggregated values

● Ratio of inorganic nutrients / total nutrients (DIN/TN, OP/TP). After inspection of the profiles with ratio > 1, change the flag of the outlier parameter to QF=3 or change the flags of all the parameters associated (for DIN: Nitrate, Nitrite and Ammonium) if no obvious outlier parameter.

● Data with QF=0 have been changed to QF=1 if no other flag was assigned during the quality check.

● Merged dataset with previous dataset and submitted merged set to Reiner Schlitzer for enrichment of previous dataset with newly introduced metadata fields. Rechecked ranges after merging of datasets.

● Marked all oxygen for cruise 14715 as QF=2, they are probably a factor of 10 too low. The data provider will be asked to verify this.

References:

Fichaut Michele, Balopoulos Efstathios, Dooley Harry, Garcia-Fernandez Maria, Iona Athanasia, Jourdan Didier, Baudet Laurent, Maillard Catherine (2000). A common

protocol to assemble a coherent database from distributed heterogeneous data sets: The MEDATLAS database experience. MAST workshop on Project Data Management, Institute for Space Applications, JRC, ISPRA, Italy, 11-13 June 2000. https://archimer.ifremer.fr/doc/00251/36211/

Giorgetti A, Lipizer M, Molina Jack ME, Holdsworth N, Jensen HM, Buga L, Sarbu G, Iona A, Gatti J, Larsen M, Fyrberg L, Østrem AK and Schlitzer R (2020) Aggregated and Validated Datasets for the European Seas: The Contribution of EMODnet Chemistry. Front. Mar. Sci. 7:583657. doi: 10.3389/fmars.2020.583657

Maillard, C., Fichaut M. (2001) MEDAR-MEDATLAS Protocol. Part I. Exchange format and quality checks for observed profiles, Rap. Int. TMSI/IDM/SISMER/SIS00-084, 1–49.

Schlitzer, R. (2020). Ocean Data View User's Guide Version 5.3.0, Ocean Data View. Available online at: https://odv.awi.de (accessed November 27, 2020).