Ocean Sci. Discuss.,
https://doi.org/10.5194/os-2018-166-RC1, 2019

Ocean Science Discussions

Interactive comment on "Using Canonical Correlation Analysis to produce dynamically-based highly-efficient statistical observation operators" *by* Eric Jansen et al.

Anonymous Referee #1

Received and published: 19 February 2019

General comments

The paper describes a method for producing efficient estimates of skin and subskin SST using the Canonical Correlation Analysis (CCA) statistical technique. This is described in the context of observation operators which are used in data assimilation schemes to provide information about the model-observation misfits in order to correct the model. The aim of the paper is to make better use of satellite SST observations in data assimilation systems compared to most existing operational systems by improving the way the diurnal cycle of SST is represented. The work is therefore valuable and the paper is well written.

Printer-friendly version

Discussion paper

Near the end of the last section, there is a reference to another paper submitted by Korres et al. which appears to address a very similar topic by the same list of authors. This other paper should be referred to in the introduction and the aims of the two papers differentiated clearly. It might not be possible to address some of the comments below without using results presented in the Korres et al. paper. If that is the case perhaps the authors should consider merging the papers or having a two part paper.

Specific comments

1. The paper describes the method in a general sense with SST as a "use case". It might help the flow of the paper if the use of the method for SST was more central to the paper so that, for instance, the introduction would have more information about the literature on SST analysis and data assimilation. The way it is currently structured, more review is needed on the various other types of meteorological and oceanographic observation operators used, for instance radiative transfer models.

2. The scheme is not tested in a data assimilation system (although that appears to be done in the Korres et al. paper). Would the aim of a scheme using the CCA method be to correct for errors in the diurnal cycle of SST at 1 m depth using all available satellite SST measurements, or to correct the model's foundation SST?

3. There are shortcomings in the design of the experiments to test the performance of the CCA method in section 4.3. The validation is performed over the same period as the model data used to generate the CCA OO, and then compared to the same model data. There is no comparison of the CCA results to real observations.

4. The comparisons with other SST assimilation methods described in section 5 assume that the GOTM model and the atmospheric forcing driving it are correct. If the skin model in GOTM had a bias for instance, could it be worse to use the CCA OO based on that model than the other methods tested? A more independent way of assessing the CCA OO method based on GOTM is needed, e.g. by comparing to real observations.

5. There is not any discussion of the need for the adjoint of the observation operator in data assimilation (which is obviously very efficient when using the CCA method).

6. The GOTM dataset used for training the CCA is a model and not based on observations. The work of Pimentel et al. 2018 describes how the model represents the skin and subskin SST, but some more information is required here to justify its use. A brief description of how the model has been made to represent the skin and subskin would help (the highest level of 1.5 cm is not at the same depth as the skin or subskin). A summary of the assessment of the model compared with real observations is also needed here, otherwise there is no link to the real world.

Technical comments

1. Pg 2, line 5. The wording “are not or not sufficiently modelled” is a bit confusing on first reading. This lack of process representation in the model is often included in the representation error in data assimilation systems. A discussion on the relationship between the complexity of the observation operator, and the inclusion of representation error in the R matrix would be good here.

2. Pg 2, line 6 -7. Does the cost of the “second” model depend on the observation number as implied here, or the (horizontal) model grid size? The cost of these models, e.g. a diurnal model, is cited as one of the justifications for implementing the CCA method. It is not obvious that a simple diurnal model is that expensive compared to the cost of the full GCM.

3. Pg2, line 7. “needs” to “need”.

4. Pg 3, eq (3). Normally matrices would be in uppercase but here you start using lowercase letters. This is particularly confusing when you use uppercase and lowercase of the same letters (e.g. u and v).

5. What are the implications of eq (11)? It is taking into account the biases in the training observations and model. These presumably are not constant in time so how can

Printer-friendly version

Discussion paper

this be applied in practice? There is no description of how these values are calculated in section 4, or their magnitude.

6. Pg 7. Where does eq. (14) come from?

7. Pg 5, 1st paragraph. There is no description of the near-surface temperature structure to introduce the reader to terms like “skin” and “subskin”.

8. Pg 5, line 11. “At the same time. . .”. Not always at the same time.

9. Pg 5, line 13. “straightforward assimilation”. I think you mean here that it is a problem when the observations contain significant diurnal cycle changes at the skin or subskin, and that is not accounted for when comparing the observation with the model. A straightforward approach could be to remove those observations from the assimilation as you mention later on.

10. Pg 6, line 8. You take the daily mean value for wind and insolation. How good is this for determining the magnitude of the diurnal cycle?

11. Fig 1. What depth is the diurnal cycle that is plotted?

12. Pg 6, line 11. How do you get skin and subskin estimates from GOTM?

13. Sec 4.3, second paragraph. Wouldn’t it be fairer to do the validation of the CCA on a different time-period to the one used to build the statistics for the CCA OO?

14. Section 5 is a comparison of the CCA OO to other methods. I think its title should be changed.

15. Pg 10, line 2. Waters et al (2015) assimilate data during the day where the wind speed is high.

16. Last paragraph section 5. The last sentence of this paragraph describes the method that should be used in the paper to assess the performance of the CCA OO.

17. Pg 11, line 24. Can you include a reference for the magnitude of the diurnal cycle?

Interactive comment on Ocean Sci. Discuss., https://doi.org/10.5194/os-2018-166, 2019.

Printer-friendly version

Discussion paper