Ocean Sci. Discuss., 6, C314–C316, 2009
www.ocean-sci-discuss.net/6/C314/2009/

Ocean Science
Discussions

Interactive comment on "Enhancing temporal correlations in EOF expansions for the reconstruction of missing data using DINEOF" *by* A. Alvera-Azcárate et al.

A. Alvera-Azcárate et al.

a.alvera@ulg.ac.be

Received and published: 6 August 2009

Thank you for your comments regarding the paper. We have answered your two main questions here (part of your text is included for clarity):

I really think that other ways to estimate the interpolation error should be included and discussed in this paper, and, in particular, at least the rms difference between the whole original and reconstructed images should be looked at.

We have already tried to force DINEOF to take more EOF modes than the number

suggested by the cross-validation. What we found is that noise is sometimes present in a mode prior to the information about the small scale variation on the data sets, or that noise and small scale features are contained in the same EOF mode. Therefore, by adding extra EOF modes we cannot guarantee that the reconstruction will be better.

However, the reviewer is right in requiring some discussion on the cross-validation technique. If the observations are error free, then the estimator of the reconstruction error by cross-validation is unbiased. Also, in the case of random observational error, the estimator should converge. Only when the observational error is correlated in space/time (which is probably the case), the estimator might be inappropriate.

The error estimator proposed by the reviewer (i.e. to calculate the rms difference between the whole original and reconstructed images), however, presents the same problem, and even an additional one: for a very small number of cross-validation points, the number of EOFs that will minimize the error between the whole image and the reconstruction will tend to the number of images used so as to exactly reproduce the original data where there are no clouds. If a large number of cross-validation points are used, the estimator there will eventually increase and overall an optimal number of EOFs will be defined. Unfortunately, this number of EOFs will depend on the number of cross-validation points retained, certainly not a desirable property.

We suggest rather the following approach:
Consider α the cloud fraction, rms^2_{cv} the error variance estimated from the cross-validation points and $\text{rms}^2_{non-cloud}$ the error variance between original data and reconstruction in unclouded regions. An estimator of the error could be:

$\alpha\ \text{rms}^2_{cv} + (1-\alpha)\ \text{rms}^2_{non-cloud}$

In this way the estimator should be robust with respect to the number (and location) of the cross-validation points and reflect correctly the contribution of the two different error sources (clouded and unclouded regions). If the reviewer agrees, we will try this approach and add a section to discuss the cross-validation technique.

I would also suggest to compare the results from both filtered and unfiltered DINEOF truncating the EOF reconstruction at the same number of modes, just to be sure that the improvement observed is effectively related to a better conditioning and not to the higher number of modes used in the filtered case.

This is a good suggestion, that we will add to the manuscript. (I would expect that fig. 7 gets improved in the non-filter version, due to the additional EOF modes used, but not fig. 8, with the warming event, which is clearly an effect improved by the filter)

Interactive comment on Ocean Sci. Discuss., 6, 1547, 2009.