Review of "Stratospheric impact on tropospheric ozone variability and trends: 1990–2009" by P. G. Hess and R. Zbinden.

Jennifer A. Logan
School of Engineering and Applied Sciences
Harvard University.

This paper addresses the question of the influence of stratospheric/tropospheric exchange (STE) on tropospheric ozone variability with a combination of data analysis and modeling. This is certainly a topic of great interest, and worthy of detailed analysis using both observations and insights from model simulations. However, I have concerns about both types of analysis in this paper. I will address them separately first, and then discuss the comparison of the two.

Data analysis

The data analysis is based on correlations of ozone at 150 hPa, in the lowermost stratosphere (LMS), and at 500 hPa. Previous studies have used such an approach, but did the correlations differently. Tarasick et al. (2005) used annual averages and found correlations of r=0.66 for mean Canadian time series for both Arctic stations and for those at 53-59 N; Ordonez et al. (2007) found correlations for high altitude sites in Europe and sonde data at 150 hPa with r=0.77 for 12 month running means; more importantly they showed that the correlations were largest in winter and spring (r=0.58-0.78) and were absent or not significant in summer and autumn. Terao et al. (2008) used 3 month running means of anomalies and found significant correlations for some Canadian and European sonde sites, and showed that the correlations were driven by the behavior in winter-spring. Given the strong seasonality in the last two studies cited by the authors, I do not understand why the present analysis relies solely on 12 month running means of monthly anomalies. It would be more interesting to examine the seasonality of the correlations, and see if it varies from region to region, given that some are more remote from anthropogenic emissions than others. We know that there is seasonality in STE and in its influence on tropospheric ozone, and this would be expected to manifest itself in the correlations.

Ozone at 150 hPa is highly variable in winter and spring (the seasonal maximum), with much less variability in summer and autumn when ozone is much lower. Thus the interannual variability (IAV) in the 12 month running average anomalies is dominated by IAV in the seasonal maximum in spring. This is less of an issue at 500 hPa, when ozone is highest in summer, and the variability is not so dominated by one season. I expect that the correlations in this paper are driven mainly by stratospheric ozone in spring. I recommend that the authors examine the seasonality of the correlation by region.

The overall time series at 150 hPa is formed from a mean for Canada (mean latitude 67 N), N. Europe (72 N), and central Europe (50 N), so it is for an overall mean latitude of 63 N. Given these relatively high latitudes, it is hardly a surprise that the three regions are highly correlated with each other. In the troposphere, a mean record is formed for Canada, N. Europe, and the eastern US (mean latitude ~40 N, but only ~12 years of data), so the mean latitude is 60 N or 70 N, depending if there is a robust data set for the eastern U.S. Clearly, the analysis of the overall record is for fairly high latitudes (~60 N). There is a huge difference in the behavior of ozone at 30 N and 60 N for both 150 hPa and for 500 hPa, when the monsoon season affects ozone at lower latitudes.

The authors find that the correlation of their mean records at 150 and 500 hPa are highly significant, with $r^2 = 0.7$, and lower for the individual regions (r^2=0.42-0.56) which can only be Canada and N. Europe (p22738, l8-12). (I hope this is what is meant, as on p. 22739 l. 6, it is said that the correlation is 0.4-0.5, and correlation means r, while explained variance is r^2; clarify). A key question is the spatial extent to which this high correlation applies. As noted above, the mean latitude of the composite time series at 500 hPa is 60 N, with the eastern US data (40 N) showing similarities to the higher latitude time series in 1995-2002, but not thereafter. If this strong correlation for the mean time series applies for 50-90 N, it covers the northern 25% of the hemisphere, but it cannot be assumed to apply to 30-50 N on the basis of the results shown here.

Model

The CTM in this study, CAM-chem, is driven by NCEP meteorology. It has a complete treatment of the troposphere in terms of emissions and chemistry, and keeps emissions the same each year. The model does not simulate stratospheric chemistry. Rather than using the SYNOZ approach, which assumes a constant ozone flux from the stratosphere each year, the authors say they specify the concentration of the "Synoz tracer" in the tropical stratosphere (details not given). It is far from clear how this works, or exactly how it was done. Since the key model results depend crucially on this aspect of the model (STE), the authors must evaluate their stratospheric tracer by comparison with observations. They must show time series of Synoz at 150 hPa, and also evaluate its vertical profile in the lower stratosphere over the sonde sites they use. The model study presented here has little validity if they do not find the model results for the LMS to be reasonable. These comparisons should be done on monthly mean ozone time series or anomalies (3 month smoothing is fine), not on 12 month smoothed anomaly time series. Their approach cannot be expected to capture the post-Pinatubo behavior, but the model should capture dynamical variability in the LMS in 1990-1991 and after 1995. If it doesn't, it corrupts the model results in the troposphere.

The model results are scattered through the paper. In Section 3, they focus on average ozone north of 30 N; ozone for 50-90 N would make more sense, in the context of the data used. A method of tagging NOx (and hence ozone produced in the troposphere) is used to back out the stratospheric contribution to tropospheric ozone, and unfortunately, the model results are only shown as the 12 month running mean anomalies. The almost monotonic increase in ozone from 1990 to 1999 (Figure 1) suggests that the model has some issues with initialization, and certainly, the model does not show the relatively high values in 1990-91 seen in the observations. The authors need to explain why the ozone produced from tropospheric chemistry (deduced from NOx tagging) has the time dependence shown in Figure 1 (an increase until 1998, then a decrease), even though emissions are constant. It is hard to know how well this model simulates extra-tropical ozone when the only comparisons shown (in Section 5) are with 12 month running mean anomalies. Does the model get the amplitude of the seasonal cycle right at 500 hPa? Many models do not.

The discussion of the tests of the “Synoz” tracer (which isn’t really SYNOZ in the McLinden et al. definition) is confusing. It is well known that the age of air in the lower stratosphere is ~4 years, so how can the model equilibrate at ~150 hPa in a year or two, unless its circulation is too fast? The circulation in a CTM with NCEP winds will not necessarily be the same as that in the parent GCM.

The statement is made in the Conclusions (p. 22746, l10-12) that “Stratospheric ozone is parameterized in these simulations assuming no interannual variability in the stratospheric ozone concentrations.” Is this really true? Surely it is true only in the source region (30 N-30 S)? Clarify.

Model/data comparisons

In Section 5 the model and data are compared only in terms of the AAMD time series, and only for 1996 onwards for correlations. The model does badly prior to 1996. Comparisons to data (absolute values) are made only in Table 2. Biases are generally small, but the reader should be shown whether the model matches the observed seasonal cycle.

Figure 7 that compares the time series (as AAMD) is so small it is hard to discern details. The panels should be expanded lengthwise (say 4 stacked panels, page width).

Given my concerns about the model treatment of STE, I do not have much confidence in the fraction of the variability in the troposphere attributed to the stratosphere. For the data, the fraction is appropriate for highly averaged and highly smoothed mid-high latitude data as noted above.

The model is also compared to 4 surface sites that vary considerably (Figure 5 and 8), so it would make more sense to show them individually in Figure 8.

Figure 7 shows that the model does best at capturing the behavior of ozone from 1997 to 2000, namely the relatively large increase from 1997 to 1999 followed by lower ozone in 2000 in Canada, northern Europe and the eastern U.S. Since this is the largest signal in IAV from 1996 to 2008, it likely that this period drives most of the correlation between the model and data. It is certainly of great interest to know what caused this signal in the atmosphere, which is apparent from Mace Head to the high Arctic sites, so I would encourage the authors to explore this further, including a seasonal analysis.

I think the authors have quite a bit of work to do before this paper is ready for publication, but I urge the authors to do it. They must address the reviewers' concerns (including mine) about their treatment of "Synoz".

Other comments:

General:

The paper is written so it jumps between data analysis and model results in a confusing way. It would make more sense to present all the data analysis first, then the model results and evaluation with data.

The nomenclature "annual average monthly deviations" (AAMD) is confusing, as there are not 12 annual averages in a year.

Using the AAMD, instead of say the real annual average, increases "n" in the correlation analysis by a factor of 12. It will thus increase the significance of any correlation present in the annual means.

In terms of presentation, it would make more sense to show and discuss Figure 4-5 before Figure 3.

What are the units on Figures 4-8? ppb? Figure 1 is in standard deviations, most others are not labeled.

Keep a consistent color for the model in all plots.

I am not repeating comments by other reviewers on correct names of stations etc.

22725, 13. This study used only BM data at the central European sites. This means that it does not use the generally more reliable ECC data at Payerne. Jeannet et al. (2007) showed that the Payerne BM data are unreliable in the early 1990s. I recently submitted a paper to J. Geophys. Res. (September,2011, copy sent to Hess) showing problem with the Payerne and Hohenpeissenberg BM data in the mid-1990s (and with the latter in the early 1990s), and showing that the data after 1998 are more reliable, based on comparisons with high altitude sites and MOZAIC data. This work was presented at the workshop in early April 2011 in Toulouse, "Tropospheric ozone changes: Observations, state of understanding, and model performances", attended by Zbinden and Hess.

22726, 3-10. Give the range of MOZAIC profiles for all 3 clusters, and note that the data are extremely sparse (often less than a few days per month) in 2005-2007 for the eastern U.S. and for some periods for Japan.

22726, 22. Some analyses are done with normalized AAMD, in which the time series are normalized by "the standard deviation of the time series". Do you mean the standard deviation of the 12 month running mean anomalies, or that of the actual monthly mean time series? Clarify.

22726, 13-14. Do you mean that the AAMD for the sites have to be correlated with each other? This section should refer to the Tables S1 and S2, and make clear if all pairs have to be significantly correlated with each other in a cluster.

22732. Discussion of Japanese sites. It is not surprising that these sites are so different from each other, they span from mid-latitudes to sub-tropics, while all other sites are in mid to high latitudes.

22733, 19. Be specific. The increase is clearly after 1993, after the post-Pinatubo minimum in the winter of 1992/93.

22734, 19. The pronounced dip is in 1992-94, not 1991-95.

22734. Why is the Boulder record not used? It is available from the NOAA/ESRL web-site.

22735, 1: Utilized is just a long word for used in this paper. Please change "utilize" to "use" throughout the paper.

22735-22726. See my comments above on problems with the European sonde data in the early 1990s. Also, note that the Zugspitze data seem more reliable in the early 1990s than the Jungfraujoch data.

22739, l4-6. Does this refer to the model? to the correlation among regions at 500 hPa? to the correlation between 150 and 500 hPa? r=0.4-0.5 means r^2 = 0.2-0.25, pretty small.

22740, l 5-6. Explain why you omit 1991-1995, and why you think there was a large effect on ozone at 500 hPa. The effect of Pinatubo in the stratosphere was first evident in extra-tropical ozone in the stratosphere in the winter of 1992-93, according to the WMO assessments. Figure 7 shows the model does not do well in 1990-1992, so including these years would reduce the correlations.

22740, l. 12-13. Of course the low values in 1992-1009 will impact trends that start in 1990 and are for periods as short as 10 and 20 years.

Table 2. What does standard deviation refer to in this table? Does “Annually averaged ozone” have one value per year or 12? Similarly, is the correlation based on one value per year or 12? There cannot be trends for the eastern US starting in 1990, as you show only data from 1995.