Geoscientific Model Development Discussions

Geosci. Model Dev. Discuss., referee comment RC2
https://doi.org/10.5194/gmd-2021-99-RC2, 2021

Comment on gmd-2021-99

Anonymous Referee #2

Referee comment on "SITool (v1.0) – a new evaluation tool for large-scale sea ice simulations: application to CMIP6 OMIP" by Xia Lin et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2021-99-RC2, 2021

The manuscript "SITool (v1.0) – a new evaluation tool for large-scale sea ice simulations: application to CMIP6 OMIP" describes a new Python diagnostic tool to evaluate sea ice models in the Arctic and Antarctic over the historical period. Although it is designed primarily for atmospheric reanalysis-forced simulations, as presented in the manuscript, it could be useful in other model frameworks as well. This tool is complementary to other climate model evaluation tools, such as ESMValTool. Comparison with multiple observational datasets allow for evaluation of sea ice concentration, extent, edge location, ice thickness, snow depth, and ice drift. The evaluation of Ocean Model Intercomparison Project runs are used here as example, but also provide results and sea ice model performance.

This manuscript is well within the scope of the journal, as it introduces a novel new tool for evaluating climate model performance on a critical component: sea ice. Consistent, repeatable methods of evaluation like this are greatly needed by the community. It also provides novel results on the impact of the atmospheric forcings on the modeled sea ice (especially that model biases are significantly reduced by using JRA-55). The title and abstract well capture the key points. Methods are generally clearly described, and the code is well-documented and easily accessible. The paper is generally well structured and clearly written, but some figures could be improved for easier interpretation. It would be helpful to be clearer about how the presented results connect with the code and outputs in the published package. I believe with minor suggested edits and demonstration of code implementation (either by an additional reviewer or by the author, within the repository), this manuscript warrants publishing. Note: this reviewer did not complete a test of the scripts, and it may be useful for this code to be checked and tested by someone who is experienced in working with these output types.

Comments:

- L23-30: [suggestion] Separate out into 4 sentences for improved readability
- L70: "either" feels imprecise here. Is it accurate to say "both atmospheric forcings, when possible, …"?
- L76: It would be helpful to more clearly state how the diagnostics of concentration and thickness in previous studies (2020) are the same or different from the diagnostics proposed here.
- L88-89: Including some sort of table summarizing the diagnostics available, in terms of variable and type (spatial map, mean, IIEE) would be helpful
- L90: Would be preferable for this to be a complete list of add'l sea ice variables, and then "e.g." is not needed
- L92-95: Please discuss somewhere (results or conclusions) the possible implications of interpolation.
- L105/Fig. 1: Can this be re-oriented such that the order progresses downwards? (So, sea ice input data above SItool?
- L105/Fig. 1: In the "Observations" portion, it would be helpful to make it clearer that extent and edge are also coming from the concentration products. In other words, it would be helpful to be explicit what observations each of the defined "metrics" are compared to
- L106: [style suggestion] I think more subsections to separate out the methods would be helpful. This will be helpful in providing a quick reference for users of SITool.
- L117: This sentence is hard to understand. Perhaps it could be clarified by separating into 2 sentences.
- L135: Why only February and September? Is it a user option to select for other (or all months)? If so, please be clear what the difference is between options for the tool and what is being shown here as demonstration.
- L187: Please describe the options for the user to select years for comparison
- L204-205: Would the authors recommend freeboard be included in future CMIP model outputs for observational comparison? If so, include this in the conclusions.
- L215/Table 2: separate dataset name and reference into separate columns
- L220: Perhaps I have misunderstood something about the calculations, but how can you determine the typical error to get the metric shown in Fig. 7b for Envisat without the second observational product (SnowModel-LG)
- L272: Is the primary difference between OMIP1 and OMIP2 protocols the atmospheric model JRA55 vs. CORE-II? If so, I suggest it may be more clear to use "J" and "C" rather than 1 and 2.
- L272/Fig. 2: Are there any significant differences in patterns between the two observational products (NSIDC and OSI)? If not, why not just use the normal error relative to the mean between the two products? I do not believe that the difference in comparison between the two products is a key point here, so I'm not sure it is useful.
- L272/Fig 2: [suggestion] It may be helpful to show this after Figures 3 and 4 (showing specific metrics), so that these values have some context and introduction already
- L272/Fig. 2: Does "Ano" in figure refer to anomaly? (i.e. interannual variability?) This somehow needs to be made more clear, such as in the figure caption.
- L272/Fig. 2: [suggestion] add some vertical space between NH and SH. "NH" and "SH" may be sufficient (rather than "North" and "South"), and would save some room
- L290/Fig. 3: The multi-model mean line is hard to distinguish. Consider using a thicker or brighter/more distinctly colored line.
- L310/Fig. 4: If possible, it would be helpful to explicitly label "std" and "trend" on these plots to demonstrate where values in Fig. 2 are coming from.
- L316: Interesting. What are the implications of this? Should typical error not be used in this case, or used with caution? Are there similarities in how some products are derived

that result in this metric having less utility?

- L341-2/Fig. 5: Separate by product in panel (c) to be consistent with other Figure. (Unless products are combined with averaging, as suggested in comment above)
- L336/Fig. 5: Perhaps include in subplot title a summary of what is being evaluated, such as "Extent: models vs. NSIDC"
- L366: I'm not sure I understand how you can have IIEE for the observational product OSI-450. Is this rather the observational products? In that case, should it be called the "typical error" here?
- 8/Fig. 9: Are these annual means/using all months? Please specificy period evaluated in figure captions.
- L489-499: Are the plots the ones that are used in the manuscript and/or the appendix? Please clarify
- It would be helpful to provide example of completed code run and diagnostic plots in the repository on Github