Geoscientific Model Development Discussions

Geosci. Model Dev. Discuss., referee comment RC2
https://doi.org/10.5194/gmd-2021-132-RC2, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on gmd-2021-132

Anonymous Referee #2

Referee comment on "Sensitivity analysis of Ocean Temperature Regressor v1.0; a data-driven model of ocean temperature" by Rachel Furner et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2021-132-RC2, 2021

Recently, the data-driven models have become a hot topic for the atmospheric and oceanic predictions for time scale from synoptic to interannual. Many cases indicate that such models have high predict skills and much less computational resource-consuming. However, as mentioned by the authors, some fundamental questions have no been answered yet, such as if the method can capture the predictability natural of the system and if we can physically explain the results. In the present study, a data-driven model has been developed for predicting the sea surface temperature in the short term. The method they used is a regression model with a non-linear term. The model has been trained using an idealized ocean model dataset as an observing system simulation experiment. They found that the model can predict the SST leading one day, and the dominant variables are also identified. After that, the sensitivity withholding experiments are conducted to identify critical physical variables and processes, like the vertical structure and the non-linear term. In general, this is an interesting and valuable paper and provides helpful information for this kind of data-driven model. Therefore, it is worth to published in GMD after the MINOR revision.

Major questions:
1 I think the most critical problem is the resolution of the model is too coarse (2 degrees) compared with the predicting time scale (1 day). Because the movement of the ocean is much slower than that of the atmosphere, the surrounding points cannot affect the central point for one day. The only fast process is the convection due to instability. That is why the coefficient magnitude of the center point is much larger than other points in Figure 4. If the horizontal resolution is increasing, I guess that the results may also be changed because some processes may transport the signal of around points to the central point during one day. Therefore, I suggest the authors test the sensitivity of the resolution further.

2 In the withholding experiments, we can find that the errors are smaller than the control experiments in some places, such as withholding the information about the vertical

structure in Figure 6 and withholding currents in Figure 7. How to understand these results? Is there some information that will bring negative effects or significant errors into the model? Can we find an optimal combination of all this information?

3 The configurations of the model or experiment are not present very clear. Please give more information about the experiment in the manuscript, such as the vertical levels, the time scale of the restoring, the observation of the restoring sea surface temperature and salinity, and the coefficients of the GM scheme.

4 The method of selecting data is also not present very clear. For instance, I do not really understand how to choose the data every 200-day and deal with the 3D variables at the surface. Please say something more about the details.

5 In the present study, the authors only show the results of one-day prediction. I am curious how the model performs when the predicted time scale becomes longer, like 5-day or 10-day. I suggest the authors further discuss the skill of the model for more extended timescale prediction.