To

Referee # 1

and Cc

Dr. Giuseppe M.R. Manzella

ESSD Chief Editor,

Dr. Alessio Rovere

ESSD topical editor

and

the ESSD Editorial Support Team

Oristano, 11/05/2023

Subject: reply to Referee's comments on "A Mediterranean drifters dataset: 1998–2022" by Alberto Ribotti, Antonio Bussani, Milena Menna, Andrea Satta, Roberto Sorgente, Andrea Cucco, and Riccardo Gerin, Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2022-344

Dear Referee,

thank you very much for your valid suggestions. We have devoted our best efforts to improve the submitted manuscript, aided by your insightful comments. We conducted a point-by-point response to your comments and queries and the manuscript has been edited and corrected, accordingly. The details of these changes can be found in the ensuing point-by-point responses to each and every comment/suggestion.

Referee's queries are shown in *italics* to differentiate our replies introduced by a **REPLY:** in **bold**.

Best regards,

Anonymous Referee #1
Referee comments
RC1: 'Comment on essd-2022-344', Anonymous Referee #1, 16 Dec 2022

This article is not fully comprehensible because of the English grammar, and as such, I deem it inappropriate for publication in its present stage. I would recommend a thorough editing and rewriting in order to improve the English.

REPLY: Thank you for the comment. The article was completely revised by a native English-speaking colleague.

I do not find the description of the processing methods to be extensive enough to understand the dataset. As an example, a kriging method is stated to be used to produce the estimated positions but nothing is said of the underlying structure functions that need to be first estimated to apply this

method. Another example is the quality control procedure of despiking: once again, no detail is given for this method (type of filter, threshold etc.).

REPLY: The 3rd paragraph (Data processing method) has been re-written improving the description of the used processing methods and quality control procedures.

"Decoded drifter data were then edited with the automatic procedure, through several QC tests, that replaced flagged time and location data with NaNs. In particular, impossible drifter positions (longitude > 180 or <-180 and latitude >90 or <-90) and the positions on land were discarded. In the latter case, about 4000 polygons, extracted from the GEBCO 1-minute resolution bathymetry data, which define the coordinates of all the coasts of the Mediterranean Sea, were used to determine drifters not in the water. For experiments extremely near to the coastline, this last QC test was not carried out to avoid the discarding of useful data. GPS data acquired before the beginning of the experiment and duplicated data due to transmission repetitions were also flagged. In general, randomly, the GPS drifter data may display duplicated positions acquired at different times. This was probably related to the buffer of the GPS module that does not correctly update the position in its memory before transmitting the data. The automatic procedure considers this issue and marks this data as incorrect. This procedure also evaluates the speed of the drifter. The first point (deployment position) was considered good and used as reference for the evaluation of the next point by computing the speed. If this speed exceeded 300 cm/s, the point was discarded and the evaluation is carried out on the further point, otherwise it was considered as a new reference and the procedure was iterated along all the available points. Additionally, a 4-degree polynomial fit was computed on a running window of 20 speed points, then speeds deviating from the fit by more than twice the total mean speed and twice the partial speed (computed considering only the points in the window) were not considered.

After the automatic editing procedure, some erroneous data still remained that required a visual check with a manual removal. In case of important temporal gaps or modification of the acquisition frequency during a Lagrangian experiment, the drifter trajectory was split into two segments and considered as two different deployments. New recovery/deployment information was included in the database and the automatic procedure relaunched. In the case of stranding, the automatic editing procedure discarded the data on land but is unable to recognise the moment when the drifter went ashore. The exact stranding time is defined by the operator through the visual analysis of the plotted drifter's track.

Edited data were then interpolated at uniform intervals using a kriging optimum interpolation technique based on the correlation of the data (Hansen and Poulain, 1996). The technique adopts a structure function and weights that were previously estimated using the drifter data collected during other experiments in the Mediterranean Sea between 1986 and 2016, included in the db_med24_nc_1986_2016 dataset (about 2000 files; Menna et al., 2017).

Drifter data with acquisition frequency between a few minutes to 2 hours were interpolated at 1-hour intervals, while those with acquisition frequency till or more than 6 hours were interpolated at 3-h and 6-h intervals, respectively. The velocities were then calculated as finite differences of the interpolated position."

The dataset website (https://doi.org/10.17882/90537) indicates that 366 trajectories (tracks) are available yet the article mentions 204? After downloading all the files, the number of track appears to be indeed 204, one per file.

REPLY: These tracks were discarded from the dataset and then drifters' numbers were recalculated accordingly.

These files do not follow a traditional data format: every single variable in these files (u,v, Lat, Lon, etc.) has its own dimension with the name of the variable. In other words, the variable "u" has dimension "u", which is odd. This does not suggest that these variables are contemporaneous or constitute time series along a common dimension ("obs" as an example). Moreover, because some variables exhibit missing values, a common software like Panoply is unable to plot time series for which missing values are present (because the dimension for that variable has missing values!). My suggestion is to reformat and recreate these files so that the variables have a common dimension (such as "obs"). There are template available for trajectory files, see as an example the one from NOAA NCEI (https://www.ncei.noaa.gov/netcdf-templates).

REPLY: The dataset has been realised following international standards used for Lagrangian data and thought to be easily comparable with similar datasets. Variables definition and dimension follow the Copernicus Marine In Situ NetCDF format manual (https://archimer.ifremer.fr/doc/00488/59938/) that specifies the NetCDF file format of Copernicus Marine In Situ TAC used to distribute ocean In Situ data and metadata. Moreover the dataset was ingested in SeadataNet following international standards and is also available at the address https://cdi.seadatanet.org/search/welcome.php?query=2610&query_code={9F00DF80-1881-42DD-9DF1-B9BD0282F2B0}. Such a link was also added in the text.

We agree with the reviewer that those variables which are all NaN are unnecessary and may be inconvenient to handle, so we have removed them from this dataset.

Moreover, some files have only two data points for each variables, and in the particular example of aarib_LCA113.nc, no valid value at all. What is the point of this set of data? This shows inadequate curation or automatic processing and editing of the data.

REPLY: The whole dataset was checked again. We have chosen to generate a dataset with three different interpolation frequencies (1-h, 3-h and 6-h) and to include all available trajectories with at least two measurements, then leaving the user the choice whether to use them or not.

The 1-hour interpolation, used for short experiments (a bit more than an hour), provides a few points only. These data may seem insignificant but, when put together with data from other drifters in the same area, they can constitute an important source of information. For example, they can contribute by describing the surface circulation in the basin by pseudo-Eulerian statistics, as described by Poulain (2001; https://doi.org/10.1016/S0924-7963(01)00007-0).