Ocean Sci. Discuss.,
https://doi.org/10.5194/os-2017-78-RC2, 2017

Interactive comment on "Forecasting experiments of a dynamical-statistical model of the sea surface temperature anomaly field based on the improved self-memorization principle" *by* Mei Hong et al.

Anonymous Referee #2

Received and published: 14 December 2017

I have four major concerns with this paper 1. Construction of the first two predictors ie T1 and T2, 2. Selection of the other predictors, 3. Structure of the model, and 4. Model validation

Detail comments on the 4 points

1 . Section 2.2 EOF deconstruction. This section requires some more detail. While the given reference describes the EOF method, we need to know how it is applied here. Is the correlation or covariance matrix used? How are the anomalies constructed - simple removal of the monthly means? How are the anomalies smoothed - how strong is the smoothing and is it applied spatially or over time? More importantly, why are only the

first 2 EOFs considered? A similar analysis has recently been reported by L'Heureux et al (Clim Dyn 2013, DOI 10.1007/s00382-012-1331-2). Their first two EOFs are similar to those described here (but with no smoothing and hence lower explained variance). Using different data sets and time periods, they show that the 2nd EOF is not stable, being entirely due to the strong trend (also evident in Figure 1d). The pattern does not appear if the data is detrended, and also becomes less important if different time periods and/or domains are used. Most importantly, they do not interpret it as indicating "the ENSO signal beginning to decay".

2. Section 2.3 Predictor selection The selection of other potential predictors is confusing. Apart from T1 and T2, the other potential predictors come from a fairly limited set, and are not well supported by the referenced works. In lines 157-160, zonal winds in the western and eastern equatorial Pacific are mentioned, and it is well known that westerly wind anomalies in the western equatorial Pacific can (and do) trigger equatorially trapped oceanic Kelvin waves. There is an extensive amount of literature on the relationship between western equatorial Pacific zonal wind and ENSO, but here no references are given and only the eastern equatorial winds is considered. Trenberth et al discuss a link between ENSO and the PNA pattern (amongst other modes of extratropical variability), but this is the context of ENSO forcing of the PNA, ie ENSO leads to PNA teleconnections, but PNA does not predict ENSO. Yang et al introduce the EAWM index, but they note that "the relationship between ENSO and the east Asian winter monsoon is relatively weak". Nowhere do they suggest that the EAWMI is closely related to any ENSO indices. It is not surprising that the east Pacific wind and PNA do not feature in the final model

3. The model The remainder of section 2.3, concerned with determining the number of predictors is difficult to follow. It is not until section 3 (page11) that it is revealed that the model is a dynamical system of four second order coupled equations, involving the products of the various predictors as well as the predictors themselves. Nowhere is the inclusion of these terms discussed or justified. What physical processes do these

terms represent? What do the predictors squared represent?, and the cross products ie what do T1 * SOI or T2 * EAWMI mean? Since the model is not a linear regression model, is stepwise regression a valid procedure for determining the significance of the predictors?

line 195. The idea that a model with the number of predictors less than 10% of the sample size can avoid overfitting is new to me. The reference given (Tetko et al) is about neural networks. Is this applicable to the system of coupled equations used here? (I could only see the first page) Also I am not sure if the discussion in 198-203 is incorrect. Even if only 34 parametres are accepted, the full set of 56 parameters must be estimated to know which to accept or reject. This may be more a problem of introducing artificial skill, which has long been recognised as a problem in statistical forecasting. It generally arises when you try enough predictors, and retain those that "work" and discard the others.

This question of the number of parametrs / predictors is exacerabated in Section 4 and 5 where the number of predictors is increased again by including lagged values. On first inspection Equations 3 and 7 involve 112 parameters. There are 28 alphas, 28 thetas, as given in lines 395 and 396. (In line 202, it is stated that there are 28 self memorization parameters beta; but there are no betas in Eqs 3 and 5, but there are in Appendix B) In addition each of the four F "dynamical cores" involve 14 parameters as shown in Equation 1, assuming that the same F is used at each lagged time. Given that the input data (the xi) are different at each lag, is the same F a valid assumption? Even with the authors 34 accepted values in the Fs, there is still a total of 90 parameters. This is well over 10%, and on the authors own criterion, this would suggest that the system is perhaps overfit. Additionally, all the 720 observations are not statistically independent. Both T1 and the SOI (and probably T2 with its strong trend) are strongly auto-correlated, and the effective sample size is probably significantly less than 720. All in all, this discussion is very confusing!

4. Model Validation

line 281-288. This paragraph took me a long time to understand, especially how one could obtain correlations and MAPE values based on a single forecast. As I understand it, "at this time" refers to the forecast at five months, and the correlation and MAPE are calculated over the first five months forecasts, and in general the values at the Nth month are based on the first N months forecast. (I assume that this is the "n" in the equation for MAPE on line 283) This method would suggest that the correlation at one month is undefined, and 1.0 (perfectly accurate) at two months? This same type of calculation appears to be used in Tables 3 and 4. line 289-298. Another confusing paragraph. January 1951 to January 1952 inclusive? is 13, not 12 months. How was the omitted section forecast, ie was it simply a 12 (or 13) month forecast starting at the last point before the omitted data? it is difficult to judge how "good" the forecast was based on Figure 3. Again it is not clear how the correlation and MAPE statistics were calculated - only one value is given, so presumeably it is taken over all (720 months) forecast? However the discussion in lines 310-312 suggest that individual 12 month forecasts were also evaluated. Overall the discussion of the forecast process and its validation in not clear.

Some minor points (There are many minor points - these are just a few that stood out to me)

In line 170, all 4 data sets range from Jan 1951 to Jan 2010, yet in at least 4 places, lines 292, 373, 402 and 416 forecasts are evaluated up to December 2010?

lines 249-253. Why does normalising the raw values avoid the overfitting problem?.

line 254. What criterion is used to determine what are "weak items" with "small dimension coefficient"

line 280 "forecast performance ... was better" than what??

Section 6.2 - Table 5 The values reported here do not make sense. By construction, EOFs (the spatial patterns) are orthogonal, and the PCs (the time series) are uncorrelated. L'Heureax et al report that the correlation between PC1 and PC2 (using the same HADISST data set) is 0.4 when the time series are detrended. This is the same value quoted in Table 5. Has T2 been detrended here also?

EOF1 is the cannonical ENSO pattern, and its time series is stronly correlated with the standard Nino indices (l'Heureaux et al give a value of 0.94 between their first EOF and the Nino3.4 index). In turn the Nino3.4 index is strongly correlated to the SOI, so that is difficult to see the correlation beteen T1 and the SOI being as small as the 0.4 given in Table 5.(This correlation is where the term ENSO ie El Nino - Southern Oscillation arises)

Acronyms need to be defined the first time they are used, eg EOF on lines 128-130

Figure caption (line 912) for figure 1 in List of figures is incorrect, and different to that given with the figure itself (line 959).

References are incomplete; there are at least 15 references that are not cited in the text, and a number that are cited but referenced.

Interactive comment on Ocean Sci. Discuss., https://doi.org/10.5194/os-2017-78, 2017.