Earth Syst. Sci. Data Discuss.,
https://doi.org/10.5194/essd-2019-253-RC1, 2020

Open Access Earth System Science Data Discussions

Interactive comment on "The Sea State CCI dataset v1: towards a Sea State Climate Data Record based on satellite observations" *by* Guillaume Dodet et al.

Anonymous Referee #1

Received and published: 3 April 2020

General comments

This data set is the output of the "Sea State" project within the Climate Change Initiative (CCI) of the European Space Agency. The paper describes the implementation of the first release of the Sea_State_CCI dataset.

The potential of a consistent long-term data set of sea state data on a global basis is un-questionable. "Sea state" is listed as an Essential Climate Variable (ECV), and is relevant to a wide variety of users, from science to engineering applications.

This project offers three product levels (L2P, L3 and L4) as deduced from satellite radar altimetry, spanning from year 1991 to 2018. L2P is intended as an expert product

Printer-friendly version

Discussion paper

containing flagged but un-edited data; L3 (along-track) and L4 (gridded) are higher level products, obtained after systematic calibration and merging between the different satellite altimetry missions used, taking Jason-2 as reference.

The data set builds on the experience of the previous Globwave project (which essentially offers an L2 product), thus carrying a mature methodological background. The accompanying documentation is adequate in describing the data organisation and methods. No problems in accessing and downloading the data that I picked up, by probing various missions, times, and processing levels.

Differently from Globwave, the Sea_State_CCI products include de-noised SWH data obtained by a non-parametric denoising method (EMD – Empirical Mode decomposition). In addition to multiple missions cross-overs and buoy match-ups, this data set also introduces an interesting idea of validation against numerical model outputs, which is described in the submitted manuscript.

A key aspect in the delivery of products destined to multiple user communities, like in this case, is a clear description of the dataset. A well documented and consistent manuscript, together with an easy accessibility to the data are fundamental when dealing with diversified users, characterised by various degrees of expertise in handling the data. In my opinion, the manuscript satisfies this requirement in general, and just needs few technical edits and some clarification, as mentioned later.

Another important aspect in such user-oriented products is the need for clear and trustable indicators of the quality of the data, which should desirably be as complete as possible. Under this point of view, I think that the calibration and validation compartment of this manuscript still has some room for further expansion.

Said that, my overall impression about this work is very positive. The main action that I recommend regards an expansion of Annex B, concerning validation. Other aspects are very minor and mostly technical.

Specific comments (scientific)

The usage of Jason-2 data set as a reference for calibration by satellite crossovers is declared to be “evolving” at page 53 (Annex A). It would be interesting if the authors could specify which missions, data sets or techniques they plan to use next. This may seed a useful discussion with the community, at the benefit of the project.

Also, it would be interesting to know if the authors plan to give access to buoy match-up data. In particular, users may take profit from single-buoy-based match-up files (i.e. all the suitable corrected SWH data from proximal satellite tracks vs a given buoy), and do their own validation exercises.

Connected with the above, authors summarise results of the match-ups for all the missions in Table 1 (Annex B) and show bias and NRMSE plots calculated on a global basis in Figure 8. It would be interesting if the authors could provide statistics on a regional basis, too. In particular, a user may be interested in checking for differences in bias and NRMSE between different regional seas. Same applies to the expected SWH uncertainties as defined in section 4.1.2.4, which seem to be averaged and globally applied in a homogeneous way.

Authors at page 57 state that “...the validation of altimeter SWH was performed on a reduced data set including only offshore buoys”, the threshold being set at 200km to the coast. Similarly to the comment above, it may be interesting to offer separate statistics for more “coastal” buoys, where users can take SWH estimations even accepting a degraded accuracy when getting closer to the coast. Again, please consider the possibility to split this analysis regionally wherever it makes sense, e.g. depending on the density and distribution of the available buoys.

I hope that these points may seed useful discussion and contribute to improve this new and relevant dataset.

Specific comments (dataset organisation and processing)

Section 4.1.4 lists a series of planned improvements foreseen for the next releases of this dataset. Please clarify if the time series extension beyond 2018 and the incorporation of SRAL data will regard SWH only or Sea Surface Height as well, which is currently limited to Feb 2016. This would be interesting for next releases.

In section 6 the SWH outlier test is described in many sub-sections. I suggest to add some brief explanation for justifying the thresholds (5 std dev and 5m).

Regarding the L3 data set, a table with the assignment of numeric identifiers to each satellite seems missing in the document. I found such information only in the .nc file in the flag_values and flag_meanings fields of the "satellite" variable attributes. I suggest to add it to Section 4.2. I suggest also to specify in the text that the flag_values field contains the identifiers of the satellites named in the corresponding positions of the flag_meanings field. This is not necessarily obvious for the non-expert user.

I enclose pictures generated by projecting data extracted from a Jason-2 track (1 January 2018) onto a kml file. This "case study" is in South America, with the track approaching the coast with a high inclination, and crossing a narrow gulf before entering inland. This was chosen to check point flagging and rejection. I enclose pictures of SWH and SWH_denoised, the latter showing a reasonable rejection of close-to-coast points, especially when the track approaches the coastline in an almost-parallel way. In L2P data, SWH adjusted and its uncertainty are calculated in all points where SWH is available, being SWH directly copied from the original GDR files. This produces estimations also in datapoints labeled as "bad" and even flagged as "not_water", as clearly seen in the pictures. The third plot shows the footprint of all the available L3 data related to the same day. Focusing on the study area, data seem to be correctly edited with only valid measurements retained. I also had the chance to check a crossover in the Pacific Ocean, West of this area, finding a good agreement between the two missions involved. This is not a proof of anything of course, being an overall error analysis already conducted by the authors. It was just intended as a random consistency check.

Printer-friendly version

Discussion paper

Technical corrections

Here are a few points on more detailed technical aspects, listed as follows.

The work by Quilfen and Chapron, identified as “2019b” in section 4.1.2.5 is fundamental to describe the denoising technique, but still unpublished at the time of writing the submitted document. Please state if it’s currently accepted and if a pre-print is obtainable from the authors.

There are still some typos around, and a further general grammar check would improve the document, such as:

page 8: “...The processing of some missions may relies on older GDR versions”–>“...may rely...”

Page 10: “...improves the GlobWave products which were...” –> “...improves the GlobWave products, which were...”

Page 23: “...adjusted indirectly by a first comparison with ERS-2, itself adjusted relatively...” –> “...adjusted indirectly by a first comparison with ERS-2, which is in turn adjusted relatively...”

Page 53: “This calibrated Jason-2 swh data is considered...” Despite the usage of “data” as singular noun is tolerated today, may I suggest the more common plural usage.

Another few suggestions regarding the text are listed here:

page 18: the variable “swh_rejection_flag” formalised in the meta-description table at page 19, is called “swh_rejection_flags” at page 18. Please remove the inconsistency and check for other similar situations through the text.

Page 23: “...ENVISAT sigma0, which seems to be stable with time...” Please try to add a couple of rows to justify this sentence, which looks too much qualitative.

Page 27: ". . .or could be matched to the L2P measurements for some reason." It may be interesting here to list briefly the main reasons considered for rejection.

Page 45: the sentence "Besides, a test on swh rms (as provided in GDR for 1 Hz measurements) is performed, checking it is below an altimeter and swh dependant threshold" is actually unclear.Please rephrase and try to be more specific.

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2019-253, 2020.

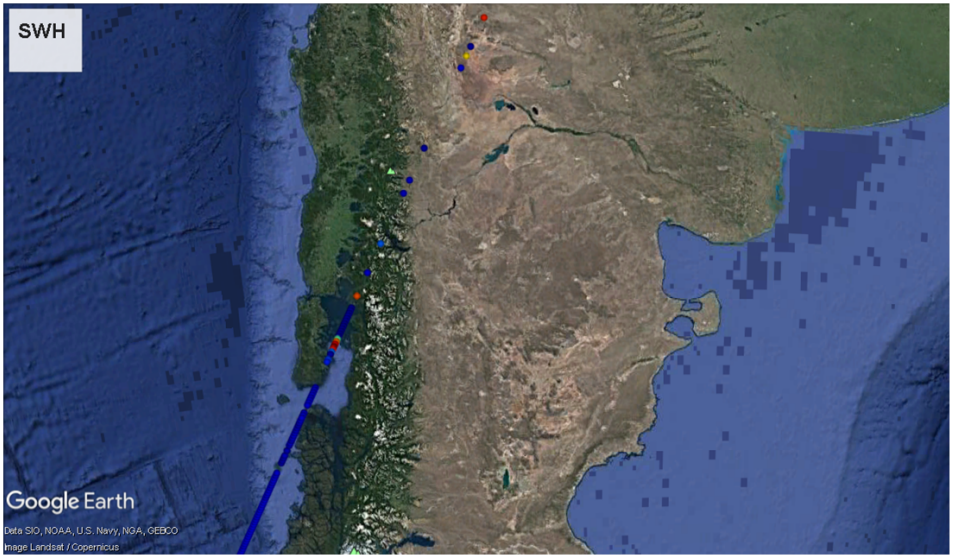

Fig. 1.

Printer-friendly version

Discussion paper

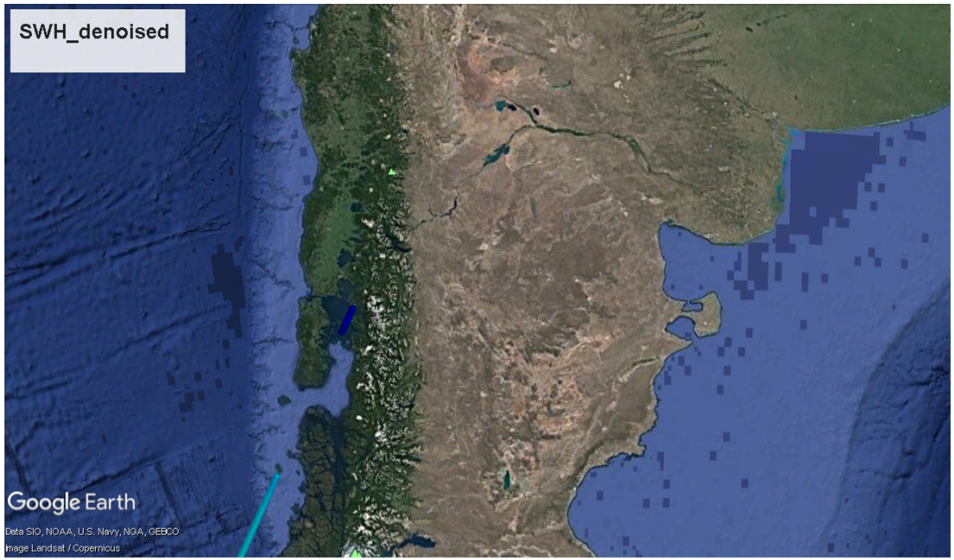

Fig. 2.

Printer-friendly version

Discussion paper

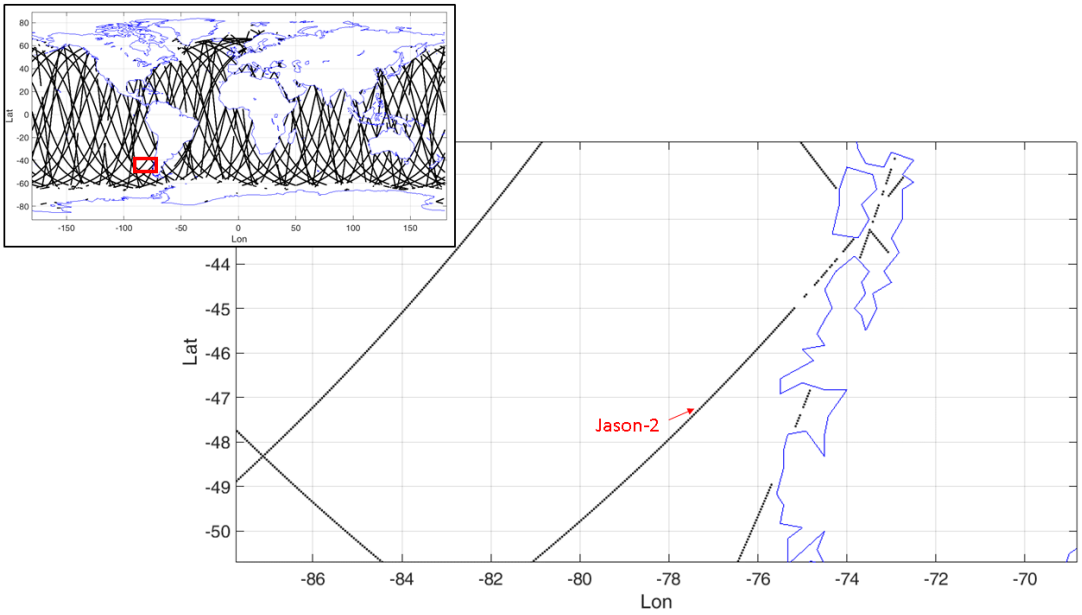

Fig. 3.

Printer-friendly version

Discussion paper