

## ***Interactive comment on “Adding value to Extended-range Forecasts in Northern Europe by Statistical Post-processing Using Stratospheric Observations” by Natalia Korhonen et al.***

### **Anonymous Referee #2**

Received and published: 16 September 2019

#### Summary

The authors use a post-processing technique based on stratospheric predictors of the Arctic Oscillation on ECMWF forecasts to try to improve predictive skill of surface temperature at weeks 3–6. Overall the paper addresses relevant scientific questions (given the lack of some stratospheric processes such as the QBO and its teleconnections in the forecast model), but certain aspects of the paper could be clarified and re-organized. I also have questions about their particular technique of only using forecasts made on the first week of each month, and using stratospheric data at 10 hPa, rather than in the lower stratosphere which is a better indicator of stratosphere-troposphere

C1

coupling. I suggest a major revision.

#### General Comments

1) More could be explained up front about how the QBO teleconnection processes in particular are not well captured in the forecast models (particularly after early winter, as per Garfinkel et al. 2018), and why. In particular, see Line 14–15, page 3: Here it should be specifically mentioned whether the ECMWF S2S model (the version used here) is able to self-generate QBO variability (most S2S models cannot- instead they are initialized with observed QBO winds and degrade relatively quickly away from that towards model climatology). This should be further emphasized (possibly showing something like the forecasted QBO winds in the tropics compared to observed values for each of the initialization dates used here, so it's clear that the model is missing this process).

2) Sensitivity testing to the parameters chosen here could be further provided, as I'm not sure I understand what was the motivation for some of these choices. For example, see Line 21–25, page 5: Why were QBO winds at 30 hPa selected and was the sensitivity to this level tested? What about 40 or 50 hPa? Also, I understand using the 60N 10 hPa metric for stratospheric variability, but it's not a great metric for coupling to the surface. I assume you might see much better results using winds at a lower level in the stratosphere, near 50–100 hPa. See for example Karpechko et al. 2017 (<https://doi.org/10.1002/qj.3017>). What happens if you use a lower level instead?

3) Furthermore, presumably this technique of identifying weak vortex periods (Line 25, page 5) misses quite a few observed SSW events in the stratosphere (because they don't all occur in the last 10 days of the month). Could you say something about this, and why your method is still valuable? It also seemed like the sample size for  $u < 3.8$  m/s was extremely small ( $n=9$ ) for the longer record period and must be even smaller for the shorter 1997–2016 period of the hindcasts. I wondered why it was so small I wondered why it was so small (144 forecasts, \*10th percentile climatology = 14 times

C2

this should happen). Can much be said for a predictor that only happens 9 times in 36 years? What happens if this threshold is relaxed to include more events? This is another reason why getting daily tropical winds so all forecast initialization dates may be used could be valuable. Along these lines, I'm confused about the number of events shown in Figure 2 for the two rightmost columns.

4) Finally, I'm curious about the choice of predictors in terms of their covariability. How many times when you had EQBO (and EQBO with maximum thresholds) did you also see a weak polar vortex at 60N 10mb? I assume these are correlated/concurrent, and the EQBO merely adds additional samples where the vortex is weak but not weak enough to meet the 3.8 m/s threshold. It would be interesting to state what the mean value of 60N 10mb zonal winds are during the various EQBO thresholds.

#### Specific Comments

1) Line 6, page 4: You don't consider forecasts in the summer here though, so why 52 weeks? Shouldn't it just be Nov-Feb forecasts? (also, because it wasn't clear- I assume you mean throughout that you use any runs initialized in Nov-Feb, but the forecasts in Feb obviously forecast into March/April, correct?)

2) Line 19-20, page 4: Are the operational forecasts used in this study? I'm not sure I understand why the CRPS is adjusted for 51 members if only the hindcasts with 11 members are shown throughout, but maybe I missed where the operational forecasts were used.

3) Line 23, page 4; line 2, page 5: Not sure what is meant here by "annual mean"- do you mean the average of weeks 1-6 across all years? Or the average across all months to get one value for each year? I'm not sure I follow what is meant in this paragraph. Also, you might present Figure 1 in section 2.1 and 2.2 in relation to what is being discussed to make it clearer.

4) Line 12-13, page 5: Isn't this data MERRA-2 reanalysis? Could you be more specific

#### C3

about what this data is, and how it was derived? Also, if you are using ERA-interim to verify the forecasts- why not also use ERA-interim for daily zonal winds, both in the stratosphere polar vortex and in the tropics for the QBO? Singapore winds may not always represent the zonal-mean tropical winds that drive QBO teleconnections. Presumably the forecast model is initialized using the winds in the reanalysis, correct? And if you need daily winds to be able to look at forecasts other than the first forecast initialized each month, you could easily get this data all from one product, rather than three different products.

5) Line 18-30, page 5: I found this a bit hard to follow without any visual explanation, and I wonder if it would be clearer to discuss Figures 2 and 3 in this section instead of later.

6) Line 13, page 7: is CRPSS above zero a reasonable metric of skill? CRPSS near zero but positive surely can't be that useful (some of these values in Figure 1 are less than 0.1).

7) Line 19, page 7: Why was only the minimum daily AO considered and not the mean? Does the mean not change enough?

8) Line 25, page 8: I think Fig 4p looks very much like Fig 4j.

9) Figure 6: How are panels (a,b) different than Figure 1? (other than being two week averages rather than 1 week).

#### Technical Edits

1) Line 19, page 1; line 22-23, page 3; possibly other locations: specify that you are referring to the previous months' tropical stratospheric wind observations.

2) Line 2, page 2: not sure what is meant by "experimented during a one year living lab"

3) Line 3, page 2: put comma after "production"

#### C4

- 4) Line 24, page 2: maybe instead “other definitions have been used”
- 5) Line 10, page 3: I would clarify that this paper looked at S2S hindcasts similar to what you are looking at here
- 6) Line 2, page 4: remove the word “scale”
- 7) Figure 1- dots seem a little blurry, might make sure it's high enough resolution for final version.
- 8) Line 29, page 7: add in “maximum” to “QBO's monthly mean zonal wind components” or it doesn't make much sense
- 9) Line 5, page 8: add in “where” between “cases” and “the”
- 10) Figure 2- might be nice to put in bold those p-values that are less than 0.05. Also, shouldn't the last column be labeled “EQBO with u winds < 10 m/s OR ZMW at 60N and 10 hPa >3.8 m/s”?
- 11) Line 24, page 9: The way this is written is confusing, do you mean p decreased so significance increased?
- 12) Line 29, page 9: add a “to” after “corresponding”

---

Interactive comment on Atmos. Chem. Phys. Discuss., <https://doi.org/10.5194/acp-2019-679>, 2019.