

Earth Syst. Sci. Data Discuss., referee comment RC1
<https://doi.org/10.5194/essd-2022-344-RC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on **essd-2022-344**

Anonymous Referee #1

Referee comment on "A Mediterranean drifters dataset: 1998–2022" by Alberto Ribotti et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2022-344-RC1>, 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.

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.).

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. 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>).

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.