Agreement in late twentieth century Southern Hemisphere stratospheric temperature trends in observations and CCMVal-2, CMIP3, and CMIP5 models

Paul J. Young,1,2,3 Amy H. Butler,4 Natalia Calvo,5,6,7 Leopold Haimberger,8 Paul J. Kushner,9 Daniel R. Marsh,6 William J. Randel,6 and Karen H. Rosenlof2

Received 18 June 2012; revised 14 November 2012; accepted 18 December 2012.

[1] We present a comparison of temperature trends using different satellite and radiosonde observations and climate (GCM) and chemistry-climate model (CCM) outputs, focusing on the role of photochemical ozone depletion in the Antarctic lower stratosphere during the second half of the twentieth century. Ozone-induced stratospheric cooling peaks during November at an altitude of approximately 100 hPa in radiosonde observations, with 1969 to 1998 trends in the range of −3.8 to −4.7 K/dec. This stratospheric cooling trend is more than 50% greater than the previously estimated value of −2.4 K/dec, which suggested that the CCMs were overestimating the stratospheric cooling, and that the less complex GCMs forced by prescribed ozone were matching observations better. Corresponding ensemble mean model trends are −3.8 K/dec for the CCMs, −3.5 K/dec for the CMIP5 GCMs, and −2.7 K/dec for the CMIP3 GCMs. Accounting for various sources of uncertainty—including sampling uncertainty, measurement error, model spread, and trend confidence intervals—observations and CCM and GCM ensembles are consistent in this new analysis. This consistency does not apply to each individual that makes up the GCM and CCM ensembles, and some do not show significant ozone-induced cooling. Nonetheless, analysis of the joint ozone and temperature trends in the CCMs suggests that the modeled cooling/ozone-depletion relationship is within the range of observations. Overall, this study emphasizes the need to use a wide range of observations for model validation as well as sufficient accounting of uncertainty in both models and measurements.


1. Introduction

[2] Observations [e.g., Randel and Wu, 1999; Thompson and Solomon, 2002] and models [Gillett et al., 2011; Polvani et al., 2011] suggest that the strong decreasing trend in late twentieth century springtime Antarctic stratospheric ozone [Solomon, 1999] is responsible for most of the colocated and contemporaneous cooling trend [see also Forster et al., 2011]. Correctly modeling the stratospheric response to ozone depletion is essential to understanding the magnitude of its effects, as well as to probe the processes that drive these effects. The influence of ozone depletion is felt far beyond the Antarctic stratosphere, and is likely apparent in modulations of the global stratospheric circulation [Garny et al., 2009; Mclandress and Shepherd, 2009], as well as in changes in the Southern Hemisphere (SH) troposphere [e.g., Gillett and Thompson, 2003; Son et al., 2010; Thompson et al., 2011], perhaps even extending to the subtropics [Kang et al., 2011].

[3] Investigating the climate role of ozone changes was a goal of the second Chemistry-Climate Model Validation (CCMVal-2) project [Eyring et al., 2008]. Analysis of the chemistry-climate model (CCM) temperature trends from CCMVal-2 [Baldwin et al., 2010] suggested that, on average, the modeled temperature trends associated with Antarctic ozone depletion were too strongly negative when compared...
with the radiosonde trends calculated by Thompson and Solomon [2002] (hereafter TS02). Moreover, the analysis found that an ensemble of climate models, from the World Climate Research Programme’s Coupled Model Intercomparison Project phase 3 (CMIP3) multimodel data set [Meehl et al., 2007], matched the same observed trend estimates better, despite their far more limited representation of the stratosphere and their exclusion of chemical processes important for ozone.

[4] Here, this temperature trend comparison is revisited, presenting an intercomparison of modeled Antarctic temperature trends and those derived from a variety of observational data sets beyond those presented by TS02. Although the focus is on trends from CCMVal-2, these are compared alongside the CMIP3 data set, as used throughout the Intergovernmental Panel on Climate Change Fourth Assessment Report [IPCC, 2007], as well as in the CMIP5 data set [Taylor et al., 2012], which will be used for the Intergovernmental Panel on Climate Change’s Fifth Assessment Report. Overall, the results suggest that there is broad consistency between the observed and the ensemble mean modeled ozone-related cooling, although the magnitude varies between different models [Austin et al., 2009] and between different observations [Randel et al., 2009], and is also sensitive to the period under consideration.

2. Model and Observational Data

[5] Table 1 summarizes the model, radiosonde, and satellite observational data used, including the acronym definitions. For the radiosondes, whereas IUK and HadAT2 are already provided as monthly mean data, monthly means for the RAOB CORE and RICH data sets were calculated from daily data (averaging both the 00Z and 12Z soundings for each station, for better temporal coverage), using only stations with less than 20% missing data at 100 hPa (for the period 1969–1998; same as TS02). Using this criterion means that 10 of the possible 18 station data were used. Both the RICH-obs and RICH-τ data sets have 32 ensemble members, derived by varying parameters in the adjustment process [see Haimberger et al., 2012]. Here, only the means of the 32 member ensembles are considered, although we note that there is good agreement between the polar cap mean (>65°S) 1969 to 1998 November 100 hPa trends calculated using the individual ensemble members, with variability at less than 0.05 K/dec (1 standard deviation).

[6] Figure 1 shows the location of the stations for each radiosonde data set, highlighting in red those stations that were included in the analysis of TS02 (the trends at the stations are discussed in section 3). As well as differing in the period covered and number of stations south of 65°S, each radiosonde data set uses different methods to adjust the temperature time series in order to account for any artificial shifts, such as from instrument or procedural changes. Furthermore, both the HadAT2 and IUK data stop at 30 hPa, whereas the RICH and RAOB CORE data extend to 10 hPa. Free [2011] showed the differences between many of the same data

![Figure 1. Locations of the radiosonde stations from the different data sets used in this study, colored by their November 100 hPa trend over 1969 to 1998 (K/dec). Symbols with black outlines indicate that the trend is significant at the 5% level. Stations in red are those analyzed by Thompson and Solomon [2002].](image)

<table>
<thead>
<tr>
<th>Data</th>
<th>Abbreviations</th>
<th>Stationsa</th>
<th>Period coveredb</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hadley Centre Atmospheric Temperatures, ver. 2</td>
<td>HadAT2</td>
<td>7</td>
<td>1969–2010</td>
<td>Thorne et al. [2005]</td>
</tr>
<tr>
<td>Iterative Universal Kriging</td>
<td>IUK</td>
<td>5</td>
<td>1969–2005</td>
<td>Sherwood et al. [2008]</td>
</tr>
<tr>
<td>Radiosonde Observation Correction using Reanalysis, ver. 1.5</td>
<td>RAOB CORE</td>
<td>11</td>
<td>1969–2010</td>
<td>Haimberger et al. [2008]</td>
</tr>
<tr>
<td>Radiosonde Innovation Composite Homogenization (obs), ver. 1.5</td>
<td>RICH-obs</td>
<td>11</td>
<td>1969–2010</td>
<td>Haimberger et al. [2012]</td>
</tr>
<tr>
<td>Radiosonde Innovation Composite Homogenization (τ), ver. 1.5</td>
<td>RICH-τ</td>
<td>11</td>
<td>1969–2010</td>
<td>Haimberger et al. [2012]</td>
</tr>
<tr>
<td>Raw radiosonde data</td>
<td>Raw</td>
<td>11</td>
<td>1969–2010</td>
<td>IGRA¹, and ERA–40 and ERA-Interim observation input [Dee et al., 2011]</td>
</tr>
<tr>
<td>MSU lower stratosphere temperatures, Remote Systems Sensing, ver. 3.3</td>
<td>MSU TLS</td>
<td>1979–2010</td>
<td>Mears and Wentz [2009]</td>
<td></td>
</tr>
<tr>
<td>CMIP3 historical (20C3M) simulation model output</td>
<td>CMIP3</td>
<td>1969–1999</td>
<td>Meehl et al. [2007]</td>
<td></td>
</tr>
<tr>
<td>CMIP5 historical simulation model output</td>
<td>CMIP5</td>
<td>1969–2005</td>
<td>Taylor et al. [2012]</td>
<td></td>
</tr>
</tbody>
</table>

¹For radiosonde data, the number of stations south of 65°S.
²The start date refers to first year used in this study. For CMIPVal-2, CMIP3, and CMIP5, the end date is the year for which all models still have output.
³Integrated Global Radiosonde Archive (http://www.ncdc.noaa.gov/oa/climate/igra/).
sets for tropical temperature trends. Further details concerning the radiosonde data sets and stations can be found in the Supplementary Material (Table S1).

[7] Lower stratospheric satellite brightness temperature data from the Microwave Sounding Unit (MSU TLS) also form part of the analysis. The weighting function for the MSU TLS data covers a broad vertical layer centered on approximately 80 hPa, with the half-power width extending from 150 to 50 hPa [see Randel et al., 2009, their Figure 1]. This weighting function was applied to the radiosonde and model data for part of the analysis. The MSU TLS data provide complete zonal coverage, but there are no data poleward of 82.5°S.

[8] Monthly zonal mean temperature and ozone data from the CCMVal-2 models were collected from the REF-B1 experiment, which was configured to reproduce the composition of the atmosphere from 1960 to 2006 [Morgenstern et al., 2010]. Ozone output was not available for the AMTRAC3 and UMETRAC models. CMIP3 monthly zonal mean temperature data were collected from the 20C3M experiment, which also aimed to reproduce past climate (from the preindustrial period to the year 2000) [Meehl et al., 2007]. For CMIP3, our study leaves out CNRM-CM3 and UKMO-HadCM3 because these models have incomplete temperature data (missing data in the lower stratosphere), as well as UKMO-HadCM3 having prescribed ozone trends twice as large as observed [Karpechko et al., 2008]. GISS-EH and GISS-ER had an erroneously low ozone forcing [Miller et al., 2006], although these models are still included in this analysis (see Figure S1). The CMIP3 models were subdivided into those that included time-varying prescribed ozone data (CMIP3 w/o ozone) and those that did not (CMIP3 no ozone), as has been done previously [e.g., Cordero and Forster, 2006; Cai and Cowan, 2007; Karpechko et al., 2008]. Monthly zonal mean temperatures from the CMIP5 historical experiment, covering the preindustrial period to 2005 [see Taylor et al., 2012 and references therein], were also considered. All these models used some form of time-varying ozone, either calculating concentrations interactively (in the manner of a CCM) or by using a prescribed data set (the data set developed by [Pollard et al., 2011] was recommended). Eyring et al. [Long-term changes in tropospheric and stratospheric ozone and associated climate effects in CMIP5 simulations, submitted to Journal of Geophysical Research–Atmospheres, 2012] describe further details related to the ozone concentrations in the CMIP5 models. More information on the individual models from the CMIP3, CMIP5, and CCMVal-2 data sets can be found in the Supplementary Material (Tables S2–S4).

[9] For the three model data sets, in which there was more than one realization available for a given model, the intramodel ensemble mean was first determined before calculating the overall data set ensemble mean (i.e., each model was weighted equally).

[10] Trends are mainly considered over the 30 year period (1969–1998), for consistency with TS02 and Baldwin et al. [2010], and over the 21 year period (1979–1999), the period that all observations and models have in common (Table 1). Observations show zonal asymmetries in SH high-latitude temperature trends for certain months [Hu and Fu, 2009; Lin et al., 2009], which likely depend on trends and variability in wave driving, not generally captured by climate models [Wang and Waugh, 2012]. As such, we only consider zonal mean trends in the models and data (albeit a sparsely sampled zonal mean for the radiosonde data). By not sampling at the radiosonde locations, this method could bias the sampling of the models to the colder deep vortex. However, we demonstrate below that the “zonal mean” radiosonde trends agree well with those calculated from the MSU TLS data, which has near full coverage of the polar cap. Trends were calculated by linear least squares regression on data binned according to month (or season), with the statistical error estimates adjusted to account for serial autocorrelation [according to Santer et al., 2000]. Unless stated, the quoted errors encompass the 95% confidence interval.

3. SH Temperature Trends Results

[11] Figure 2 shows the high-latitude (>65°S) mean temperature trends for the radiosondes, the CMIP3 w/ozone ensemble mean, the CMIP5 ensemble mean, and the CCMVal-2 ensemble mean for the period 1969 to 1998 as a function of pressure and month (similar to TS02, their Figure 1). Trends for the RICH-τ and RAOB CORE data (not shown) are very similar to RICH-obs, and arise from those data sets using the same stations and the same break detection algorithms (see Table S1 in the Supplementary Material for more information).

[12] Both the radiosonde and model data show a strong and significant cooling in the lower stratosphere, extending from approximately 200 to 50 hPa, and from at least October to December, as also reported by TS02. The maximum cooling trend occurs for November at approximately 100 hPa, although the trends differ both in magnitude and (less so) in spatiotemporal patterns. Comparing the radiosonde data shown in the top row of Figure 2, the maximum cooling trend is found in the IUK data set (−4.7 ± 2.8 K/dec), followed by the RICH-obs data set (−4.1 ± 2.4 K/dec), with the HadAT2 data showing the weakest cooling (−3.8 ± 2.4 K/dec). These are all stronger values than the −2.4 K/dec (−7.1 K/30 a) trend reported by TS02, but more comparable with the −3.8 K/dec peak value determined from radiosonde data by Thompson and Solomon [2005], although this covers a different period (1979–2003). Like TS02 and Thompson and Solomon [2005], both the IUK and RICH-obs data suggest that the significant lower stratospheric cooling trend persists into March, whereas the trend stops in December with the HadAT2 data.

[13] Why are the trends different between the radiosonde data sets? Figure 1 shows that the SH mean temperature for each data set is composed of different stations, covering different longitude and latitude ranges. At the lower latitudes, the stations may be occasionally sampling air outside of the cold vortex [Hassler et al., 2011a], which weakens the trends derived at these locations. For example, November 100 hPa trends at the South Pole station (90°S) are in the range of −6.0 to −7.1 K/dec, whereas they are between −2.3 and −2.5 K/dec for Casey (66°S; see Table S1). Temperature trends at the continent edge may also be affected by the zonally asymmetric nature of the trend patterns [e.g., Lin et al., 2009], although the degree of asymmetry depends on the month. Hence, part of the reason for the stronger cooling seen with the IUK data is that its mean
is weighted toward higher-latitude stations. However, the RICH-obs, RICH-t, and RAOBCORE data sets include more stations toward the edge of Antarctica compared to HadAT2, yet they still have stronger cooling trends, suggesting that the spatial distribution of stations cannot explain all the differences between data sets.

Figure 1 also indicates a range of values for the trend at a given station, depending on the data set (see also Table S1). In general, the HadAT2 data show the weakest cooling for a given station. This is particularly the case for McMurdo, Novolazarevsk, and Mawson stations, where the HadAT2 data is a factor of 1.5 to 1.7 lower than the maximum cooling trend. Restricting the RICH-obs data to just the HadAT2 stations (also including SANAE, which does not meet the <20% missing data requirement) results in a trend of $-4.7 \pm 2.2$ K/dec, i.e., more cooling than that calculated from the HadAT2 data. Restricting the RICH-obs data to the IUK locations results in a trend of $-4.6 \pm 2.6$ K/dec, i.e., slightly less cooling than that for the IUK data.

Figure 2 shows that the peak cooling trends for the CMIP5 and CCMVal-2 model ensembles are comparable to the radiosonde data, with November 100 hPa trend values of $-3.5 \pm 0.3$ K/dec and $-3.8 \pm 0.7$ K/dec, respectively. The trends are weaker for the CMIP3 w/ozone ensemble, in which the November 100 hPa trend is $-2.7 \pm 0.3$ K/dec. The smaller confidence interval in these trends is due to the substantial reduction in interannual variability from averaging several models together, and it is not comparable to the confidence intervals for the observed trends discussed above. Based on the data from TS02, Baldwin et al. [2010] concluded that the CMIP3 w/ozone models had a more favorable comparison with the observed trends than the CCMVal-2 models. However, the broader range of radiosonde trends in Figure 2 suggests that the CCMVal-2 and CMIP5 ensemble means compare more favorably with observations than CMIP3 w/ozone.

In addition to the ozone-induced cooling, the RICH-obs and IUK data also show a higher altitude warming trend, occurring after the ozone cooling in IUK and at approximately the same time as the ozone cooling with RICH-obs. A warming trend similar in magnitude and timing to that of IUK is apparent in the CCMVal-2 ensemble mean trend, and (more weakly) in the CMIP5 ensemble mean trend. From the individual models, a warming trend is found in more than half of the CMIP5 models (Figure S2) and in 16 of the 17 CCMVal-2 models (Figure S3), although it is not always significant for either set of models. Manzini et al. [2003] described a similar feature in their CCM study, attributing it to increased downwelling (and compressional warming) due to enhanced gravity wave propagation, itself due to the ozone-induced cooling. Thus, the presence of such a trend could be an indicator of how well models perform in terms of middle atmosphere dynamics, although further study is required.

Figure 3 explores the modeled and observed polar cap ($>65^\circ$S) temperatures in more detail. Figure 3a shows the time series of November temperature anomalies from 1969 to 2010 (or the latest date for the given data; see Table 1). Figure 3b shows the time series of October to January (ONDJ) averaged anomalies for the MSU TLS data, MSU
TLS-weighted radiosondes, and MSU TLS-weighted models. The ONDJ mean is used here because this corresponds to the months in which Baldwin et al. [2010] found that the CCMVal-2 models agreed better with the observed trends of TS02. Notwithstanding that the TS02 trends are smaller in magnitude than those from the other radiosonde data sets, including summer months to calculate the mean trend goes some way to counter model biases in the timing of the SH vortex breakup [Hurwitz et al., 2010; Butchart et al., 2011]. Anomalies are computed relative to the 1979 to 1999 climatology. Note that using anomalies removes systematic biases, identified as a particular issue for SH spring temperatures in the CCMVal-2 models [Butchart et al., 2011].

For the individual CCMVal-2 models and observations, both sets of time series show the large year-to-year variability characteristic of springtime lower stratospheric temperatures [e.g., Young et al., 2011], although this is damped in Figure 3b by averaging over more months. Time series for the model ensemble means show far less variability, as the noise from individual models tends to cancel. The correlation between the anomalies for the different radiosonde data sets is very high ($r > 0.96$), despite the different stations and different adjustment methods. Furthermore, for Figure 3b, the correlation between the independent MSU TLS data and radiosondes is also very high ($r > 0.93$). As the MSU data set has complete zonal coverage (although stopping at 82.5°S), this suggests

Figure 3. (a) Time series of 100 hPa temperature anomalies for November, for the different radiosonde data sets and the unadjusted radiosondes at the Thompson and Solomon [2002] locations (red), the CMIP3 “w/ozone” and “no ozone” ensemble means (green), the CMIP5 ensemble mean (purple), and the CCMVal-2 models and ensemble mean (blue). (b) Similar to Figure 3a, but time series of MSU TLS-weighted temperature anomalies for the October to January average and including MSU TLS data (black). (c) Trends for the anomalies in Figure 3a over the period 1969 to 1998 (K/dec). (d) Trends for the anomalies in Figure 3b over the period 1979 to 1999 (K/dec). For Figures 3c and 3d, the error bars indicate the 95% confidence interval for the trends, except for the CMIP3, CMIP5, and CCMVal-2 ensemble means, in which it indicates the range of the trends from the individual models. The black cross on the “TS02 locations (raw)” trend in Figure 3c indicates the trend reported by Thompson and Solomon [2002].
that the radiosonde data sets are representative of the polar cap temperature, despite their more limited spatial coverage.

[19] Figure 3c shows the 1969 to 1998 trends for the time series in Figure 3a, including the trends calculated for the individual CCMVal-2 models as well as the radiosonde and model ensemble mean trends discussed above. The error bars encompass the 95% confidence interval for the trends for all cases except the model ensembles. Here, due to the aforementioned low variability in the ensemble mean time series, the error bars indicate the range of the trends found from the individual models that comprise the ensemble mean. For CCMVal-2 models with more than one ensemble member (CCSRNIES, CMAM, LMDZrepro, SOCOL, and WACCM), the data shown are just from the first simulation (run 1) so as not to dampen the contribution of interannual variability to the trend uncertainty. The trends calculated for the radiosondes and the CMIP3 w/ozone, CMIP5, and CCMVal-2 ensembles all agree with each other within their uncertainty ranges. The figure also shows the trend calculated using unadjusted (raw) radiosonde data, for the same stations used by TS02. As seen with the adjusted radiosonde data sets, this trend reflects a stronger cooling than that calculated by TS02 (indicated by the black X), although, again, all are within each other’s uncertainty.

[20] Figure 3d is similar to Figure 3c, but shows the 1979 to 1999 trends for the time series in Figure 3b, which is the longest period common to all model and observational data sets. Again, the trends for the observations and CMIP3 w/ozone, CMIP5, and CCMVal-2 ensembles all agree within their statistical uncertainty. The close agreement of the MSU TLS and radiosonde trends further underlines the representativeness of the radiosonde data for the SH high latitudes, although the coverage of the MSU TLS data, which does not extend to the pole, could be affecting the strength of the cooling.

[21] Figures 3c and 3d show that trends for the individual CCMVal-2 models cover a wide range of values, wider than that for the other ensembles. Furthermore, although all the trends are negative, they are not all significant at the 5% level. For many CCMVal-2 models, the error bars for the trends are greater than those for the observations, suggesting a larger interannual variability in these models compared to the observations, and a topic for further study. The range of trends from these models is discussed further in section 4. That there is a spread of trend estimates is not unique to the CCMVal-2 data set: figures in the Supplementary Material show individual model trends from 1969 to 1998 as a function of month and pressure for the CMIP3, CMIP5, and CCVal-2 models, emphasizing the model diversity in this regard.

[22] Figure 3 also includes the time series and trends from the CMIP3 no ozone ensemble, which is markedly weaker than the other ensemble mean trends and the observations. (Note that the “CMIP3 no ozone” trend in Figure 3d is not significant; the error bar only indicates that all the models in this ensemble produce a negative trend.) The absence of any significant negative trend for this set of models suggests that cooling from CO2 increases cannot explain the observed trends, confirming the role of ozone depletion as the dominant driver of the SH lower stratospheric cooling at the end of the twentieth century [e.g., Shine et al., 2003; Karpechko et al., 2008].

[23] We also note that the magnitude of the trends depends on the period over which they are computed. Trends calculated from 1969 are generally lower in magnitude (i.e., less cooling) when the end year is later in the record. For example, using the RICH-obv November 100 hPa data, trends are in the range of $-3.7$ to $-4.3$ K/dec with an end year from 1998 to 2002, and $-3.2$ to $-3.6$ K/dec when the end year is from 2003 to 2010. A similar behavior is evident from the CCMVal-2 and CMIP5 ensemble mean time series, in which there is a monotonic decrease in the magnitude of the cooling trend when the end year is after 1998. All of these trends are still significant at the 5% level.

[24] Visual inspection of the observed time series in Figure 3 suggests a flattening out of the trend toward the end of the record, although a longer time series would be needed to diagnose a statistically significant change point. Although there might be some expectation of a reduction in the cooling trend since we have now passed the peak CFC concentration [Newman et al., 2007] and ozone loss rates have been reduced [Hassler et al., 2011b], there is large year-to-year variability in the stratospheric circulation, and hence, temperatures. For example, the thus far unique SH sudden stratospheric warming in 2002 [e.g., Newman and Nash, 2005] is a notable anomaly in the record that can affect trend assessments.

4. The Relationship of Ozone and Temperature Trends With CCMVal-2 Models

[25] As well as model-observation comparison for the magnitude of the trend, Baldwin et al., [2010] considered how the austral spring temperature and ozone trends were related to one another, in the individual models and in the observations. They reported (their Figure 10.13) that CCMs qualitatively reproduced the observed correlation between weaker (stronger) ozone depletion and weaker (stronger) 100 hPa cooling trends. But they also suggested, based on TS02 and Halley station total ozone column observational data, that the models overestimated the cooling for a given ozone loss. We revisit their analysis in this section, using the expanded set of observations from the radiosonde data sets as well as an expanded set of ozone column data.

[26] Figure 4 shows the relationship between the September to December ozone column trend and ONDJ temperature trends for the individual CCMVal-2 models, the CCMVal-2 ensemble mean, and observations. Figure 4a shows the relationship for the 1969 to 1998 trends, using the 100 hPa temperature [according to Baldwin et al., 2010], and Figure 4b shows the relationship for the 1979 to 1999 trends, using MSU TLS and MSU TLS-weighted temperature data. The MSU TLS data are useful for this case because there is substantial overlap between the weighting function and the vertical region of greatest Antarctic ozone depletion. Observed ozone column trends are from the mean of the four Antarctic ozone stations with the longest records: Faraday/Vernadsky (65° S, 64° W), Syowa (69° S, 40° E), Halley (76° S, 27° W), and South Pole (90° S, 25° W) [according to Hassler et al., 2011a].

[27] Error bars on the observations show the range of the trend values for the different temperature data sets (horizontal bars) and different ozonesonde stations (vertical bars). All the observed trends are significant at the 5% level. Error bars on
the CCMVal-2 ensemble mean trend overlap in both panels of Figure 4. This further strengthens the case made in the previous section, that the CCMVal-2 ensemble mean matches the observations within the uncertainty for both ozone and temperature.

[28] Individual CCMVal-2 models show a large scatter in their ozone and temperature trends, as also shown in Figure 3. However, the dashed line in each panel of Figure 4 indicates the positive (and significant) relationship between the modeled trends in lower stratospheric temperatures and ozone column. Note that the regression is forced through the origin, i.e., assuming zero temperature trend for zero ozone loss. If the regression is not forced through zero, the resulting intercept suggests a positive temperature trend for ozone loss of less than 20 DU/dec, which may not be physical. The regression coefficients for each panel are very similar. The regression coefficient for Figure 4a is 12.0±1.8 DU dec⁻¹/K dec⁻¹, and for Figure 4b it is 12.1±2.3 DU dec⁻¹/K dec⁻¹; i.e., an ozone trend of 12 DU/dec is predicted to be accompanied by a temperature trend of 1 K/dec.

[29] The gray shaded area in Figure 4 shows the 95% confidence interval for the linear regression, estimated from the variance of the regression residuals [e.g., see Wilks, 2006]. The area estimates the uncertainty in the relationship between the trends in ozone column and lower stratospheric temperature, as derived from the models; i.e., if we have a certain trend in the ozone column, the gray area indicates where we are 95% certain that the corresponding temperature trend will lie (or vice versa). Although the observed trends in Figure 4 sit below the regression line, and therefore suggest a smaller temperature trend for a given ozone change, they are within the shaded area. This would suggest that, at the 5% level, the CCMVal-2 models do not produce a stronger cooling for a given ozone depletion compared to observations. Note that the observations still lie within the 95% confidence interval if the regression is not forced through the origin. (Interestingly, the 95% confidence interval does not include the origin if the regression for Figure 4a is not forced through it.)

[30] Finally, Figure 4a also indicates the temperature trend calculated by TS02. For ONDJ, their estimated trend falls within the range of those estimated from the other data sets, but only due to the inclusion of the HadAT2 data, which has a weaker cooling trend than the other radiosonde data sets. From Figure 2, we see that the HadAT2 cooling trend does not continue through the austral summer, as it does for the IUK and RICH-obs data. Restricting the RICH-obs data to the HadAT2 stations does not change the month-pressure trend pattern markedly from the full RICH-obs data in Figure 2a, suggesting that the more limited cooling in the HadAT2 data is related to the adjustment method, rather than the spatial distribution of the stations.

5. Summary and Conclusions

[31] We have presented late twentieth century, high-latitude SH stratospheric temperature trends from several radiosonde data sets, satellite measurements, and multimodel ensemble data from climate and CCMs. Except for models that do not include ozone depletion, all the trends show strong cooling during austral spring and summer, peaking in November at approximately 100 hPa. Although the observed cooling is dependent on the data set, the magnitudes of the trends
calculated here are more than 50% stronger than those presented by **Thompson and Solomon** [2002], whose work is often used as a benchmark for modeling studies. However, once the statistical uncertainties are taken into account, the trends from the multimodel ensembles and observations (including TS02) agree with one another. Overall, the results suggest that there is no systematic bias toward excessive cooling in the ensemble mean SH lower stratosphere temperatures determined from the latest generation of stratospheric CCMs. Furthermore, our results also suggest that ensemble mean temperature trends determined from climate models also compare well to observed trends, provided that the models include some representation of late twentieth century ozone depletion [**Cordero and Forster**, 2006].

Although trends calculated from ensemble mean temperatures match observed trends well, trends from the individual models show a large spread. Despite including the drivers of ozone depletion, some of the CCMVal-2 models do not have significant austral spring temperature trends. This range of model skill for stratospherically relevant parameters in the CCMs has been explored in several other studies [e.g., **Gettelman et al.**, 2010; **Heglin et al.**, 2010; **Butchart et al.**, 2011], and several shortcomings have been identified. In particular, **Butchart et al.** [2011] highlighted a pervasive poor performance of the models for the SH during austral spring, and several studies have indicated the large range of modeled ozone trends [**Eyring et al.**, 2006; **Austin et al.**, 2009; **Austin et al.**, 2010]. For the climate models, the spread of the trends is generally less, likely due to the prevalence of a prescribed ozone data set rather than the more complex online calculation of ozone in the CCMs. Nevertheless, for the CCMs, there is a significant positive linear relationship between the magnitude of the cooling trend and the magnitude of the ozone depletion, and the observed temperature and ozone trends seem to conform to the same relationship, within statistical uncertainty, i.e., our results suggest that the models do not systematically overestimate the cooling for a given ozone depletion.

Overall, we would recommend that multiple temperature data sets are used to understand the evolution of stratospheric temperature, both for observational studies and for model evaluation, echoing the similar sentiments of **Free** [2011], **Thorne et al.** [2010], and **Calvo et al.** [2012]. The latter study is especially pertinent here; as they evaluated SH late twentieth century temperature trends in a CCM using a range of data sets and found that the modeled cooling trend agreed better with data sets other than TS02.

**References**


YOUNG ET AL.: MODELED STRATOSPHERIC TEMPERATURE TRENDS


Kang, S. M., L. M. Polvani, J. C. Fyfe, and M. Sigmond (2011), Impact of


Kang, S. M., L. M. Polvani, J. C. Fyfe, and M. Sigmond (2011), Impact of


Kang, S. M., L. M. Polvani, J. C. Fyfe, and M. Sigmond (2011), Impact of