Atmos. Chem. Phys. Discuss., 12, C13617–C13625, 2013
www.atmos-chem-phys-discuss.net/12/C13617/2013/

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Comparison of three vertically resolved ozone data bases: climatology, trends and radiative forcings" *by* B. Hassler et al.

B. Hassler et al.

birgit.hassler@noaa.gov

Received and published: 27 March 2013

We would like to thank Ted Shepherd for the comments, praise and concerns about our manuscript. We tried to answer these with detailed explanations below. Original comments are printed in black, answers are printed in blue.

Short comment from Ted Shepherd

It is certainly useful to document the differences between these three ozone data sets and the reasons for those differences, and the authors are to be commended for the thorough fashion in which they have done this.

Thank you, and we appreciate your commendation.

The problem I have with this paper, from a user perspective, is that it is a mixture of apples-to-apples and apples-to-oranges comparisons, and it is important for the reader to know which is which. Yet they are treated in essentially the same fashion by the authors. The apples-to-apples comparisons involve issues such as using different ozone observations, different vertical resolution, etc. There is only one atmosphere and uncertainties in reconstructing the past behaviour of the atmosphere are uncertainties of an apples-to-apples nature. The well-known issues with RW07 in representing the Antarctic ozone hole certainly fall into this category. However, the use of different basis functions is an apples-to-oranges comparison, because the explained signal and the trend mean different things in each case. For example, it is said several times that BDBP fits the observations better in an RMS sense than the other data sets; well of course it does, because it includes more of the variability!

All three data sets are individual things. In this manuscript we only compare the different available data sets to each other, but we don't create them. We state what is available, and it is not our intention and not our task to change the data sets to a point where they would be comparable on an apple-to-apple basis, e.g. all using the same basis functions. The three data sets were made differently for several reasons, which we state in detail in the manuscript in Section 2 (data set description section) and Section 8 (Discussion and Summary section). RW07 was designed to be a data set, which includes much but not all of the natural and anthropogenically-caused variability, and which could be created by conventional regression methods. The SPARC data set was specifically created to serve as boundary conditions for the planned CMIP 5 climate model runs, and therefore had to cover the period 1850 to 2100. The choice to "only" include the EESC and solar cycle basis functions for the observational period was made, as far as we know, because the creation of, for example, a realistic QBO or ENSO time series before observational period and after the observational period proved to be very problematic. The small amount of variability included in the SPARC data set was therefore a choice of the people creating it, to ensure a smooth transition between several different time periods (historical, observational, future) to facilitate use in climate models. The BDBP data set was specifically created to describe as much of the observed variability as possible over the period of direct observation. For this reason a larger collection of ozone measurements and basis functions was used, as well as a specifically created regression method to deal with large data gaps was created.

We do not agree that a comparison of these three data sets as we show it in our manuscript is an apple-to-oranges comparison. We compare the data sets as they are. This is what you get if you want to download the data sets from the place where they are freely available. With our comparison we are helping users understand the differences between the data sets, regarding their creation, the observations they are based on, and how they compare to independent observations and already published analyses. Users will still have to decide for themselves what data set would be best suited for the purpose of their intended study.

We compare the three data sets in several different ways with observations: as climatologies (Section 3), as anomalies (Section 4), as individual time series (Section 5), and as a sum of overall distance to observations (Section 8). Only for the latter comparison do we describe the BDBP as being closer to the observations in the RMS sense. For the comparison of the climatologies and the comparison of the individual time series, the exact variability of the different data sets is not important.

Most seriously, the use of different basis functions leads to different conclusions with respect to the 1979-1996 trends. In describing the ozone changes, the authors use "decrease", "depletion", "loss", and "trend" synonymously. It is very confusing to use multiple words for the same thing, and there seems to be an implicit assumption that all these quantities are deterministic responses to forcings rather than stochastic noise. The authors need to be crystal clear on what they mean by these terms, and if they think they represent deterministic effects that should be reproducible by models. Especially because the paper is ostensibly being written for modelers, it really needs an upfront discussion of the philosophy behind the different choices of basis functions, and how these might affect the use of the different data sets.

Since this paper describes only a comparison of three different data sets, and not their creation, it is not possible to provide a discussion about the philosophy of the basis functions chosen for the creation of the different data sets. The only real information about the selection of the basis functions is available in the respective publications:

SPARC: Cionni et al., 2011, page 11269: "The regression includes terms representing EESC and 11-yr solar cycle variability. Thus, other sources of interannual variability, including volcanic eruptions and the Quasi-Biennial Oscillation (QBO), are removed."

RW07: Randel Wu, 2007, D06313: "Decadal trend changes in ozone are modeled using the equivalent effective stratospheric chlorine (EESC) proxy shown in Figure 2a [. . .]. The solar cycle proxy is the standard F10.7 radio flux (Figure 2b), and QBO1 and QBO2 are two orthogonal QBO time series based on observed equatorial stratospheric winds [Wallace et al., 1993; Randel and Wu, 1996]. We include a constant plus annual harmonic term for each of the regression coefficients in equation (1)."

BDBP: Bodeker et al., 2013, page 33, two paragraphs. In summary: The regression model includes an EESC basis function that is dependent on the age of air, a linear trend term to "to account for linear changes in ozone, for example, that may result from secular changes in green- house gases", two QBO basis functions that are orthogonal to each other "to permit fitting of the phase", and El Niño–Southern Oscillation (ENSO), solar cycle, and Mt. Pinatubo basis functions.

The reason why some basis functions were not included in the regression models for the data sets can only be guessed. And it might not even be a philosophical question, but rather an adjustment to circumstances and data availability (see comment above). We therefore decided not to add a discussion about the selection of the different basis functions for the three data sets to the manuscript.

We agree that we use many words ("trend", "decrease", "depletion" and "loss") to basically describe the same thing and that this might be confusing for the reader. We therefore checked all occurrences of the four words and adjusted them when appropriate. We decided to remove the term "loss" altogether since it implies the reduction of ozone due to enhanced chemical depletion, and while describing regression model results we cannot be sure that the ozone decrease in the data sets is only due to chemical reactions and not also partly due to dynamical interactions.

Most variability seen in the three data sets are a result of a combination of effects caused by specified basis functions and is therefore mainly deterministic. The fits of the basis functions can also be influenced by noise on the data. How much of the overall variability should be reproduced by the models depends, in our opinion, strongly on the scientific question that is being addressed: case studies of short time periods and a specific region of the globe might want to reproduce more of the variability than global, multi-decadal runs. The determination of what part of the variability can be reproduced by models is, in our opinion, beyond the scope of the paper.

In particular, by including ENSO and volcanic basis functions, BDBP obtain a much larger Arctic ozone trend over 1979-1996 than the other two data sets. What is the user to make of this? If the effect was due to ENSO, which is known to affect Arctic ozone, then one might wish to include that effect in a model run with prescribed SSTs, but one certainly would not want to include it in a coupled model run or a run with prescribed modeled SSTs. If, on the other hand, the effect was related to the volcano term, then this raises the issue of whether the observed very low Arctic ozone values in the mid-1990s were really a deterministic response to the volcanic forcing. The authors seem to imply this (p. 26572, lines 8-11), but none of the three studies cited addressed polar ozone changes. It is well established (e.g. the last few Ozone Assessments) that the low Arctic ozone values in the mid-1990s were due to a combination of a series of very cold winters combined with high chlorine loading (Tegtmeier et al. 2008 GRL is probably the most comprehensive attribution study), so the argument would need to be made that the series of very cold winters was the result of the volcanic forcing. I am not aware of any solid literature basis for such a claim. At the very least, the reader needs to be told that this is the implicit assumption behind the claim that the BDBP

Arctic ozone trend is really a trend and not just some natural variability that happened to alias onto the volcanic forcing term in the MLR model.

We agree that our explanation of why the Arctic ozone trend is much larger than the trend seen in RW07 and SPARC was not sufficient enough. While the negative Arctic anomalies in the winters following the eruption of Mt. Pinatubo are most likely influenced by the volcano, the negative anomalies in the late 1990s and early 2000s are most likely not. Bodeker et al. (2013) suggest that the underestimation of the integrated ozone in the Arctic is attributed to dynamically forced increases in ozone that are not reproduced by the regression model ("The integrated total column ozone from the Tier 1.4 database agrees well with the 4 independent time series over the Arctic. The suppression of total column ozone following the eruption of Mt. Pinatubo, seen in the independent observations, is tracked well, although the onset may be slightly too early. Through the late 1990s and early 2000s the spring maximum in total column ozone is occasionally under-estimated in the Tier 1.4 database, particularly in 1998, 1999, 2001, 2002, 2004 and 2006. This underestimation is likely the result of dynamically forced increases in ozone (Hadjinicolaou et al., 2005) which cannot be tracked in the regression model since it does not include an appropriate basis function.").

We added a few sentences about this to the manuscript and reworded the paragraph about the effects of Mt. Pinatubo eruption on Arctic ozone decreases slightly to be more precise in our description.

In a similar fashion, the authors introduce the use of a "hockey-stick" fit as an alternative to EESC (p. 26579, lines 4-5), arguing that perhaps the dependence on EESC is nonlinear and therefore not properly captured by an EESC term in MLR, even though there has never been any solid evidence that this is the case. But then it seems a bit disingenuous to only examine the pre-1996 and not the post-1996 part of the fit; one cannot ignore "inconvenient truths"! Vyushin et al. (2007 JGR) compared EESC and hockey-stick fits to extra-polar total column ozone. There was hardly any difference for the SH, but for the NH, there were big differences between the two, similar to what is seen here (i.e. with the hockey-stick fit picking up much more of the pre-1996 decline). One then needs to think very carefully about the implications of the differences. Vyushin et al. did not make a judgement as to which was better, but noted that if the hockey-stick fit was to be believed, then ozone recovery (in the sense of a statistically significant increase) was on the verge of being detected in 2005, which seemed physically unlikely. None of this sort of discussion occurs here, which leaves the user helpless in deciding which data set to use. What I would have found interesting would be to compare the component of the trend that was congruent with EESC in each case, as this would be an apples-to-apples comparison. This would be very useful for the modeling community.

In Section 6 (Trends) of our manuscript we describe trends that we determine using a multiple linear regression model on the three different data sets. Each of the three data sets, however, is already a construct of a multiple linear regression that was used to fill data gaps between available measurements to create gap-free representations of the ozone distribution from 1979 to somewhere in the 2000s (depending on the data set), and from the earth surface or the tropopause to the upper stratosphere or lower mesosphere (depending on the data set). For all three data sets an EESC basis function was used in their regression model to describe the ozone decrease that is primarily due to anthropogenic emissions of chlorine and bromine compounds. In the regression used to create RW07 and SPARC, only one, unchanging EESC time series basis function is used, whereas in the regression used to create BDBP the EESC time series is dependent on the age of air and therefore changes shape slightly depending on latitude and atmospheric region.

Section 6 of our manuscript was not intended to be a trend analysis to reproduce a realistic description of the observed ozone decrease, or to serve as a causality trend analysis. Our intention was to find a measure to describe the ozone decrease that is contained in the three ozone data sets to be able to compare that decrease with the ozone decrease in observations.

Since the three data sets used different EESC time series, we decided to not use EESC in our regression. It would not have been possible to describe the variability of the three data sets that was assigned to the EESC basis function at their creation correctly. Therefore we decided to use a piecewise linear trend approach instead. We emphasize that our analysis is not intended to be a trend study, rather it is just a reality-check for the trends contained in the data sets when compared to trends derived from measurement time series. As such, we don't think it is necessary to show the post-1996 part of the fit,, or add a discussion about the problems of a piece-wise linear trend approach in comparison with an EESC approach to the manuscript. All three data sets are constructs and not measurements, and we only want to give the reader an idea of how the ozone decrease in these data sets compare to observed decreases.

This all said, our explanation for analyzing the trends of the data sets might seem confusing. We therefore went through the manuscript again (especially the Introduction and Section 6) to rephrase or expand explanations where necessary.

p. 26564, lines 17-19: I find this statement odd. MLR does indeed specify the time evolution of different components of the signal, but it does not specify their amplitude, so the overall trend is not imposed, as you suggest. Note that this statement contradicts the recent past practice of the WMO Ozone Assessment, which expresses trends in terms of EESC, so you are basically saying, without any evidence to back it up, that what they have done is wrong ("should not be used" is very strong language). In fact, aren't you also saying that the detection-attribution approach used heavily in IPCC is also wrong? Frankly I don't see anything wrong with asking what component of past behavior is attributable to a particular forcing. The key in any statistical analysis is to also look at the behavior of the residuals, which confirms the goodness of the fit. If, for example, the actual dependence on EESC is nonlinear, then this should become clear by looking at the residuals.

What we wanted to say in the sentence on p. 26564, lines 17-19 ("Although all three data sets are based on regression model output, and should therefore not be used for trend studies, as trends are effectively imposed by the fitting techniques used."), is that if somebody wants to do trend studies they should use measurements (from satellite or ground-based instruments) rather than using a regression model output to do so. Because the trend that is contained in the data set that is created by a regression model will not necessarily describe exactly the trend that is present in the original measurements, but rather the fit of the trend (or EESC) basis function to the original measurements. We understand that we might not have been clear enough in our expression to capture the intended meaning of our thoughts and therefore rephrased the respective sentence in our manuscript to 'The purpose of this is to assess the extent to which the data sets capture the observed ozone decreases, although we emphasize that the data sets are not suitable for assessing the true ozone trend as they are all based on regression model output and any trends are effectively imposed by the fit to a linear trend or EESC basis function in the regression model.'

Interactive comment on Atmos. Chem. Phys. Discuss., 12, 26561, 2012.