Ocean Science
Discussions

Open Access

EGU

Interactive comment on "Using feature-based verification methods to explore the spatial and temporal characteristics of forecasts of the 2019 Chlorophyll-*a* bloom season over the European North-West Shelf" *by* Marion Mittermaier et al.

Anonymous Referee #1

Received and published: 11 December 2020

The authors apply a feature-based comparison technique to sea surface chlorophyll snapshots in order to compare a model forecast to satellite observations and data-assimilative model output. While potentially useful, the particular forecast that is used in the study performs so poorly that the sophisticated comparison metrics cannot be assessed properly.

general comments

I have two major comments and the first likely affect the second. While the manuscript

Printer-friendly version

Discussion paper

starts out well written, sections 3 and especially 4 are very difficult to follow. Some statements seem to contradict earlier ones and I found myself second-guessing what I had just read in light of the next paragraph.

A good example of this occurs in section 4.3: "All objects are identified using the 2.5 mg mˆ-3 threshold." (l 594). Just as I was about to ask why a constant threshold was used again (when it does not work in earlier examples), the caption of Fig 11 states that "For (a) the thresholds varied to but were anchored to "truth" threshold of 2.5 mg mˆ-3". So it appears that variable thresholds were used but this is far from immediately clear. Yet this information is important to understand the results.

Generally, terms need to be introduced better and used more consistently: In section 4.1, an "observed threshold" is introduced, followed by a "forecast threshold", and a "seasonal threshold", but these terms are not used consistently. Similarly, "forecast" and "analysis" need to be introduced better: Initially the manuscript states that "In order to assess the European NWS Chl-a concentration forecast (AMM7v8), a satellite based gridded ocean colour product (L4) product and model assimilative analysis (AMM7v11) are considered as gridded "truth" sources." But then, without introduction, an AMM7v8 analysis appears and from then on "AMM7v8" may refer to the forecast or the analysis. To add to the confusion, the plural term "AMM7v8 forecasts" or just "forecasts" is often used, which (I think) is referring to the sequential nature of assimilative forecasting but is not helpful here to the reader.

To improve the description of MODE/MTD I would also recommend to include some early examples of what objects typically look like in the context of sea surface chlorophyll. Currently, the MODE algorithm is introduced, object attributes are described, MODE tuning is examined, difficulties applying MODE are identified, and an object's minimum size is discussed (pages 8-10) without ever mentioning what a typical object would be. It would be very useful for the reader to include an example image, a bit more zoomed in than those in Fig 2.

Interactive comment

Printer-friendly version

Discussion paper

As a reader interested in models and model assessment, my main interest in the manuscript was to see how the comparison technique works, and performs in a typical scenario. Unfortunately, it appears to me that the the model that was selected here underperforms so much in terms of recreating surface chlorophyll (both in magnitude and timing) that any far less sophisticated comparison would show it. Worse yet, the large discrepancy between AMM7v8 and the data appears to hinder a high level comparison of the two. So in the end, as a reader, I know that AMM7v8 does not recreate surface chlorophyll well but I am not sure if MODE or MLT are well suited for assessing surface chlorophyll output.

Here, I would suggest the following: rather than focusing on the comparison of AMM7v8 to L4 and AMM7v8 to AMM7v11, the manuscript could focus much more on the comparison of AMM7v11 to L4. A drawback here would be that L4 was already used to inform the AMM7v11 analysis but AMM7v11 performs much better thus permitting a much more interesting high level comparison. Perhaps a AMM7v11-based forecast could be employed to assess the capabilities of MODE and MTD.

specific comments

l 21: "whilst several forecast blooms did not materialise in the observations": This sounds like the data is to blame for not showing the bloom, maybe rephrase to something like: "while the model forecasts also showed blooms that do not appear in the observations".

l 22: "Whilst the model...": Most of this has been said in the previous sentence, would be more useful to move it one sentence up as an intro.

l 34: "double penalty effect": For readers not familiar with this term, it would be beneficial to describe this term a little better. Currently it sounds like not in the right place would be a single penalty and additionally not at the right time would be the double penalty.

Printer-friendly version

Discussion paper

l 114: "the models at 7 km resolution cannot resolve the coasts": It is not entirely clear what exactly is meant here: are you referring to the coastal chlorophyll a dynamics?

l 126: "is" -> "are"

l 133: It would be good to introduce the "Atlantic Margin Model" the first time "AMM" is used in l 117 or l 98.

l 135: I have a little trouble understanding this: is "Day 4 for the period of 1 March-31 July 2019" simply March 4th or are multiple 4-day forecasts created?

l 136: Mention that both the analysis and the forecast are used in this study. The previous sentence is confusing to the reader, it appears to state that the forecast is hereafter referred to as AMM7v8, when later on AMM7v8 forecast and analysis are used.

Fig 1: I would prefer to have the log scale included in the colorbar ticks ("10^{-3}" instead of "-3") rather than written out in the title. This would also eliminate potential confusion, as the values in (d) are not exponents, although the title may indicate that.

l 150: Use (a) and (b) instead of "left" and "right". Same for the next sentence.

l 158: "can be comparable": That is a very imprecise statement, as a reader I am not sure what this is telling me. Are the observation errors of the same magnitude as the model bias?

l 166: The instrument to measure ocean colour is satellite-borne, the ocean colour is not. Maybe use "remotely-sensed" or "satellite-derived"? Also, I would not refer to colours as concentrations.

l 167: Out of curiosity, why is a different satellite product used for the comparison and is there a significant difference between the two satellite products?

l 172: Is this referring to the coupling between physical and biological model components?

l 173: Is the AMM7v11 data assimilation also based on 3Dvar?

l 181: The "sequence of forecasts" is a bit confusing here. The next sentence refers to "a forecast". I know that data assimilation can create a sequence of forecasts but here I think it would be much easier for the reader to stick with "a forecast" throughout the description of MODE.

l 184: The mention of a model-based analysis is not helpful to the reader here, and it also not used again in this paragraph. I would suggest to rephrase to something like: "in this context, one is typically a model forecast, the other an observed field, i.e. observations regridded to match the locations of the model grid." Later on it can be mentioned that it can also be used to compare two model fields.

l 189: "observed objects" -> "observed field"

l 198: "and is based on a disk": Is the convolution kernel really a (flat) disk? The "based on" is confusing here.

l 201: Maybe use "objects" again, as above, instead of "areas".

l 204: Are the observed fields not smoothed? In step 2 it sounds like both fields are smoothed.

l 212: This is unclear: the first sentence in (6) "which together are expressed as the so-called "interest" score" makes it seem like one interest score is computed summarizing the fit of all objects. In this sentence, "interest scores are computed for all" objects. Please rephrase.

l 252: "minimum volume of 1000 grid squares": This should be "area" or is this applied below the ocean surface? Why is it 1000 grid squares here and 10 above? Reading it a few more times, it seems like this describes a "volume" in space-time, which needs to be made more explicit. "1000 grid squares" is not a volume, is this 10 grid squares times 100 time steps?

l 255: I am guessing here that paired objects are those for which an equivalent was found in the other field while all others are called single. And clusters are two or more objects in one field, classified as belonging together. Please add a bit of description here, to define these 4 terms.

l 262: Using the log-transformation seems like a sensible choice for chlorophyll. It would be good to know if the transformation was applied before smoothing the fields.

l 266: "mg." -> "mg"

l 267: The units here would be "$\log_{10}(mg\ m^{-3})$". In my opinion, it would be clearest to just provide the values in linear space, e.g.: "For this study, a range of thresholds were applied to the log10-transformed chl-a fields, corresponding to chl-a concentrations between 1.62 and 25 mg m^-3."

l 268: Why use 25 mg m^-3 here when the values of interest are in the range 3 - 5 (previous sentence)?

l 274: "This radius" -> "The 5 grid point radius"

l 276: Why so much discussion of different thresholds above when only a single one will be used here? Or is "here" only referring to the sensitivity analysis? Please be more specific.

l 304: What is "this", maybe use "The effect of bias"

l 310: "would yield no useful information": One could argue that having no matched pairs contains the useful information that the model solution cannot be very good.

l 315: Why is the L4 product referred to as an analysis here?

l 324: "the two that clearly differ more dramatically from Fig 2": What is meant here is probably "the two fields that show the largest discrepancies in Fig 2".

l 336: What about using different thresholds for the different fields?

l 339: "forecasts": Which forecast, only analysis solutions were considered thus far.

Fig 3: Please include the AMM7v8 distribution after the quantile mapping. And why not include the AMM7v11 distribution as well?

l 356: "the observed threshold": Do you mean the threshold applied to the observations?

l 358: After a lot of reading the next paragraphs over and over to figure out what the thresholds are, this is the position where I lost track. Here, it needs to be stated that the "value that has the equivalent rank in the forecast distribution" is the "forecast threshold". They are currently not linked and initially I thought that the forecast threshold was the same as the (slightly misnamed) observed threshold.

l 362: Based on this description it is not clear what this threshold is. It is also called "seasonal" here when Fig. 4 shows daily variations in a threshold. Are these the same thresholds, why is it called seasonal? If this is the threshold used to identify objects, please mention this explicitly.

l 365: Is this procedure applied to both v8 and v11?

l 374: "the latter": is this the L4 product? "the latter" is not really referencing anything at this point, please just use the product name.

l 375: Please explain better what this threshold means. It is impossible to understand this paragraph without knowing what the threshold signifies.

Fig. 4: It would be good to include a grid and reduce white space, the threshold never exceeds 6. I would further suggest to merge this figure with Fig 5.

l 390: "Error! Reference source not found." Something appeared to have gone wrong with the automated submission system.

l 391: "using the built-in functionality in MODE as for Fig 4.": This is unclear, please rephrase.

l 394: What is the forecast lead time?

l 405: "the forecasts are very active": What exactly does this mean, phytoplankton blooms?

l 406: "The latter object is not identified in the L4 ocean colour product." It is not clear what "object" this is referring to. I do not think this paragraph is very helpful to the reader at this point. So far the reader has only seen an example of what object identification should not look like (Fig. 2) and now here is a lengthy explanation of chlorophyll-a blooms leading to peaks in a threshold without showing an example of that these variable thresholds improve object identification.

l 434: "this work was done without accounting for the concentration differences": Does this imply that the variable threshold from the previous section was not used? But it seemed to be very import for successful identification.

l 444: Are "quilt plots" the same as "quilt "difference" plots"? I would suggest to stick with one term.

l 444: Here are two nearly identical sentences two (short) paragraphs apart: "Figure 6 provides a selection of quilt plots derived from using the L4 ocean colour products and AMM7v8 analyses during July 2018, using one of the merging options which was tested." (l 435) and "In Figure 6 some quilt "difference" plots are shown to focus on the individual characteristics of the AMM7v8 analysis and the L4 ocean colour product based on a set of initial data that was available for July 2018."

l 451: How useful is this analysis if it is already clear from the previous section that a comparison at the same threshold is not sensible? The 2.5 mg m^-3 threshold can be established without considering the model output.

l 475: Again, how valid are these results if different thresholds would be used for model and data? Would the conclusions about the smoothing radius hold?

Fig 6: The font is much too small.

l 489: Is this analysis done at the same threshold for data and model?

l 491: Is this "total area" of identified objects or ocean grid cells?

l 534: The lead time has still not been properly explained.

l 545: So these percentiles are characterizing the chl-a distributions among different areas. How can they contain values below the 2.5 threshold applied to the observations? The "(in this case 2.5 mg mˆ-3)" makes it sound like the same threshold was applied to the two products but this seems to contradict earlier sentences.

Assuming now that the threshold are different: From Fig. 3, we already know the distribution of chl values, what new information does Fig 9 give the reader? We know of the bias and have a rough idea of the distribution, and we also know that a higher threshold will likely be used for AMM7v8, which appears to be the main contributor to the differences between the distributions in Fig 9.

l 574: It would be nice for comparison to see good results for comparison. How would Fig 10 look for a AMM7v11 to L4 comparison?

l 623: "AMM7v8 day 0 forecast to L4 and AMM7v11 (labeled AMM7v8 vs AMM7v11, and AMM7v8 vs L4)": The labels are either switched or cleverly selected to confuse the reader.

l 653: Is the spatial centroid the average location of the center of a series of identified objects? This needs to be explained.