

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

Anonymous Referee #1

Received and published: 22 March 2007

This paper deals with the assessment of forecast fields produced by the Mediterranean Ocean Forecasting System, using RMSEs and 1 skill score, considering the period August 2005 - January 2006. I am sceptical about the selected scores and the way the authors looked at them. The general remark below explain my thoughts. I believe the methodology used in this paper is confusing, and the resulting considerations (and conclusions) are often unjustified (or even misinterpreted). I suggest the authors to choose a more suitable set of scores to carry out their analysis. In addition, no observations are included in the assessment, so everything is rated using analyses as ‘sea truth’, and this hide the actual skill of the model forecast.

– General remarks –

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

The authors claim to design this new score, the SSP, following Murphy's papers.

Briefly, Murphy's MESS is

$$\text{MESS} = 1 - \text{MS}(\text{forecast-observation}) / \text{MS}(\text{reference forecast} - \text{observation})$$

In this paper, analyses are considered to be the 'truth' instead of observations (by the way, that may be a severe assumption, and not really justified in the manuscript), and persistence is considered as the reference forecast. Following Murphy's theory, MESS should then be read:

$$\text{MESS} = 1 - \text{MS}(\text{forecast-analysis}) / \text{MS}(\text{persistence} - \text{analysis})$$

The reason for that is: $\text{MESS} > 0$, then forecast is better than persistence, i.e., add more information, $\text{MESS} < 0$ means the forecast is worse than persistence (generally you don't want this).

Instead the author choose the score in the following way (I write it disregarding the percentage):

$$\text{SS} = 1 - \text{RMS}(\text{forecast} - \text{analysis}) / \text{RMS}(\text{forecast} - \text{persistence})$$

or, equivalently,

$$\text{SS} = 1 - \text{RMS}(\text{analysis} - \text{forecast}) / \text{RMS}(\text{persistence} - \text{forecast})$$

So, the authors are rating analysis vs. persistence considering the forecast as it was the true state. It doesn't sound that close to Murphy's approach. If $\text{SS} > 0$, it means that analyses are closer than persistence to forecasts, which is not a useful information. If you show that forecast are closer than persistence to analyses it may help more. Using Murphy's MESS, you can judge if the forecast is good/bad compared to a reference forecast. Using the score presented in this paper you can't. So what are you assessing then? I'm confused.

In addition, I don't agree with the interpretation of the relative weight between FA and

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

FP (and therefore the resulting considerations throughout the manuscript), which affects also SSP estimates (by definition).

For example, in section 4, para 3 and 4, the authors state that forecast is better than persistence since $SSP > 0$ (or in other words, $FA < FP$). Now pretend that in one location the analysis (supposed to be the sea truth) is 10 degC. The forecast is 12 degC. Persistence is 9 degC. Then $FA = 2$ degC, $FP = 3$ degC. In this example, persistence is more accurate than forecast, but $FA < FP$ and consequently $SS > 0$. On the opposite, if persistence is 11 degC, then it is still more accurate than the forecast, but now the $FP = 1$ and therefore $SS < 0$. . . I don't understand why you want such a score! If you want to rate the forecast compared to persistence, I suggest to use $RMS(\text{forecast-analysis})$ vs. $RMS(\text{persistence - analysis})$. . . and not to use what you call 'persistence error', $RMS(\text{forecast-persistence})$.

– Minor remarks –

- 1) in the manuscript actually MFS analyses are not assessed (as stated in the title).
- 2) Sometimes it seems that the authors actually computed time/space averages of SSP scores (for example, the statement in the first para of 4.2). I recommend accumulating separately the numerator and the denominator first and then to compute the final resulting SSP. Otherwise the median is preferable as measure of central tendency.
- 3) about the SSP, I don't see positive outcomes in taking the root of MSE, instead of just sticking with the MSE. But I see a drawback, since RMSE skill score decomposition would not be as straightforward as MSE skill score decomposition is.
- 4) add reference for FGAT
- 5) ch.3, 3rd para: what do you mean with 'best' analysis? How did you rate them?
- 6) I would drop figure 1, it doesn't seem relevant. If not, please state in the figure captions what's 'J'. It is explained in the text but well below your first pointing to figure 1.

7) ch 4.3 second para. different scores from summer 2005 to winter 2006 may be associated to a seasonal signal ? and not necessarily an improvement due to data assimilation?.

– Editorial remarks –

1) font-size of the labels in figure 7-8 is too small

2) ch.4 first para. It is not figure 4, it is 3. 2nd para: it is not figure 3 it is 4.

3) references are not in alphabetical order

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

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper