

Interactive comment on “How to improve your metocean datasets” by Erik Quaeghebeur and Michiel B. Zaijer

Nikola Vasiljevic (Referee)

niva@dtu.dk

Received and published: 9 October 2019

General comments

This manuscript represent an exhaustive review of three metocean datasets. The review is well presented and provides important recommendations for both the data creators and data users. The manuscript does not fall under the category of traditional research paper. The manuscript can be categorized as a mix of data paper (newly emerging type of scientific communication) and as a classical review paper. I am accepting the publication of the paper if the authors take into account the following feedbacks and include them in their manuscript:

It is important that the authors communicate their finding by taking into account

[Printer-friendly version](#)

[Discussion paper](#)



the FAIR data principles. Accordingly, I would like them to try to score the FAIRness of three datasets before and after the implementation of their work. Use the following simple sheet to do this:

https://www.cessda.eu/content/download/3845/35038/file/20170707_How_FAIR_are_your_data_Jones.pdf

What would be also helpful in this work is to review the FAIRness analysis of DTU Wind Energy:

<https://zenodo.org/record/1493874>

It seems that authors are not aware of EERA-JP wind endorsed wind energy taxonomies, metadata and restricted vocabularies:

<https://github.com/wind-energy/taxonomies-and-vocabularies>

I request that these are introduced as a part of NetCDF metadata and also reviewed in their paper. Also, they can see / review the implementation of the taxonomies in DTU Data (data publishing platform of DTU):

https://data.dtu.dk/DTU_Wind_Energy

I am not sure about the authors background, but considering the tone authors use that somewhat 'prosecutes' data creators of these three datasets points out that the authors themselves are not experimenters, ie. people who are designing field campaigns, setting up measurement equipment in field, collecting and providing data. Otherwise the manuscript would contain discussion of potential reasons why the three datasets are in the condition as they are. My point here is that the authors need to make effort and understand and discuss the reasons why the original data creators did not spend additional time improving the datasets. Also, they need to stress out what would be a benefit general benefit for data producers to spend additional time/resources to improve their datasets? Why should they do it? Is there a reward for them? Why their managers should allocate resources to improve data? What their for

[Printer-friendly version](#)

[Discussion paper](#)



them? Otherwise the manuscript they are trying to publish will not have desired effect on the community.

This leads me to the comment related to the section of the paper 'Code and data availability'. If the original data creator allow you to published the improved datasets you must include them as authors of those new and improved datasets.

Minor comments

The following is the list of minor comments:

Pg 1 Ln 24 remove brackets

Pg 2 Ln 5 - 6 : 'In brief: (i) Yes, there are shared issues, but, not unexpectedly, not all of them in all datasets. (ii) Dataset producers can address the issues with a few non-onerous additions to their creation practice. ' - this text belongs to conclusion or discussion of the paper.

Pg 3 Ln 2-1 Rephrase sentence since it is not very clear what you are trying to say.

Pg 3 Ln 3 Provide url to the website

Pg 5 Ln 7 Again, provide url to the website

Pg 5 Ln 30 Faulty data and quality flags are interrelated

Pg 6 Ln 5 It is not clear whether you converted each dataset to both formats or not

Pg 6 Ln 7 Be more concrete regarding the automatic data issues detection

Pg 7 Ln 1 what is the 'normal range' for the series ? values please!

Figure 3 Axes don't contain any information, it is hard to understand plots. Include some of the text from Pg 8 in the figure caption.

Pg 11 Ln 26 Be aware what term accuracy really means. I think you are talking about the uncertainty

Pg 17 Ln 29 - 31 Include reference to the wind energy taxonomies, metadata and restricted vocabularies.

Pg 19 Ln 27 How did you help to fix the bug ?

[Printer-friendly version](#)

[Discussion paper](#)

