To
Referees # 1-2
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 Referees,

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).

Anonymous Referee #2
Referee comments
RC2: 'Comment on essd-2022-344', Anonymous Referee #2, 04 Apr 2023

The article entitled "A Mediterranean drifters dataset: 1998–2022" contains a series of files with trajectories of different kinds of drifters deployed between 1998 and 2021. Lagrangian drifters are a very useful tool for understanding the surface circulation of the ocean, as well as for determining mesoscale and submesoscale structures that are related to areas of accumulation of larvae or plastics, among others. However, in this work, the dataset needs to be revised in order to be used by other users in the scientific community.

On one hand, title of the paper does not seem appropriate, since, as indicated in the manuscript, the first buoy deployment experiments were carried out in the years 1998-1999 but were not repeated until 10 years later, so this title is misleading.

REPLY: We agree with the referee. We have changed the title in "A new released Mediterranean drifters' dataset"

And on the other hand, some files (e.g. hrib_LCE230) are not related to Mediterranean tracks (they show latitudes of 68 – 70 ºN). The work needs a great deal of effort to improve its quality in order to be published.

REPLY: We carefully checked the tracks and eliminated those outside the Mediterranean. Numbers of drifters' trajectories were recomputed accordingly.

ABSTRACT

Here, and along the manuscript, you say that the dataset belongs to 138 experiments but there are 204 tracks, could you explain why? trajectories of which drifters were divided?

REPLY: We have simplified the text just mentioning the experiments without numbers that confuse the reader.

INTRODUCTION

Are there previous Lagrangian experiments in the study area? Why is this dataset important?

REPLY: There have been a lot of Lagrangian experiments in the Med area which are published by OGS in the Mediterranean and Black Sea dataset (approximately 1700 drifter data since 1986): see for example 10.6092/B40CD642-9555-44FA-8B91-3CD88B6C225B, 10.6092/7a8499bc-c5ee-472c-b8b5-03523d1e73e9.

This dataset is important because it is composed of new data generated following the same procedures as the above-mentioned OGS datasets. This permits researchers to integrate and compare these data with previous released ones for scientific Mediterranean studies on circulation, climate, etc. We mentioned this in the Introduction leaving understood its importance, then highlighted in the Discussion and conclusions.

In the Introduction (page 2, lines 68-70 of the file with revisions) we wrote: "Recently, the OGS in Trieste has re-elaborated all drifters' experiments following standard and state-of-the-art procedures (editing and interpolation) already adopted for previously released Lagrangian datasets, then creating a new one freely available online."

In the Discussion and conclusion (page 12, lines 279-283 of the file with revisions): "Further, this dataset is also compliant and can be interfaced with the other drifter datasets produced by OGS in the Mediterranean and Black Sea which collect about 1700 drifter data starting from 1986 (Menna et al., 2017; Menna et al., 2018a; Menna et al., 2018b; Menna et al., 2019; Gerin et al., 2020), thus facilitating the use of a huge amount of drifter data available for scientific purposes in the Mediterranean basin (circulation, climate, etc)."

THE DRIFTERS

Table 1 and Figure 1: As above mentioned, some files are not from the Mediterranean Sea. Please remove them from this dataset and article, or include them but change the study area and so on.

Some of these drifters are: hrib_LCE230; mrib_LCE227; mrib_LCE229; prib_LCE229; qrib_ LCE229; urib_ LCE229; vrib_ LCE229; (...).

REPLY: We have deleted tracks outside the Mediterranean.

In my opinion, in order to be able to draw conclusions about the Lagrangian motion, it is necessary to have several points of the buoy's motion, which allow for concluding how the circulation in an area is. In this work, many trajectories are presented with little data (less than 10), so I do not consider that they give information. I attach a table with all the available buoys and the number of data. Therefore, I recommend removing the files with a such small number of locations.

REPLY: We are conscious of this but we think that also the very short tracks can be useful to describe the surface circulation in the basin by, for example, pseudo-Eulerian statistics as described by Poulain (2001; https://doi.org/10.1016/S0924-7963(01)00007-0). We left the user the choice to use them or not.

Table 1 shows information about hybrid and temperature (LCH and LCF, respectively). Nevertheless, when I downloaded the dataset I did not find information related to this kind of drifters. Besides this, please, remove temperature information in the NetCDF information (SST) because there is no data about that (as well as along the manuscript).

REPLY: We have eliminated this information from the dataset. We have explained the missing of these data with the sentence "Ancillary data like temperature, battery level or drogue presence were not considered as not available for all platforms" in the first lines of the 3rd paragraph.

Following the data, you do not specify the units of zonal and meridional velocity in the files. I think you should explain how you calculate.

REPLY: We have added the units of the zonal and meridional velocities in the metadata of the dataset. The explanation of the velocity computation is provided at the end of paragraph 3 with the following sentence "The velocities were then calculated as finite differences of the interpolated positions."

Tracks 2009-100 2010: the CODE

Line 102. Measurements of CODE 110 cm x 15 cm? These values are confusing attending to Figure 2B.

REPLY: The mentioned dimensions were indicating just the housing of the drifter, not well visible in the figure. We have decided to update the sentence, then considering the dimension of the whole instrument, sails included.

Line 115: Could you check it with your data? And not only with this kind of a drifter.

REPLY: Different types of drifters acquired with different sampling intervals from a few minutes to several hours but the dataset includes just processed interpolated data and not raw ones. As stated in the text, drifter data with an 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.

Tracks 2015-2022: coastal and offshore Nomad drifters

Line 156. Same here, in the dataset, there are no LCH and LCF files.

REPLY: We carefully checked and there are just a few LCH and LCF files in the dataset. They are visible at columns 2, 3 and 4 of the table in the Referee's comments.

Line 157. Were the Southtek drifters drogue? Why is not shown in Figure 2?

REPLY: All types of drifters, Southtek ones included, were equipped with a drogue. This information was missing at paragraph 2.1 on CLD drifter where we added the sentence "A sail (0.5 m length and diameter) was attached below the drifter to enhance the drag below the water surface.".
Drogues figures and their dimensions have been added to Figure 2.

Line 207. I do not understand: "A dedicated custom MATLAB tool..." What is this?

REPLY: We are sorry for the misunderstanding. We meant script and not tool. Anyway we modified the sentence to better detail the process, as follows: "The exact stranding time is defined by the operator through the visual analysis of the plotted drifter's track."

Line 213. "From the original 138..." Again, why? Which tracks?

REPLY: In order to avoid any confusion, we have rephrased the sentence and simplified it as follows: "At the end of the whole procedure, the final dataset consists of 158 interpolated drifter tracks (Figure 4) with at least two data points.".

Figure 4. Kind repetitive... you explain this in Table 1. There is a mistake in "th".

REPLY: We have re-plotted the figure by considering only the tracks.

Line 225. Which figure?

REPLY: We have modified the sentence by indicating the number of the figure mentioned in the text (Figure 5).

CC1: 'Comment on essd-2022-344', Adam Gauci, 10 Jan 2023

This work presents various drifter experiments that were carried out between 1998 and 2022. In Section 2 of the paper, the authors put forward the characteristics and specifications of the four types of drifters used. Very good and relevant comparisons between the instrument dimensions, battery duration, data transmission type, and data formats, are made. Clearly, the authors have extensive knowledge and experience in using and deploying such equipment. The points mentioned are useful to anyone that plans to carry out similar experiments in the same region or elsewhere. In Section 3, details about the data quality control procedures are provided. This has also been carried out meticulously and follows the current state-of-the-art methods. Details of how the data is saved and what information is stored are also mentioned. The 138 trajectories that are discussed are also made freely available. This dataset is useful to several users working with data assimilation or HF radar validation. Well done to the authors. Excellent work!

REPLY: We thank Dr. Adam Gauci for his assessment of our work. We hope that the dataset will prove useful and that the manuscript in its final form, after careful revision, will be of more interest.