***Interactive comment on* "Influence of Arctic Stratospheric Ozone on Surface Climate in CCMI models"** ***by* Ohad Harari et al.**

Anonymous Referee #2

Received and published: 21 November 2018

Summary This paper evaluates relationships of Arctic stratospheric ozone (ASO) and surface climate, and ASO and El Niño-Southern Oscillation (ENSO), using model out- put from the Chemistry-Climate Model Initiative (CCMI). They find that the connection between ASO and surface climate may arise due to dynamic variability in the lower stratosphere, which has a strong correlation with both ASO and surface climate. They also find a weaker-than-observed (though "still significant") relationship of ASO and ENSO, with ASO leading ENSO by one to two years.

General Comments In general, I think the writing is clear and the analysis is well- explained. I think there is definitely utility in looking at these relationships in state-of- the-art chemistry-climate models, particularly since it's apparent that some of the ob- served relationships might be an artifact of sampling. My one major comment relates
to the relationship of ASO to ENSO. While the authors do due diligence and attempt to make sure this relationship isn't a factor of auto-correlation, I'm confused about their method. For example, I don't understand why the "minimal useful correlation" is basi- cally zero (so basically anything greater than zero is useful?). See specific comments below too. I think this technique at a minimum needs to be better explained, but I would also recommend addressing the issue using Granger causality techniques, as in Mc- Graw and Barnes (2018).

Thank you for your helpful and constructive suggestion. We have removed the minimally useful correlation and replaced it with a new analysis based on causal effect networks, as suggested by the reviewer. This rather new technique is methodically outlined and implemented in Runge et al 2017 on a climate dataset. The specific technique we use is based on Pearl causality (Pearl 2009) and is somewhat different from Granger causality techniques (i.e, Mcgraw and Barnes (2018)). The benefits and drawbacks of Granger vs Pearl causality are discussed in Runge et al 2017. Note that Pearl causality has been used in climate science by Kretschmer et al 2016.

The causal effect networks analysis is based on a two-step algorithm. The first step is the PC algorithm (Spirtes and Glymour, 1991). This step is used to find the "parents" (i.e. of a time-series) while the next step is used to quantify the causal strength of the first step. The full analysis description is outlined in the revised manuscript. In short, we produced three variables, each one is a time-series (ASO, ENSO and Zpole), from the SWOOSH and CCMI models. The PC step of the analysis was used to find the parents of each variable, within a lag of 10 to 27 months prior to it. The significance threshold that was used is 0.05 (note that this threshold has a different statistical interpretation than that used in e.g. Student t tests- see Runge et.al 2017). In contrast to Runge et al 2017, we use the PC step with q_max=10 (maximum combinations of conditions) as the original PC algorithm suggests. In the second step of the analysis we used two different methods to evaluate the parents' causality strength. One method, named Partial Correlation, is used to calculate the correlation between two sets of residuals – that of a variable and that of its parent. The residuals are obtained by regressing out the influence of all other parents identified from the PC step (i.e. the first step). (see ParCorr alg. In Runge 2017). The partial correlation result was tested with a two tailed T test with alpha=0.05. The second method that is used to quantify the causal strength is called linear mediation (Runge 2015).

While this method can be used to compute different causal strength scores, we used it to calculate the beta coefficients of a multiple linear regression with the parents of the variable (e.g., ENSO) as regressors. The results for the ENSO variable are shown in the figure below copied from the revised manuscript:

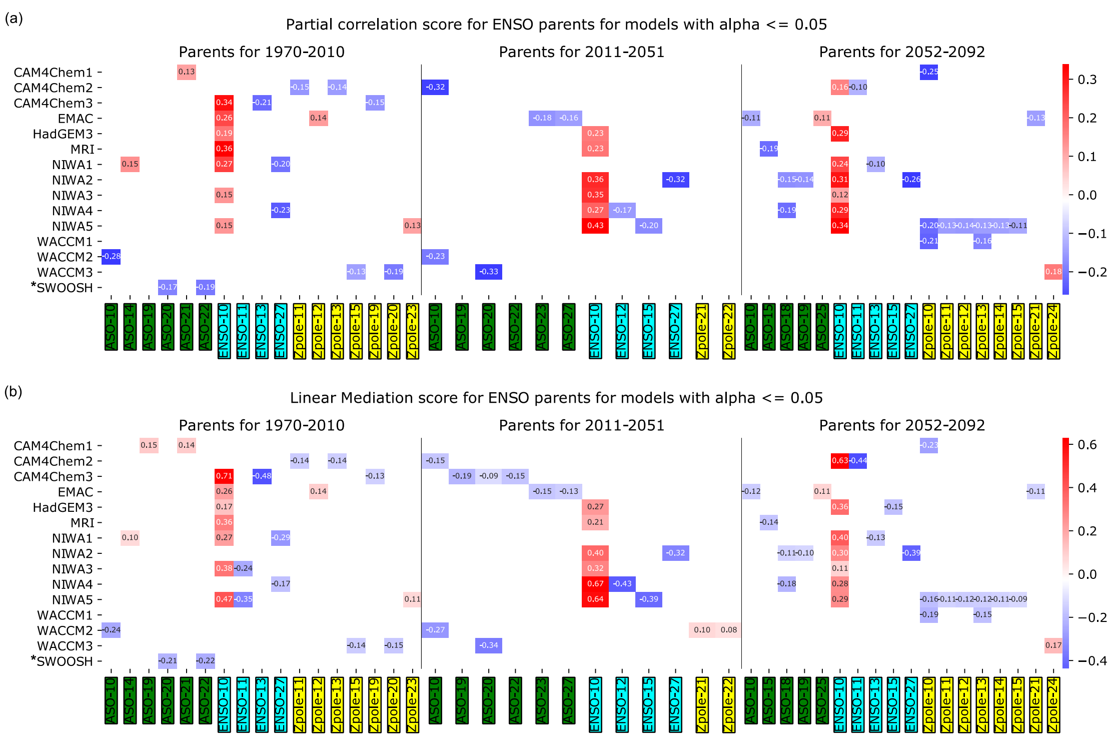

Fig12: Results of the (a) PCMCI analysis and (b) PC algorithm with liner mediation. Both started by finding ENSO's Parents using the PC algorithm. Then, two methods were used to estimate the connection's strength: (a) partial correlation with 95% confidence level, and (b) computing the beta coefficients of ENSO's parents in the different time periods.

In observations/SWOOSH, ASO is a robust parent of ENSO at lags -20 and -22 months. However, for the historical period in the models, ASO is a robust parent of ENSO only for a few selected models but inconsistent in sign as compared to SWOOSH. In the periods of 2011-2051 and 2052-2092 more models show ASO as a parent of ENSO, from lags -10 to -27 months, with more sign consistency as compared to SWOOSH. The main parent of ENSO in the models is ENSO -10 months (i.e. auto-correlated), but not for observations (SWOOSH).

Runge J, Sejdinovic D, Flaxman S. Detecting causal associations in large nonlinear time series datasets. arXiv preprint arXiv:1702.07007. 2017 Feb 22.

(https://arxiv.org/abs/1702.07007)

Runge et al. (2015): Identifying causal gateways and mediators in complex spatio-temporal systems. Nature Communications, 6, 8502. http://doi.org/10.1038/ncomms9502

Pearl, J., Causality: Models, Reasoning, and Inference (Cambridge University Press, Cambridge, 2009, 2nd edition)

Kretschmer, M., Coumou, D., Donges, J. F., & Runge, J. (2016). Using causal effect networks to analyze different Arctic drivers of midlatitude winter circulation. Journal of Climate, 29(11), 4069-4081.

P. Spirtes, C. Glymour, An Algorithm for Fast Recovery of Sparse Causal Graphs. Soc. Sci. Comput. Rev. 9, 62–72 (1991).

Additionally, I think in parts the authors are attempting to use their results to support observational results in Xie et al. but I think if anything their results more clearly indicate the weakness of this relationship between ASO and ENSO, and that it's likely a sampling artifact in the observations. I would encourage them to remove statements such as in the abstract that state that ASO "may also influence the surface in both polar and tropical latitudes"- my impression from the results was actu- ally the opposite, that ASO has very little influence on either polar or tropical surface climate, and that significant correlations can occur randomly for 40-year subsets, which is probably what we are seeing in the observations.

As will be discussed in our new discussion on causality, the observed apparent connection between ASO and ENSO is causal within the framework of Pearl causality, though we agree it is very weak. It is even weaker in the models. We suspect that this weak effect may not be particularly useful in an operational sense, and we plan to discuss this in the revised text.

In general, I think understanding these possible relationships between Arctic ozone and surface climate is important, and the CCMI dataset provides a new tool to do so (particularly since it's hard to find models that are both coupled and have interactive chemistry and decent stratospheric processes). But I would like to see the authors address these concerns. I recommend a major revision.

Specific Comments Page 1, Line 16- maybe mention why ASO has been "spared from the worst ozone destruction", e.g., the relatively warmer polar temperatures due to stronger wave forcing.

We have added " due to the relatively stronger wave forcing from the troposphere"

Page 3, Line 12 – Can you explain what Ref-C2 is, i.e., what radiative forcings, specifi- cations do these runs use.

Full details of the Ref-C2 simulations are described in Eyring et al. (2013); briefly, these simulations span the period 1960–2100, impose ozone depleting substances as in (World Meteorological Organization, 2011), and impose greenhouse gases other than ozone depleting substances as in RCP 6.0 (Meinshausen et al., 2011).

Page 3, comment about Data- you need to include more information about SWOOSH- how long is the record, which version are you using, which data goes into SWOOSH, are you using the “anomaly filled” version, which latitude resolution, etc. One thing I’d be interested to know- how much data does SWOOSH have polewards of 60N to calculate the ASO? Is there any data from 80-90N? If not, do you need to take that into account when comparing the model ozone?

For SWOOSH, we now define ASO as an area weighted mean ozone from 60-degree N to 81.25-degree N and mass-weighted average from 150hPa to 50hPa. The poleward limit of the region used to define ASO is set at 81.25N to match the data available from SWOOSH. We now also clarify that we use the combinedeqfillanomfillo3q product at 2.5-degree resolution with 31 vertical levels, and focus on the period 1984-2014.

Page 4, lines 5-7- is the multiple linear regression performed before or after sub- division into 40 year chunks? (does it matter?). Is this removal of the GHG/ODS effect done for both dynamic (T,Z) and ozone time series? Is this also performed for MERRA2/SWOOSH data?

We perform the MLR before dividing into 40-year chunks, now clarified. It is done for both dynamic and ozone time series, also now clarified. It is also applied to MERRA/SWOOSH, also clarified.

Page 6, line 5-7- change to "ASO and polar cap SLP than is observed in March" (the connection is stronger than observed in April).

Changed as suggested

Also, is the statistical significance true for the r=0.09 value? Or just the April r=0.17 value? Even if it's significant. . . is a value like r=0.09 very useful? It's implying that only 0.8% of the variance in polar cap SLP is explained by ASO. Stating only that it's significant statistically may be misleading (and, I think, not strong support for statements in the abstract or conclusions that suggest such a relationship provides useful information about predictability).

Even the 0.09 correlation is significant due to the >1600 models years available. However, we agree that while these correlations are statistically significant at the 95% level, the variance explained is low and hence ASO may not be particularly useful for prediction of surface climate. This has been clarified.

Page 6, lines 24-27- might acknowledge here that there could be non-linear feedbacks at play that linear regression would not remove

We agree. We have added a new figure where we compute the correlation between polar cap height at 100hPa and polar cap SLP from 1970 to 2010, from 2011-2051, and from 2052-2092 We do indeed find differences among these three periods. The accompanying discussion for this figure highlights that while the linear relationship between ozone and polar cap SLP is indistinguishable from that associated with polar cap height, there is certainly the possibility for nonlinear feedbacks.

Page 7, line 7- in general, I find the authors to be trying too hard in this section to rein- force Xie et al (2016) results; this statement is an example- “This relationship is nearly statistically significant at the 95% level”. This should be changed to either the specific significance level that it meets, or it should say “this relationship is not significant at the 95% level”. Particularly since, if anything is striking about Figure 6, it’s that almost no correlations shown (not even the observed ones) meet significance levels. These results to me more strongly argue that the relationships suggested by Xie et al. are artificial.

We now state that these correlations are not statistically significant at the 95% level.

Xie et al in contrast claims that it is significant, and we now note that there is a difference in the level of significance between our results and those of Xie et al.

Page 8, line 29-30 more effectively state what I think the results of this study conclude- that there is no strong evidence of this mechanism/relationship in the CCMI models. But this message isn’t clearly reflected in either the abstract or the conclusions (such as point (3) on page 11, line 1)

When looking at the CCMI models from 1970-2010, we do not see strong evidence of the relationship between ASO and ENSO nor to the mechanism suggested by Xie et al. The observational data (SWOOSH 1984-2014), however does seems to be compatible with their results. In addition, for the CCMI models for the year 2011 and on, the prediction power of ENSO based on ASO seems somewhat larger. Hence while the relationship between ASO and ENSO has not yet been proven to actually be helpful in an operational sense, the relationship does appear to exist.

The revised version of the text will include a better discussion of this issue in the abstract and conclusions.

Page 7, Line 22- could this have to do with the power spectra of ENSO in models having higher amplitude at periods of 12 months instead of 24 months? (e.g., AchutaRao and Sperber 2002). Again, this would suggest these lead-lag relationships are more a reflection of ENSO auto-correlation than physically-based relationships.

The causality argument now takes into account ENSO autocorrelation explicity. The models indeed do not show any causal influence, though the observations do.

Page 9, Line 1-10- I commend the authors for trying to deal with the auto-correlation is- sue. However, I'm not sure I understand their method. Why is ASO at a lag of 3 months chosen? Why not at the lag where the relationship peaks, in either the observations or the model? Then the "minimally useful correlation" is shown in equation (2) and plotted in red in Figure 6/7, but it's not clear to me what the right side of that equation has to do with the correlation values in plotted in black (yet it's then stated that "for both observations and the CCMI models the actual correlation between ASO and ENSO far-exceeds the minimally useful one". . . but the actual correlation r(ENSO,ASO) is not part of the criteria in equation 2). Also confusing is that the minimally useful correlation appears to be nearly zero at all lags, so how is the criteria in eqn(2) satisfied? This should be clarified to better explain what this analysis tells us, but I also would recom- mend that instead of this method, additionally consider applying the Granger causality techniques as detailed in McGraw and Barnes (2018).

We have accepted the reviewer's suggestion, and now use causality techniques. The minimally useful correlation arguments have been removed.

Page 9, line 13- this relationship at zero lag is true only in the multi-model mean CCMI, right? It seems to be the opposite sign in the observations.

Over the period of 1984 through 2014, the correlation of ENSO with polar cap geopotential height was indeed quite weak. See Hu et al 2017 and Domeisen et al 2019. We now note this apparent decadal variability.

Domeisen, D. I., Garfinkel, C. I., & Butler, A. H. The Teleconnection of El Niño Southern Oscillation to the Stratosphere. *Reviews of Geophysics*.

Hu, J., Li, T., Xu, H., & Yang, S. (2017). Lessened response of boreal winter stratospheric polar vortex to El Niño in recent decades. *Climate Dynamics*, *49*(1-2), 263-278.

Page 10, lines 1-17- this part of the conclusions was nicely written and well-phrased.

Thank you!

Technical Corrections Page 3, Line 6- Capitalize the appropriate letters for MERRA

fixed

Page 3, Line 18- remove repeated “the”

Fixed

Page 5, Line 8- remove comma after heights and put it instead after the first word “stratosphere”.

fixed

Page 5, Line 17-18- should be “with a single x, and the. . .”. Also it should be yellow asterisk, not green?

Fixed

Page 5, Line 27- remove “polar cap” and change to “sea level pressure anomalies” (I assume climatology is removed?). Could also add “anomalies” in line 31 (and else- where throughout paper). Figure 5a shows the correlation of sea level pressure anoma- lies at each grid point with the ASO in March, right?

fixed

Page 5, Line 30- semi-colon instead of comma after

Ivy et al. 2017.

fixed

Page 9, line 5- should be “3 months”

fixed

Page 10, line 10- change “an” to “a”
fixed

Page 11, line 5- capitalize “acknowledgement”

fixed

Page 14, caption- misspelled “stratosphere” on line 3.

fixed

Should these be stated as anomalies in T and Z or are these full fields? (also true in other captions)

Fixed

Also the y-axis on panel (h) seem mislabeled (should be ZApr?).

fixed

Note also in y-axis of (a) and (b) that the correlation value is 0.36

on one and 0.35 in the other; I believe this should be the same number.

It is not the same number – one is for March the other for April. In any event we have removed this from the figure, and now include a large black X to represent the multi-model mean.

C
5