**Ocean Science
Discussions**

Interactive comment on "Forecast and analysis assessment through skill scores" *by* M. Tonani et al.

Anonymous Referee #2

Received and published: 16 April 2007

Review of "Forecast and analysis assessment through skill scores" by Tonani, Pinardi, Fratianni, and Dobricic

In this paper the authors present an assessment of the skill of the ten day operational ocean forecasts produced for the entire Mediterranean Sea once per week for the period from 16 Aug 2005 - 10 Jan 2006 (a total of 22 forecasts). In order to assess the forecasts skill they use the root mean square error (rmse) and a "percentage" skill score (ssp) based on rmse of the forecasts as compared to the rmse of persistence. The truth or observations for the validations are the analysis fields produced by the data assimilation - forecast system based on the model. While this is an important topic in terms of assessing the value of such an operational forecast system, the results presented in this paper fall far short of what should be, and could be, a more rigorous

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

EGU

assessment of the model's performance. Furthermore, the paper presents only the "forecast assessment" but not the "analysis assessment" as indicated in the title (more on this below in comment 2). In my opinion the paper requires major revisions before it can be considered acceptable for publication in Ocean Science. Following are some specific points that need to be addressed.

1. Considering the fact that the paper was submitted in Feb 2007, is there some particular reason why the authors restricted their assessment to only 22 forecasts when they could have used a much longer period (perhaps even an entire year)?

2. Since the forecast assessment is based on a comparison to the analyzed fields (as opposed to real data), it is very important to include here the "analysis assessment" (note the title of the paper) in which we are given some indication of the magnitude and behavior of the errors associated with the data assimilation system. If this assessment is given in a companion paper, then at least a clear reference to that paper and a brief summary of the performance of that system should be repeated here. I have no problem using the analysis fields as truth as long as we know something about their quality. Alternatively, they should include a validation based on the data, but this too has its pitfalls as the data are incorporated through the data assimilation system (unless there are some independent data available).

3. The validation of the forecasts is restricted to the rmse and the rmse based ssp. Model evaluations of this type usually include, in addition to rmse, some measure of the bias (mean error) and some measure of the anomaly correlation score (acc). While rmse measures the domain averaged accuracy of the forecasts, acc tells us something about the spatial skill and the ability of the model to forecast the anomaly patterns.

4. Their definition of "persistence" (Eq. 2) is rather strange. Usually persistence is used as a minimum skill forecast which the model must beat in order to have predictive value. Thus rmse(persistence) should be defined (in the authors' notation) as

Rms(t) = SQRT [SUM (Xa(t=1) - Xa(t))**2 / N],

i.e., persistence is a forecast in which the initial conditions are persisted, but which should be compared to appropriate analysis at the later verification time. I do not understand the meaning or significance of FP as they defined it in Eq. 2 in which they compare the evolving forecast fields to the initial conditions. Subsequently, all of the evaluation of model performance with regard to persistence is highly questionable.

5. In the first two paragraphs of Section 4.1, the references to Figures 3 and 4 are reversed.

6. In the paragraph that begins at the end of p.195 the authors present the normalized rmse of the near surface atmospheric fields that are used to force the model. They note that the decay of the atmospheric forecast skill (i.e., increase in rsme) is nearly linear and based on the similar growth of the ocean model rsme (FA) they conclude that the inaccuracies in the atmospheric forcing are the main source of error. I am not at all convinced by this simplistic explanation. One must recall that the model is forced by wind stress and heat flux components, all of which are nonlinear functions of the meteorological variables that they assess. This is probably further compounded by comparing normalized rmse of the atmospheric forecasts with absolute rmse of the ocean forecasts

7. On p. 199 line 9, I believe they mean "effective" and "ineffective" rather than "efficient" and "inefficient".

8. Page 199, line 14. I would prefer to see a somewhat more rigorous definition of "predictability limit" rather than a simple eyeball assessment of the "saturations" of the rmse curve.

9. The list of references should be arranged alphabetically.

10. Some of the figure, such as the right panels of Fig.2 are unreadable.

11. The entire manuscript can use a good editorial reading for spelling mistakes, grammatical errors, and missing words.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

EGU

Interactive comment on Ocean Sci. Discuss., 4, 189, 2007.

Interactive
Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

EGU