Atmospheric
Chemistry
and Physics
Discussions

Open Access

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 #3

Received and published: 14 August 2014

The manuscript "Tracing the second stage of Antarctic ozone hole recovery with a "big data" approach to multi-variate regressions" by De Laat et al. describes and analysis with which the uncertainties of basis functions for a multi-variate regression for determining Antarctic spring ozone trends are estimated. The focus of this study lays especially on the statistically significant detection of the onset of ozone recovery in this region. Historically, the uncertainties of trend estimates derived from multiple linear regression methods are based on the uncertainty of the fit of the trend or EESC basis function to the measured ozone time series. However, normally the uncertainties of the basis functions themselves are not considered in this method. In this study,

uncertainties for the most commonly used basis functions for a regression model applied to Antarctic ozone are estimated, and their effects of the detection probability of a significant ozone trend is described and discussed.

Some studies have shown already that the ozone recovery has already begun in the Antarctic region in spring. Statistically significant trends in total column ozone have been reported, based on multiple-linear regression models. In these cases the uncertainties in the basis functions used is not considered in the regression. A careful analysis of basis function uncertainties and their effect on trend detection is therefore very useful, especially since ozone recovery is a highly interesting topic for the public and policy makers. In my opinion, the "big data" approach of this study is an interesting and necessary try to achieve this.

The manuscript is in mostly well written, and the grammar and style are fine. However, there are a few concerns I have with the description of the approach, the description of the results, and the discussion on the their relevance. I would recommend publication after the major and minor concerns that are listed below are addressed:

Major concerns:

- It is stated in Section 2.1 (page 18597, last paragraph) that the main focus of this study is "to investigate trend significance, not specifically what parameters can best explain Antarctic ozone", however, in Section 3 many individual combinations of basis functions are pointed out if they have the good explanatory power. This is extremely confusing for the reader, and makes it very hard to understand the main message of the analysis. In fact, it actually hinders the reader to focus on the main message. In my opinion, the whole manuscript has to be checked to achieve the right balance of a detailed description of the main message (the effect of uncertainty of basis functions on trend detection), and some side notes about good basis function combinations.

- Most sections of the manuscript provide much detail about the specific topic of the section. However, in almost all sections, the main message of the section is hidden

somewhere in the details (very hard to grasp as a reader), or the details are focused on something only somewhat related to the main message, so that too many details are provided for the related topic, and not enough for the main message. I strongly suggest re-evaluating each section, and checking if the main message of the section is clearly described, with enough information provided.

- The determination of a trend (given in DU/yr) based on an EESC time series is not directly straightforward. It should be explained in the text how the trend is derived from applying the EESC basis function.

- This could then also clarify the dismissal of EESC as valuable basis function for trend detection as it is suggested here in this study. EESC is widely used in regression models, and the discussion and explanation on why EESC should not be used for trend detection is way too short in this manuscript. It is shown that the use of EESC in the regression has the best explanatory power (page 18614, Sect. 3.6), but in the first paragraph of Sect. 4 EESC is dismissed as valuable basis function. The paragraph is not detailed enough to provide a satisfying explanation.

- There is no further discussion about the choice of ozone values, although it is stated in the introduction that this is one of the points that are supposed to be systematically analyzed. The different ozone scenarios are explained and described, however, the choice of using ALL months in the regression and extracting seasons/individual months per Fourier pair fitting is never mentioned and not discussed, although this is one of the main methods used in regression analyses.

Minor concerns:

- The ozone data set that is used in this study is only briefly mentioned in the introduction, but never explained in detail in the ozone scenarios section (section 2.9). The ozone data is the whole basis of this analysis and should therefore be discussed in more detail. The authors mention in Section 2.10 that the uncertainties in the trend detection dominate the uncertainties due to different ozone data sets. However, it is

important for the reader to know exactly which data set is used, since (1) most total column ozone data sets do not provide data during polar night, and (2) the results can be reproducible, and do not suffer, like stated by the authors in line 26 on page 18617, from lack of information about how the ozone record was derived.

- It is confusing why QBO and Solar Cycle basis functions get a separate section although they are not used in the regression model separately. They should be described, but I think it would be clearer to describe them under the header "Mixed solar-QBO index" (which could have sub-headers "QBO", "Solar Cycle" and "Construction so the mixed index")

- I would recommend providing the URLs for the sources of the basis functions in the text, not the supplementary material, since it helps to understand the description in the text where the data came from.

- Page 18598, line 17: ". . . the period for which. . ." -> period is confusing in this sentence. I think it might be better to use "months" or "seasons". . .

- Page 18600, line 24-25: statement ". . . since various studies have shown. . ." needs references (e.g. Hood, JGR, 1997).

- Page 18601, line 11-12: First statement of this paragraph needs a better reference for the effect of solar radiation on stratospheric ozone.

- Page 18603, line 12: "Both time series. . ." -> it is not clear what the "both" refers to.

- Page 18607, line 11: ". . . the analysis presented in the previous section shows. . ." -> "section" does not make sense here, do you mean "paragraph"?

- Page 18607, line 16-23: the explanation of the shifting of the peak aerosol loadings is not clear here. Please clarify what you did exactly.

- Page 18608, line 28: reference "Newchuch et al." should probably mean "Newchurch et al.", and it is not cited in the reference list

- Page 18610, line 6: it is not clear to which regression you refer as "basic regression". Please clarify.

- Page 18610, line 13: "…on the PWLT is a bit larger…" -> there is no mentioning or explanation of the uncertainties on the PWLT trend which are ~3.5 times larger than reported in Kuttippurath et al. (2013)

- Page 18611, Section 3.3: Figure 4 shows the results for the piecewise trend regression 1979-1999, and 2000-2012. But in the text this figure is discussed as if the break year would be flexible. I understand that there regressions with different break years have been analyzed, however, when discussing Figure 4, all mentioning of "BREAK" is highly confusing and does not help the reader understand what is important here.

- Page 18611, line 26-28: The statement in the sentence starting with "In addition,…" needs more explanation than is given at the moment.

- Page 18612, line 9-13: This paragraph should not be a "passing" comment: Treating auto-correlation in ozone records used for trend detection is an important part of the statistical analysis. It should be explained here in more detail why the authors think it is not important in any of their 35 million regressions. Might even be worth a separate section.

- Page 18612, line 24: Reference to "Fig. S1" should really be "Fig. S2", I think.

- Page 18613, line 1: Reference to "Fig. S2" should really be "Fig. S1", I think.

- Page 18614, line 24-26: Sentence starting with "For all regression models…" -> the meaning of the sentence is not clear, please rephrase.

- Page 18615, line 19-23: Two sentences starting with "As discussed in Sect 2., …" -> The reasoning for dismissing all EESC-based ozone recovery results are not clear. In Section 2 it was stated that the uncertainty of the EESC basis function would be based on the results of three different EESC scenarios. Is that not valid anymore?

C5884

- Page 18616, line 1-15: Figure 9 is not very well described in this section. This figure is complex, and definitely needs a more careful description in the text.

- Page 18619, line 11-14: Sentence starting with "The EESC scenario…" -> it is not clear why it would be expected that one particular scenario reflect that the ozone trends are highly sensitive to the chosen scenario. Please explain/rephrase.

- Page 18629, Table 4: It is not clear to what "start year" and "end year" given in the table refers. This has to be explained with a little more detail. As well, it might be worth adding the unit "year" to "length".

- Page 18630, Table 5: Please state in the figure caption more clearly for which regressions these numbers are relevant (Period ending?)

- Page 18632, Figure 1: In caption, second to last line, "Data is…" must be "Data are…"

- Page 18633, Figure 2: In caption, it is not clear to what the 16 combinations refer. Please explain in more detail either here or in the text.

- Page 18634, Figure 3: Legend for stratospheric aerosol: "Pt=2.5 El C" is supposed to be "Pt=2.0 El C"? At least that's what is stated in the text.

- Page 18636, Figure 5: it is not clear what the thick dark blue lines in graph [A] and [C] represent, and it is not clear what the thick red lines in graph [B] and [D] represent. Sum of all counts?

- Page 18639, Figure 8: lowest line of right panel is labeled "EEASC". Should be "EESC".

References: Hood, L. L.: The solar cycle variation of total ozone: Dynamical forcing in the lower stratosphere, J. Geophys. Res., 102, 1355– 1370, 1997.

———————————————————

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