Geoscientific Model Development Discussions

Geosci. Model Dev. Discuss., referee comment RC2
https://doi.org/10.5194/gmd-2021-353-RC2, 2022

Comment on gmd-2021-353

Anonymous Referee #2

Referee comment on "DINCAE 2.0: multivariate convolutional neural network with error estimates to reconstruct sea surface temperature satellite and altimetry observations" by Alexander Barth et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2021-353-RC2, 2022

General comments

The paper by Barth et al. (2021) presents a modification of the DINCAE method for reconstructing missing data from satellite sources. The main advantages of this updated procedure lie on the ability to handle multivariate and non-gridded data, on the superior performance given by an innovative form of the skip connection and the addition of a refinement pass. Moreover, in order to prove the improvements with respect to the previous version, some applications are shown. The main result of the paper concerns the accuracy of the reconstructed error variance used in the refinement of the cost function when dealing with non-gridded datasets. The computational speed improvements are probably due to the change in the programming language. Overall, the paper shows interesting results concerning the accuracy of reconstructed sparse data but more attention to the scientific rigor and the clarity of concepts and figures is needed. I recommend publication with minor revisions, although I suggest paying careful attention in resolving them.

Specific comments

- I think that you use the words “variance” and “standard deviation” in an equivalent manner. The standard deviation is the square root of the variance that you call both σ throughout the paper (see for example in Section 2.2 or Section 4). Also I think that sometimes the same variable is also mistaken for the error associated with that variance. Please, revise their definition and be consistent with the name of the variables.
- Pg. 4: In the definition of the loss function (Eq. 3) a term is missing, which I think is negligible for your purpose but needs to be there (or, at least, you should mention that you are neglecting it). Also it is a good practice to specify what represents each variable before or after every equation.

- Pg. 5: Equation (5) is incorrect. It should be:
  $g(x,y) = \int w(x-x',y-y')f(x',y')dx'dy'$.
- In Section 3.2 you explain that in preliminary studies you found that including the data from adjacent seas can help a neural network to better generalize and to prevent overfitting. Can you cite some valid reference or say a little bit more about that? Do you have some results that you can show or are you able to give a reasonable explanation on this topic?
- I strongly suggest to spend more words analyzing the results of Table 2 and all the figures thereafter, including (average, max or min) values of the calculated errors with respect to the original data in the region.
- The last sentences of Sec. 5.2 (lines 6-8 of page 21) seem to state that even if the SD of the DINCAE reconstruction is higher, the RMS is lower. I do not see the contradiction in that, in fact the reconstructed SD from the DINCAE method is on average closer to the observed one than the DIVAnd's one. The point is not that the SD is higher or lower but simply more or less accurate with respect to the one derived from the DIVAnd method, and it would be nice to show a quantity that describes this improvement. Maybe you could further describe this improvement showing a whole map of the errors over the whole time period.

Technical corrections

- In general the paper presents a few errors (for example, several parentheses in citations are missing, singular/plural agreement in the sentences needs to be corrected and the adverbs position in several sentences is wrong). In particular, the word "data" is plural, please check the corresponding verb throughout the paper.
- Pg. 2:
  - At the beginning of line 2 it should be "which" instead of "who".
  - At line 13, since "*deeper*" is a comparative, the object of the comparison should be explained (deeper than what?). This is something that can be found several times in the paper.
- Pg. 3:
  - It is not recommended to start a sentence with an equation (line 20-21). Also, it is a good practice to explain each variable when showing an equation, saying what X, f and l stands for in Eq (1). Same for equations (6) and (7).
  - At the end of Section 2.1 (line 28) I think "basis" should be "bias".
  - Line 6: The word "error" appears twice.
  - Line 10: I think "expect" should be "except".
  - At the end of Section 2.2 (line 21) I think "gage" should be "gauge".
- Pg. 6 (line 26): The dataset SEALEVEL_EUR_PHY_L3_REP_OBSERVATIONS_008_061 changed its name in SEALEVEL_EUR_PHY_L3_MY_OBSERVATIONS_008_061 after some corrections. Can you please correct the name or declare the date in which you downloaded the data?
- Pg. 10 (line 15): I would rather say "sine and cosine".
- Pg. 11 (line 16): Either you are missing a word when you say "as explained in the following…" or you should change it to "as described further on" or similar.
- Pg. 17: In the caption of Figure 7 you state that the map for the expected error standard deviation by DIVAnd is without adjustment while the title of the panel suggests "Adjusted error std. dev.". Could you correct either one of them? Also in both captions of Fig.s 7 and 8 you mention the panels (a), (b), (c) and (d) but those letters do not appear in the images.
- In Figures 9 and 10 there are no units. Also, the title of the left panels suggest

“Number of data points” which you never explain nor comment in the text. Moreover, there is no explanation of the colors or the dashed lines in both plots.