Earth Syst. Dynam. Discuss.,
doi:10.5194/esd-2016-14-RC1, 2016
© Author(s) 2016. CC-BY 3.0 License.

Earth System Dynamics Discussions

Interactive comment on "The use of regression for assessing a seasonal forecast model experiment" *by* R. E. Benestad et al.

Anonymous Referee #1

Received and published: 26 July 2016

The authors employ factorial regression to assess the influence of model set-up and initial conditions on seasonal forecasts of the EC-Earth2.1 earth system model; particularly the influence of sea-ice, the stratosphere and NAO SST's initial conditions on monthly mean air temperature anomaly forecasts. The Walker test is employed to assess the global statistical significance of the model response to these three factors. The authors conclude that seasonal predictability of monthly mean temperature in the mid-to-high latitudes is sensitive to the various choices of model set-up and initial SST conditions, with the most sensitive response to the stratosphere and SST initial perturbations, and sea-ice to a lesser degree. Also, the authors argue that the model response in the mid-to-high latitudes is partly due to the non-linear interaction of these three factors.

Major:

The problem studied by the authors is relevant and the methodology seems appropriate. However, the results shown are hard to follow at times and the conclusions are not totally convincing. This is partly because of the disconnect between figures, table and explanations, as well as seemingly contradicting results. Specifically:

1. When describing Figure 2, the authors state that "A comparison between the different panels in Figure 2 suggests that the different choices for model set-up had similar magnitude on the predicted outcome for all these factors", which seems to contradict the conclusions of this paper. On the other hand, when describing Figure 3, the authors observe that response-to-error ratio for the different factors are less than unity for sea ice in most regions, and greater than unity in more extensive regions for SST initial perturbations and stratosphere. This is more in tune with the conclusions of the paper. Then, if the response-to-noise ratio of the coefficients is what matters, why showing Figure 2? It is a bit confusing. Figure 3 seems to be showing exactly the ratio of the values shown in the panels of Figure 2. If so, what useful information or conclusion can we draw from Figure 2 alone?

2. The authors obtain minimum p-values for sea ice, SST-perturbation and stratosphere larger than the Walker p-value (**page 6, lines 155–156**). This implies that the Walker test for field significance fails, and so it suggests that the difference between temperature responses for the different model set-ups and SST initial conditions considered here are not statistically significant at the global α level employed. Thus, how robust are the conclusions of this paper? Did the authors consider air temperature forecasts at lead times other than 3 months and/or vertical transects at latitudes higher than 60°N? Did the authors consider more traditional tests than factorial regression for local assessment to check whether they lead to minimum p-values smaller than the Walker p-value? Could the authors expand on this?

3. The authors address the non-linear dependence on sea-ice, SST and stratospheric effects by comparison of the difference of temperature response of two model configurations with the linearized difference of temperature response, relative to these three factors, of the two model configurations - that is, by comparison of the left and right hand side of Equation 1 (**Page 7, lines 164–165**). This comparison led the authors to conclude that the non-linear model response is confined mainly to the mid-to-high latitudes where the difference between left and right hand sides of Equation 1 is the largest. Could the interpretation of this result be misleading if other factors are taken into account? For example, could the model response still be approximately linear (i.e., left and right hand sides of Equation 1 approximately equal) if other factors (e.g., snow cover) are considered?

4. According to the authors, factorial regression "offers an alternative to traditional ways for estimating statistical significance", and has not been widely applied to seasonal forecasts. The reader would greatly benefit from a more in depth exposition of this methodology in the context of the problem examined by the authors. For example, the authors state that (**page 4, lines 94–95**) "If the final response ΔT is a linear function of sea-ice, SST, and stratospheric effects, then (...) "

$$\Delta T = x_1 C(sea\text{-}ice) + x_2 C(SST) + x_3 C(stratosphere)$$

What is $C(\cdot)$? What does $C(\cdot)$ represent? Should it be interpreted as a categorical predictor depending on its argument? What does this actually mean in this context? Could the authors give a bit more detail about the implementation of this multiple-linear regression model? Should there be a residual added to the right hand side to account for other sources (i.e., factors) of changes in mean temperature?

5. **Page 3, line 76:** The authors state that "An overview of the model simulations are listed in Table 1". I could not find Table 1.

6. **Page 5, lines 107–133:** The authors refer to Figure 1-2 but only describe Figure 1.

7. **Page 5, line 108:** "Figure 1 shows horizontal transects at 200 and 50 hPa levels respectively". Is this correct? It seems that Figure 1 only shows horizontal transects at 200 hPa for different model set-up and SST initial perturbations.

8. **Page 6, line 123:** "The horizontal picture at 50 hPA (Figure 1) suggests ...". Is this shown?

9. **Page 6, line 136:** When referring to Figure 2, it says "Panels b-f show difference in the forecasts due to different choices in the model set-up in terms of the regression coefficients β, ...". Should it be Panels b-e instead?

10. **Page 6, line 137:** When referring to Figure 2, it says "... panels g-e show error estimates for these coefficients." Should it be panels f-i instead?

11. **Page 6, lines 137–139:** When referring to Figure 2, the authors state "Regions with large values estimated for the coefficients and large errors suggest a high sensitivity ...". Large values relative to what? Are these coefficient and errors given in units of temperature? Is not the response-to-noise ratio what actually matters?

12. **Page 7, line 164–165:** The first term on the right hand side of Equation 1 is "(DyIce-NoDyIce) nNAO L62". Is this correct? I would expect "(DyIce-NoDyIce) pNAO L91" to be consistent with the difference in the left hand side. This also applies to the title of Figure 4b and the caption to Figure 4.

13. **Page 7, line 170–171:** "The comparison shows that the non-linear response ... presented in Figures 3–5". I could not find Figure 5.

14. Labels in figures are generally too small, particularly in Figure 2. In most cases, axes titles and legends in colorbars are missing, e.g., y-axes and colorbars in Figures 1, 2, 3. Also, should the title of Figure 4 be “difference of monthly mean temperature" instead of “monthly mean temperature"?

Minor:

The authors should also address the following issues:

1. **Page 1, line 8** Consider replacing “demonstrate". Perhaps, “show"?
2. **Page 5, lines 110–11** At the end of line 110 and beginning of line 11, the word “positive" seems to be missing (i.e., “positive NAO").
3. **Page 7, lines 176–178** “We can also test ... if the model response is highly non-linear". Is this assertion accurate? What if more factors are needed for linear dependence?
4. **Page 8, line 191** Remove “has"
5. **Page 8, line 193** Remove “presented"?
6. **Pages 9–14, lines 214–343** The authors should use a unified style for the references.

Interactive comment on Earth Syst. Dynam. Discuss., doi:10.5194/esd-2016-14, 2016.