Atmos. Chem. Phys. Discuss., 14, C6221–C6230, 2014
www.atmos-chem-phys-discuss.net/14/C6221/2014/

Atmospheric Chemistry and Physics
Discussions

Interactive comment on "Tracing the second stage of Antarctic ozone hole recovery with a "big data" approach to multi-variate regressions" *by* A. T. J. de Laat et al.

Anonymous Referee #1

Received and published: 26 August 2014

1 Overall

The paper investigates the possible turnaround of Antarctic spring ozone depletion using multiple linear regression analysis to establish the underlying long-term trends. Similar previous papers typically use one fixed set of predictors / proxies in the regression, and determine statistical significance from the difference between observed and regressed time series. Different from this standard approach, the authors now vary their input /predictor/ proxy time series, and their observed time series. Variation is done over a wide range using random Monte-Carlo simulations. The probability

distributions resulting from this "Big-Data" Monte-Carlo simulation approach are then considered and checked for significant or insignificant ozone trends before and after a possible turn-around point in the late 1990s.

Since the major fraction of their simulations does not show a significant ozone increase after the late 1990s, the authors conclude that there is no robust statistically significant positive trend in Antarctic springtime ozone since the late 1990s, and the 2nd stage of ozone recovery, which requires such a trend, has not been reached.

Overall, I think this is a worth-wile new investigation of the uncertainty of regression models. The paper is generally well written and well suited for the scope of ACP. I do have some reservations though, where I feel that the presented evidence is not conclusive or does not fully support the conclusions.

2 Major Questions

Clearly, the number of regressions with significant trends depends to a large degree on the range of uncertainties sampled by the Monte-Carlo simulations. If a large number of regressions use unrealistic proxies, or even "bad" proxies, these regressions are unlikely to find significant trends - even if there really is an underlying trend, that would be significant in a "good" regression. So I feel that the authors need to be more careful about the impact of varying the proxy ranges/ uncertainties on their results. Is there a way to do this? Can the tested proxy ranges / uncertainties be reduced and the effect on trend significance be investigated? Is there a way to "weigh" regressions/ proxy choices, e.g. by the R^2 they achieve? Clearly regressions that only explain a small fraction of the observed ozone variation are not as "good" as regressions that explain 90% of the observed variation.

My largest concern in this respect is the combined QBO-solar proxy. Here the authors assign uncertainties of ±200%, which in terms of correlation means correlations less than 0.4. To me it seems that a large fraction of the tested QBO-solar proxies then has little or no correlation with ozone anymore, and probably should not be in the regression in the first place.

Construction of the combined QBO-solar proxy (Section 2.5) is also not clear to me. QBO phase is very important. How is the correct QBO phase (or pressure level) chosen in this proxy? How is it varied in the Monte-Carlo simulations? QBOs out of phase with ozone are not useful and should be omitted. Why is only the complex combined index tested? Why are simple solar cycle and simple QBO not tested also?

The effect of including / excluding extraordinary years should definitely be tested as well. Clearly 2002, the year of the vortex split is a key year. One glance at the top panel of Fig. 2 shows that including / excluding 2002 potentially has a very large effect on trends after 1999, depending on how well the other proxies can "explain" the high ozone values in 2002. Most likely, in my experience, they will not fully explain the 2002 high anomaly. When 2002 is included, most post-1997 or post-1999 trends will be biased low because of the positive outlier at the beginning of the trend interval, and will be more uncertain, because 2002 lies far away from the other data points. So I feel that the authors should also consider scenarios where 2002 is excluded.

Similar arguments could be used for the "volcanic" years 1983, 1984, 1992, 1993. Like the authors, I have reservations about using the aerosol proxy.

To summarize these points: I think the paper needs additional discussion and additional consideration of the impact of the wider or narrower range of tested proxies / predictors and the effect of that range on the fraction of significant trends. See my "cherry picking" remarks below. Including/excluding 2002 should also be tested.

3 Minor Points

I do agree with the other reviewers: A revision of the entire text, with the goal of bringing out the main points clearer and crispening things up would be beneficial.

Abstract I wonder if the abstract is a bit too pessimistic. To me it seemed that October ozone and September EP flux gave a lot of significant trends - a combination that I would expect to be good for showing turnaround. Also not only the 95% confidence level, but also the 67% confidence level should be considered for the main conclusions.

pg. 18594, line 13: Suggest to replace "in fact" by "surprisingly"

pg. 18594, line 21: Suggest to insert "of most ozone trend analyses and" before "of the suggested"

pg. 18594, line 22: Suggest to insert "Antarctic" before "ozone"

Section 2: Suggest to include the URLs for the regressor time series in the normal text

pg. 18597, line 1: I do not understand why solar cycle and QBO are not also included (or tested) as single regressors.

pg. 18597, lines 7 to 10: Are the PWLT and LINT not fitted at the same time as the other regressors? If not, then I have to doubt the entire paper!! If you first fit C_1 to C_4 and then fit trends to the reminder, results may be very different!!

pg. 18597, line 16: maybe add " R " after "observations".

pg. 18599, line 12: "Kleiciuck" should probably be "Klekociuk"

pg. 18601, lines 1-3: I think this requires a short discussion that the phase of the QBO / the ozone variations is important. As QBO phase varies with height, the phase can be changed by using a different height. This is not really discussed / explored here. ALSO: Most regressions use two orthogonal QBO proxies, so that the regression can "pick" the right phase. This is also not done here, and as mentioned above simple

QBO and simple solar cycle regressors are not tested at all, instead only the complex QBO * solar cycle regressor (with the correct phase??) is used. I think this needs more discussion, and maybe additional regressions are required.

The implicit assumption of the entire paper is that regressor uncertainty translates directly to ozone uncertainty. This is not really the case, none of the regressors is usually perfect. The chain of processes that translates a QBO wind variation to Antarctic spring (or other) ozone variations is very complex, partially random and certainly not linear. So there is no "perfect" regressor, and regressor uncertainty is not a source of "error" in the regressions. It is just a way of exploring uncertainties of the performed regressions. I think some text needs to be added, maybe in the introduction, and most likely in the description of the regressors and their uncertainties.

pg. 18601, lines 10 - 27: My last comment also applies to the Solar Cycle proxy discussion here. Even though solar UV most directly influences ozone in the upper stratosphere, the chain of processes that affect Antarctic Springtime column ozone is very complex. So F10.7 may be just as good, or even better than Lyman-α as a proxy.

Section 2.5: What QBO level is used? How are the different time series in Figure 2 constructed? Needs more explanation! Why are the uncertainties assumed so large? What happens if just QBO and / or just solar cycle are used? What happens if no QBO / solar proxy is used at all? I think these tests should be done and should be described.

Section 2.8: I think the recent aerosol measurement papers by Vernier et al., and Trickl et al., should be cited here as well.

pg. 18609, line 12: For explanation of the "100 x 100 x 8 x 8 x 6 x 3" scenarios, it would be good to add a reference to Table 1.

Table 2: I have several points about the numbers in Table 2:

- As mentioned by the other reviewers, a brief description of how the EESC trends are determined is needed somewhere.

- I assume the following: The regression using EESC gives DU/ppt, and using the EESC ppt/yr before and after turnaround this can be multiplied to give DU/yr before and after turnaround.

- First question: Why do the EESC trend uncertainties in this study not scale as the trends themselves? -5.26 ± 0.21 does not fit together with 1.02 ± 0.12; If the 0.12 is correct, the 0.21 should be $\frac{5.26}{1.02} \times 0.12 \approx 0.6$, not 0.21! Same thing for the 2000 to 2012 trend uncertainties. The Kuttipurath et al., results in Table 2 seem correct though.

- 2nd question: Is the large 9.88 DU/yr uncertainty of the 2000-2010 PWLT trend correct? It is 3 times the uncertainty from Kuttipurath. Looking at Fig. 2 top panel, I have a hard time believing a $\pm 100 DU/yr$ uncertainty. A -$60 DU/yr$ trend seems very unlikely from that figure, if the regression does a halfway good job taking out variability. What happens if 2002 is omitted?

- Generally: Why are all 2000-2010/2012 PWLT uncertainties from this study so much larger than the Kuttipurath et al. $2.73 DU/yr$ uncertainty? Are wrong numbers given? Since uncertainties are so important here, this issue needs to be clarified and needs to be discussed.

- The LINT uncertainties seem much more reasonable and believable. I think it would be very important to also give LINT trends and uncertainties for the case where 2002 is removed from the data. LINT trends should be very sensitive to the 2002 outlier. (Or not sensitive, if the EP-flux regressor accounts well for 2002.) PWLT trends should be less sensitive, because the PWLT low point is determined by the slope before 2000.

- As mentioned, if they turn out to be correct, the large differences in PWLT trend uncertainties between this study and Kuttipurath need to be discussed.

Section 3.3 / Figure 4: As mentioned, a brief statement is needed to explain how the EESC trends in DU/yr were determined.

Very early on, there also needs to be a clear statement that the trimodal structure of EESC trends is purely an artefact of using three EESC curves. The current text on page 18611, lines 19 to 22 is not very good/ clear and should be improved. I would suggest to replace both sentences by something like: "The EESC trends show a tri-modal distribution, because only three different EESC curves were used (see also Fig. 3). These three EESC curves differ less in the pre-1999 slope, much more in their post-1999 slope."

pg. 18611, lines 15 to 27: BREAK is not consistent with Figure 4 where only 1999/2000 appear in the plot title and in the caption. What has actually been done? Could Figure 4 not show results for different BREAK years as well? Would those results be much different from 1999 only as a BREAK year.

Throughout the manuscript, the varying use of BREAK years (sometimes 1997, sometimes 1999, sometimes 2000, sometimes 1997, 1998 and 1999) needs to be cleared up!!

pg. 18611, lines 23 to 24: I don't see/ understand why PWLT would necessarily result in larger trends. The pre-BREAK PWLT trend is actually smaller, only the post-BREAK PWLT trend is larger. I suspect that the smaller post-BREAK LINT trend has a lot to do with 2002. It would be (again) interesting to see what would happen without 2002!

pg. 18612, line 3: Add "In the bottom panel, " before "Figure 4"

pg. 18612, lines 5-6: I would expect that the LINT correlations are larger than the PWLT correlations, because they have an additional degree of freedom. I am slightly surprised that the EESC correlations are larger than LINT or PWLT correlations. I guess this means that the smooth EESC turnaround fits ozone better than the linear trends. Wonder what a parabolic trend would do and would tell us? Would a significant

positive coefficient of the quadratic term and a minimum sometime around 2000 mean ozone recovery? Maybe comment on those ideas.

pg. 18612, line 9: How small? Give numbers!!

pg. 18612, lines 10-11: I think these papers talk about a different auto-correlation, e.g. the one between spring ozone anomalies and summer/ fall ozone. Here you talk about auto-correlation of ozone residuals, after much of the other auto-correlation has been removed by the regression.

Figures 5 and 6: I assume that the one much higher curve in all these plots is the overall distribution / sum of all the other curves. I wonder if it is necessary to plot these overall curves. To avoid the extra-explanation (which is missing right now), and for clarity it would probably be better to omit these extra-curves, which I found confusing.

Figure 6, pg. 18613, lines 6-18: The discussion that aerosol has virtually no effect on the ozone regression, and essentially gives random results is basically fine, and could even be tightened up more. There are only 3 or 4 data points that are affected by aerosol!

I do not understand what is shown in the bottom panel of Fig. 6. Are aerosol regression coefficients shown or EESC regression coefficients? Why are there three EESC scenarios in the bottom panel, but 5 aerosol scenarios in the top panel? Should the 1st EESC in the title of the bottom plot not be dropped? Should the 1st "EESC" in the figure caption not be "aerosol"?

Overall I would suggest to drop Figure 6 completely. All it does is further elaborate on the point that aerosol is a fairly meaningless regressor and should probably not be included at all. The existing text essentially says that already. The Figure in confusing as it is now, and I would omit it.

pg. 18613, line 25: Replace "leads to" by "correlates with". The SAM index is a concept, it does not lead to ozone depletion.

pg. 18614, line 13: What is meant by "explanatory power"? Is that R^2? If so, then that should be added.

pg. 18614, line 14-15: Where does that come from? Is that shown in a Figure or Table? Needs more explanation.

pg. 18614, line 17: Maybe add "(see Fig. 5A, 5B)" after "deficit"

pg. 18615, lines 1 -4 : It would be good to have a table showing the trends and their uncertainties for these cases.

pg. 18615, line 6: The solar-flux / QBO variations are definitely not "random". If they are random they are wrong, and they should not be in the regression!! Please reword.

Entire Section 3.6: I agree with the other reviewers that this section is a little bit at odds with the general concept of the paper. Section 3.6 is picking the "good cherries". The paper in general, however, says you should check "all cherries". Now to me the question is "what about the bad cherries"? What "cherries" should be thrown out? I think finding and describing this balance is an important aspect that is currently still missing.

pg. 18615, lines 13-22: I find this EESC discussion awkward and confusing. The point is that you cannot use EESC to identify increasing ozone, because the EESC curve shape is prescribed. There is no degree of freedom allowing different trends before and after turnaround. Please reword! Probably best to shorten these lines.

pg. 18616, line 10: I took me a long time to understand, why the cumulative fraction of significant trends in Figure 9 does not reflect the numbers given in Table 4. I think Table 4 needs 2 extra rows giving the cumulative fractions for all start years and 2010 and 2012 end year, respectively. These extra rows will then jive with the red shaded curves in Figure 9. Also, for consistency with the sequence in Table 4 (2010 end year, then 2012 end year), I suggest to swap the two panels of Figure 9. Also, the 2010 curve should be deleted from the 2012 panel of Figure 9, because it is already shown

in the 2010 panel. Little things, but helpful to a reader!

pg. 18616, line 23, caption of Table 5: What is a "statistically significant regression"? You probably mean regressions with statistically significant post BREAK trends. 1 σ? 2 σ? Please define.

Tables 5 and 6, and their discussion bring us right back to the "cherry picking" question. Clearly these tables show some good cherries and some bad ones? Which ones can we use for the "cake" of possibly increasing ozone? I think the authors need to come up with a better answer than they are doing right now.

pg. 18620, line 20: "Shephard" should be "Shepherd". (I'm not Ted Shepherd, though.)

References: Need to be checked. Some of the discussion papers are published now. I think some references are missing, e.g. Newchurch et al., 2003.

Figures 4 and 7: I think it would be good to compare the width of these distributions with the typical (average?) uncertainty for the regressed trend (or other coefficient) coming out of the regression residuals. Maybe add a horizontal error bar in a corner of the plot. Also discuss in text. The middle panel of Fig. 4, for example gives a width of about 2 DU/year (FWHM), roughly equivalent to $\pm$ 1 DU/yr (1σ) or $\pm$ 2 DU/yr (2σ) - quite comparable to the Kuttipurath PWLT uncertainty in Table 2.

Overall: Quite a good paper - needs a bit more work and then it will be good to go through to ACP.

Interactive comment on Atmos. Chem. Phys. Discuss., 14, 18591, 2014.