ESSDD

Interactive
comment

Interactive comment on "KRILLBASE: a circumpolar database of Antarctic krill and salp numerical densities, 1926–2016" *by* Angus Atkinson et al.

T. O'Brien (Referee)

todd.obrien@noaa.gov

Received and published: 5 December 2016

Dear Angus et al.,

What an excellent effort and compilation of data. It is nice to have another "crustacean-named zooplankton data collection" available to the world. ;)

The paper / data technical reference is well written, and is careful to present the materials and concepts clearly to the reader, and to likewise give warnings and caveats and usage-cautions where they arise.

Two minor things that were not clear to me:

(1) How are the salp numbers being presented in the data column? If their chain ("combined aggregate") is intact, do you count it as a single item (one (1)) or do you estimate it into its individuals? The difference will obviously influence the numbers found in the database. Is this a question that goes unanswered into the original data itself, perhaps? (Did the original data forms not specify, perhaps?)

(2) Is the standardization equation (PDF page 12, line 292) showing up correctly? Specifically the ".2.255X." in the middle of the equation seems like a typographic error. Is that suppose to be "* 2.255X *" (the solid black and bold "multiplication" symbol)? See the PDF version of the equation if my question is confusing.

Because of these two minor things, I marked "minor revisions", but most likely they can be solved by adding a clarifying sentence and/or correcting a symbol in the equation. Other than that, an excellent product! Thank you for assembling it.

As I hope to see the work continued, two brief suggestions for the next release/update of KRILLBASE:

Future Suggestion #1: In Section 2.4 (data processing and error checking), if you log10(x+1) transform the abundances, it eliminates the "right-skewed frequency distributions" that you mentioned. Usually you get a nice bell curve with the log10 data, which makes finding anomalously high and low values easier (e.g., they approach the edge or fall outside of the bell curve). Also, by using log10, any calculation unit-related translation errors like "1000m3 vs m3" show up as a clear shift of exactly three (log10(1000) = 3) in a histogram plot. COPEPOD plots log10 histograms of new data on top of log10 histograms of the entire database, match by month and ocean region, looking for any data that shifts far left or right of the main histogram.

Future Suggestion #2: The current data set offers two options: "raw" and "standardized". There is also a lot of text mentioning what data to not use, or to be very careful with (e.g., shallow areas, data from the seasonal or latitude limits). Have you thought of creating a third data option, which would only include those standardized data that

you feel were safe to use? You could do this by adding an "okay to use" data column (flag) in the spreadsheet, for example. It would be really nice to see what your group felt were safest to use. If not a simple "yes/no" flag, you could do a scale like "good, questionable, use-caution".

Thank you for creating this excellent resource,

Todd O'Brien COPEPOD

Interactive comment on Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2016-52, 2016.

Printer-friendly version

Discussion paper