Reply to Reviewer #2

We would like to thank Reviewer #2 for the constructive and helpful comments. Reviewer's contribution is recognized in the acknowledgments of the revised manuscript. Our response to the reviewer's comments follows point by point.

1) The reviewer notes " *1. The results are based on the evaluation of temperature values that are computed from the thickness of atmospheric layers. It is argued that this method provide better temperature values than the temperatures themselves but little evidence is given to support this statement. A comparison of the presented temperatures with the initial temperatures of the used data sets should thus be provided. Also it would be interesting to have an idea of how the presented temperature anomalies compare to those of satellite data in the lower stratosphere (e.g. MSU channel 4 and SSU channel 1) in the 1980-2011 period. Indeed ,these datasets are widely used for the evaluation of recent temperature trends in the lower stratosphere.*"

Our argument for the preference to use thickness to obtain layer mean temperatures referred mainly to the case of the FUB dataset as it was pointed in our earlier studies based on this dataset (Zerefos and Crutzen, 1975; Zerefos and Mantis, 1977; Mantis and Zerefos, 1979). In the papers by Zerefos and Mantis (1977) as well as Mantis and Zerefos (1979) it was emphasized that the approximate geostrophic balance of the upper winds insures that the contour analysis will be more representative than the temperature analysis which for the FUB data it was based on map fields based on scattered radiosonde locations. Please note also that the FUB data used in this study were derived from daily stratospheric map analyses of isobaric surfaces prepared by the Stratospheric Research Group of the Free University of Berlin (http://www.geo.fu-berlin.de/en/met/ag/strat/produkte/fubdata/index.html). The FUB dataset consists of 35 years of daily (or bi-daily in summer) geopotential height and temperature fields at 50, 30 and 10hPa in the northern hemisphere.

The hemispheric analyses were produced in real time by a subjective analysis technique, using the 00UT radiosonde reports from the observational network, by a team of experienced meteorologists. Both geostrophic and hydrostatic balance were assumed in the analysis procedure, and the wind observations were given a high priority. These balance conditions ensures a consistent dataset; further, temporal continuity was assured by meteorological inspection. Note that these balance conditions can result in layer temperatures which deviate from the local radiosonde reports, which include meso-scale structures as well as any random or systematic observational errors. The Berlin analyses thus represents the synoptic-scale structure of the lower and middle stratosphere.
However in the submitted manuscript we are not arguing that this method provides better temperature values than the temperatures themselves in a re-analysis dataset such as NCEP. But rather for purposes of comparison of the FUB thickness temperature data with the NCEP reanalyses we calculated similarly the mean thickness layer temperatures in NCEP data set as well.

Following the reviewers comment we modified the text accordingly in order to avoid any misunderstanding and the following paragraph was added in section 2.1:.

"The FU-Berlin is an independent stratospheric analysis data set which is based on earlier subjective hand analyses of temperature and geopotential height fields at 50, 30 and 10 hPa for the northern hemisphere, using the 00UT radiosonde reports from the observational network by a team of experienced meteorologists (http://www.geo.fu-berlin.de/en/met/ag/strat/produkte/fubdata/index.html). Hydrostatic and geostrophic balances were assumed, and observed winds were used to guide the height and temperature analyses. The imposition of these balance conditions ensures a consistent dataset. In addition temporal continuity is assured by meteorological control. Note that these balance conditions can result in layer temperatures that deviate from the local radiosonde reports, which include meso-scale structures as well as any random or systematic observational errors (Labitzke et al., 2002; Manney et al., 2004). Earlier studies using the FU-Berlin dataset point that the approximate geostrophic balance of the upper winds ensures that the contour analysis will be more representative than the temperature analysis based on scattered radiosonde locations (Zerefos and Mantis, 1977; Mantis and Zerefos, 1979). The FU-Berlin analyses thus represent the synoptic-scale structure of the lower and middle stratosphere and the layer-mean temperature derived from the thickness is well suited for an investigation of large-scale climatic fluctuations of temperature. The analyses are provided as gridded data sets with a horizontal resolution of 10o x 10o before 1973, and 5ox 5o thereafter. FU-Berlin geopotential height data are available from July 1957 until December 2001 at 100, 50, and 30 hPa (Labitzke et al., 2002). Hence, from the FU-Berlin dataset we calculated layer-mean temperatures for the two lower stratospheric layers, 100-50 hPa and 50-30 hPa. It should be noted that the FU-Berlin dataset provides geopotential height data already since 1957, but temperature at the same levels since 1964. Hence, aiming in this study at presenting the stratospheric temperature trends from the earliest possible time, we have used the independent FU-Berlin stratospheric dataset with the choice of layer-mean temperature derived from geopotential heights thus extending the records in the past. The variability and trends derived using this dataset have been compared in the past to stratospheric data from other sources, both observations and reanalysis. The overall comparison is good, with differences in the variability (in the earlier period before 1980) that can be attributed mainly to the close match between the FU-Berlin analysis and the to radiosonde observations. (e.g. Randel et al., 2009; Labitzke and Kunze, 2005; Manney et al., 2004; Randel et al, 2004; also in Labitzke et al.,2002 and references therein)."

Furthermore following the suggestion of the reviewer we put our trend calculation in context with MSU channel 4 and SSU channel 1 trend calculations from previously published work. The following text was added in Section 4:

" Our post-1980 year round stratospheric temperature trends at layers L4 (100-50 hPa) and L5 (50-30 hPa) are in the range of calculated trends in Microwave Sounding Unit (MSU) channel 4 (15-20 km) and Stratospheric Sounding Unit (SSU) channel 1 (25-35 km). MSU channel 4 trends derived from RSS and UAH data show cooling trends over the Northern Hemisphere ranging from -0.2 ^{o}C/decade to -0.5 ^{o}C/decade over the period 1979-2007 (Randel et al., 2009). Comparable cooling trends were obtained for MSU channel 4 after reprocessing by NOAA with the trends at polar latitudes revealing higher uncertainties. The SSU channel 1 trends as processed by the UK Met Office and reprocessed by NOAA show cooling trends ranging from about -0.5 ^{o}C/decade (Met office) to about -1.1 ^{o}C/decade (NOAA) over the period 1979-2005 (Thompson et al., 2012)."

2) The reviewer notes " *2. Considering the parameters used (QBO, stratospheric aerosol optical depth), the regression model seems to be best suited for the evaluation of temperature trends in the stratosphere. Although it is quite clear that the study focuses on stratospheric temperature trends, results are also presented for the troposphere. Can the authors comment on the validity of the temperature trends in the troposphere?* "

The following text was added in Section 3.1 addressing the point of the reviewer.
"It should be also noted that the year-round tropospheric temperature trends in the post-1980s period calculated in NCEP (see supplementary material SMT3), RICH (see supplementary material SMT4) and WACCM model (see supplementary material SMT6) for the three latitudinal belts are within the range of respective calculations in previously published work based on different radiosonde datasets (Randel et al., 2009).

3) The reviewer notes " *3. Some more information should be provided on the multiple regression analysis. Since trends are calculated for two time periods, what is the sensitivity of the temperature to the other parameters (QBO, solar cycles) in both these periods? How the model reproduce this sensitivity?*"

The following paragraph was added in Section 3.1:
"The effects of natural forcings derived from our multi-linear analysis are in a generally good agreement with previous studies (e.g. Randel et al., 2009), given that we use layer-mean temperatures and different latitude band averages. The effects of solar and volcanic forcing are found to be more pronounced after 1980. Although the QBO signal is very small and insignificant in the troposphere, we have used the same regression model throughout the atmosphere for uniformity and consistency. "

4) The reviewer notes " *4. The trends are computed in specific latitude bands (e.g. 5-30_N, 30-60_N and 60- 90_N). Considering the position of the tropical barrier in the stratosphere, the former latitude band mixes tropical air with mid-latitude air. Can the authors comment on*
this point? Also, how representative are temperature trends in winter and spring in the 60-90_N latitude band, considering the presence of the polar vortex during these seasons? Could the formation of the vortex influence the large cooling trends found in February, especially during the earlier period in the polar regions?"

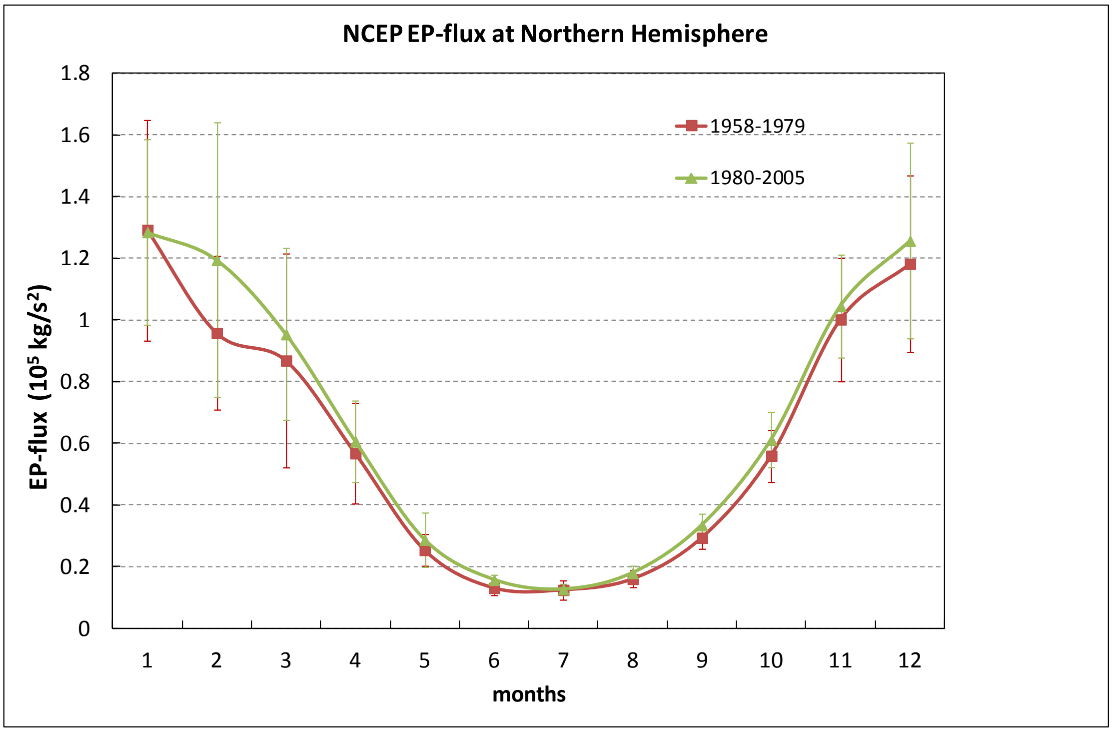
We agree with the reviewer that there is mixing of mid-latitude air with tropical air in and out of a tropical barrier in the stratosphere which according to previous model and observational studies usually ranges at Northern Hemisphere between 20 ^{o}N and 30 ^{o}N with the barrier being weaker in winter and stronger in summer in response to seasonal variation of the tropical zonal-mean flow and wave spectrum (Bowman and Hu, 1997). In our study we consider as NH tropical barrier the 30 ^{o}N which is within the range of estimates from previous model and observational studies. Furthermore the 30 ^{o}N tropical barrier has been also used in other previously published studies (e.g. Randel et al., 2009; Seidel and Randel, 2006).

As far as it concerns the temperature trends at polar latitudes especially during the earlier period we agree with the reviewer that should be considered with caution due to the presence of the polar vortex and the high natural variability. We have already mentioned in the manuscript (at the discussion section) that "At polar latitudes (60^{o}N-90^{o}N), though, the lower stratosphere cooling trends are either non-statistically significant or marginally significant at the 95% confidence level for all datasets. This finding could be related to the competing dynamical and radiative processes that may reduce the statistical significance of these trends."

At the polar latitudes we find a characteristic abrupt decrease (or elimination) of the cooling trend in from winter to early spring for the pre-1980s which is a common feature in all three datasets (NCEP, RICH and FUB) which could be related to dynamical characteristics and the strength of the polar vortex. We have looked the EP-flux through the tropopause at 100 hPa and 45-75 degrees north as calculated from the NCEP data. The following Figure shows the mean seasonal cycle of the EP-flux for the pre-1980s period and the post-1980s period. In the pre-80s period we can clearly note that in February the EP-flux has significantly lower values than the respective value in the post-1980s period. This indicates that February in the pre-1980s period the polar vortex should be stronger and colder due to the weaker BD circulation compared to the post-1980s period. This may have an impact on the strong cooling trend during February in the pre-1980s period and the characteristic abrupt decrease (or elimination) of the cooling trend in from winter to early spring.

Following the reviewer's question the text was modified as follows in Section 3.2:

" At polar latitudes, we find non-statistically significant (at 90% confidence level) cooling trends for all months in NCEP, except in February-March with a characteristic abrupt enhancement of the cooling trend for the pre-1980s (Fig. 3e).In the post-80s period the cooling trends are non-statistically significant for all months except in March-April with strongest cooling signal which might be associated to the Arctic ozone depletion by ODSs (Figure 3f). In the lower stratosphere over polar latitudes for the pre-80s period, both RICH (Figure 4e) and FU-Berlin (Figures 5a and 5c) datasets do not show statistically significant (at 90% confidence level) negative trends. However, it should be noted that the abrupt shift in trend from winter to early spring is a common feature in all three datasets which could be related to dynamical processes and the related variability of the polar vortex."



However as we discuss within the paper the emphasis is given in summer when the stratosphere is less disturbed because it is characterised by lower vertically propagating wave activity from the troposphere, it has smaller interannual natural variability than winter and it is also not influenced by chemical ozone depletion due to ODSs at high latitudes.

5) The reviewer notes " *5. More information should be given on the validity of the FU-Berlin record, which seems to be quite noisy in the early period. Results from this data set also show significant positive values in some months in the early period, in contrast to results based on the other data sets. A more detailed discussion of the various monthly trend results is thus recommended.*"
The variability and trends derived using this dataset have been compared for the past to stratospheric data from other sources, both observations and reanalysis.

The overall comparison is good, with differences in the variability (in the earlier period before 1980) attributed mainly to the close matching of the FUB analysis to radiosonde observations. (e.g. Randel et al., 2009; Labitzke and Kunze, 2005; Manney et al., 2004; Randel et al, 2004; also in Labitzke et al.,2002 and references therein).
Following the reviewer's comment the above mentioned paragraph was added within the revised manuscript in Section 2.1.

6) The reviewer notes "*Significance of trends and correlation coefficients should be indicated in the contour figures.*"
The Figures were revised accordingly.

7) The reviewer notes "*In section 3.3 the significance of correlation coefficients is not provided.*"
The correlations in Figure 7 with $\rho > 0.3$ or $\rho < -0.3$ are statistically significant at the 95% confidence level. This was added in Figure caption of Fig. 7. The contours indicate the statistically significant correlations at 95% significance level with .$\rho > 0.3$ or $\rho < -0.3$.

8) The reviewer notes "*P1078, l16: what is meant by "low frequency variability of the BD circulation"?*"
The low frequency variability of BD circulation was changed to "from inter-annual to decadal variability" in order to be more specific.

9) The reviewer notes "*P1080, l7: Do the derived quantities correspond to age of air? The text should be more specific.*"

The text was modified accordingly:
In contrast, other studies using balloon-borne measurements of stratospheric trace gases over the past 30 years to derive the mean age of air from sulfur hexafluoride (SF6) and CO2 mixing ratios found no indication of an increasing meridional circulation (Engel et al., 2009). Furthermore, Iwasaki et al. (2009) pointed that the yearly trends in BD strength, diagnosed from all re-analyses products over the 10 common period 1979–2001, are not reliably observed due to large diversity among the reanalyses.

Reply to Reviewer #3

We would like to thank Reviewer #3 for the constructive and helpful comments. Reviewer's contribution is recognized in the acknowledgments of the revised manuscript. It follows our response point by point.

1) The reviewer notes "*Page 1075, line 12: the authors mention the time period 1958-2011 here, but the Figs.1-7 all refer to some different time periods ending either 2001, 2005, or 2010. Would be nice to have some clarification.*"

The different time periods used in Fig. 1-7 results from the different time periods of the used datasets. For example as pointed in Section 2.1 the NCEP/NCAR reanalysis data cover the whole referred period from 1958 to 2011. The FU-Berlin dataset from 1958 to 2001, RICH dataset from 1958 to 2006 and the historical simulation with CESM1-WACCM from 1958 to 2005. So in Figures 1 and 2 as well as in Table 1 the NCEP data span the whole period 1958-2011. Mind also in Table 1 that for comparability reasons with the other datasets the trend calculations in NCEP cover also the periods 1980-2001 and 1980-2005. Similarly, for comparability reasons among the different datasets we decided to show in Figures 3,4, 6 and 7 the trends in common periods 1958-1979 and 1980-2005.

We added the following sentence in Section 2:

" Finally, it should be noted that the selection of various time periods is related to the different time periods of the used datasets aiming to a more representative comparison among them."

2) The reviewer notes "*Page 1080, line 29: Please add some points why mean temperature from thickness would be expected to improve homogeneity in both space and time.*"

Our argument for the preference to use thickness to obtain layer mean temperatures refereed mainly for the case of the FUB dataset as it was pointed in our earlier studies based on this dataset (Zerefos and Crutzen, 1975; Zerefos and Mantis, 1977; Mantis and Zerefos, 1979). In the papers of Zerefos and Mantis (1977) as well as Mantis and Zerefos (1979) it was emphasized that the approximate geostrophic balance of the upper winds insures that the contour analysis will be more representative than the temperature analysis which for the FUB data it was based on scattered radiosonde locations.
Please note that the FUB data used in this study were derived from daily stratospheric map analyses of isobaric surfaces prepared by the Stratospheric Research Group of the Free University of Berlin (http://www.geo.fu-berlin.de/en/met/ag/strat/produkte/fubdata/index.html). The FUB dataset consists of 35 years of daily (or bi-daily in summer) geopotential height and temperature fields at 50, 30 and 10hPa in the northern hemisphere. The

hemispheric analyses were produced in real time by a subjective analysis technique, using the 00UT radiosonde reports from the observational network, by a team of experienced meteorologists. Both geostrophic and hydrostatic balance were assumed in the analysis procedure, and the wind observations were given a high priority. The imposition of these balance conditions ensures a consistent dataset; further, temporal continuity is assured by meteorological inspection. Note that these balance conditions can result in layer temperatures which deviate from the local radiosonde reports, which include meso-scale structures as well as any random or systematic observational errors. The Berlin analyses thus represent the synoptic-scale structure of the lower and middle stratosphere.

However in the submitted manuscript we are not arguing that this method provides better temperature values than the temperatures themselves in a re-analysis dataset such as NCEP. But rather for purposes of comparison of the FUB thickness temperature data with the NCEP reanalyses we calculated similarly the mean thickness layer temperatures in NCEP.

Following the reviewers comment we modified the text accordingly in order to avoid any misunderstanding an the following paragraph was added in section 2.1:.

" The FU-Berlin is an independent stratospheric analysis data set which is based on subjective hand analyses of temperature and geopotential height fields at 50, 30 and 10 hPa for the northern hemisphere, using the 00UT radiosonde reports from the observational network by a team of experienced meteorologists (http://www.geo.fu-berlin.de/en/met/ag/strat/produkte/fubdata/index.html).

Hydrostatic and geostrophic balances were assumed, and observed winds were used to guide the height and temperature analyses. The imposition of these balance conditions ensures a consistent dataset. In addition temporal continuity is assured by meteorological or control?. Note that these balance conditions can result in layer temperatures that deviate from the local radiosonde reports, which include meso-scale structures as well as any random or systematic observational errors (Labitzke et al., 2002; Manney et al., 2004). Earlier studies using the FU-Berlin dataset point that the approximate geostrophic balance of the upper winds ensures that the contour analysis will be more representative than the temperature analysis based on scattered radiosonde locations (Zerefos and Mantis, 1977; Mantis and Zerefos, 1979). The FU-Berlin analyses thus represent the synoptic-scale structure of the lower and middle stratosphere and the layer-mean temperature derived from the thickness is well suited for an investigation of large-scale climatic fluctuations of temperature. The analyses are provided as gridded data sets with a horizontal resolution of 10o x 10o before 1973, and 5ox 5o thereafter. FU-Berlin geopotential height data are available from July 1957 until December 2001 at 100, 50, and 30 hPa (Labitzke et al., 2002). Hence, from the FU-Berlin dataset we calculated layer-mean temperatures for the two lower stratospheric layers, 100-50 hPa and 50-30 hPa. It should be noted that the FU-Berlin dataset provides geopotential height data already since 1957, but temperature at the same levels since 1964. Hence, aiming in this study at presenting the stratospheric temperature trends from the earliest

possible time, we have used the independent FU-Berlin stratospheric dataset with the choice of layer-mean temperature derived from geopotential heights thus extending the records in the past. The variability and trends derived using this dataset have been compared in the past to stratospheric data from other sources, both observations and reanalysis. The overall comparison is good, with differences in the variability (in the earlier period before 1980) that can be attributed mainly to the close match between the FU-Berlin analysis and the to radiosonde observations. (e.g. Randel et al., 2009; Labitzke and Kunze, 2005; Manney et al., 2004; Randel et al, 2004; also in Labitzke et al.,2002 and references therein).

3) The reviewer notes "*Page 1081, line 7: Please mention explicitly what variables you are referring to when using the term "data", e.g. temperature, pressure etc.?*"

We clarified this point in the revised version as follows: "Tropospheric and stratospheric temperature, pressure and geopotential height data used in this study are based on the following sources:"

4) The reviewer notes "*Page 1084, equation 1: Please explain what "t" and "T" stands for?*"

We clarified this point in the revised version by adding the following phrase: "Where Mt is the monthly deseasonalized zonal mean temperature and t is the time in months with t = 0 corresponding to the initial month and t = T corresponding to the last month." Also the other terms of equation 1 were also clarified in the revised version.

5) The reviewer notes "*Page 1084, equation 2: Please double check, whether it is the variable "Nt" or if "t" is an index (same for "et")? Please explain variable '.*"

Nt is the unexplained noise term assumed to be autoregressive with time lag of 1 month. It is specified within the text. The "t" is like a time index in the Nt residual time series. We changed Nt to N(t) as a time function of the noise.

6) The reviewer notes "*Page 1085, line 17 and associated figures: The text refers to the layer mean temperature, but the figures do not have any units associated with either the temperature or the layer thickness.*"

The units were added in the figure captions.

7) The reviewer notes "*Page 1089, lines 4-7: I think what is shown in the figures is a decrease in the cooling trend, but not a shift from a negative trend in winter to a positive trend in early spring as the values are still negative.*"

The sentences were modified accordingly:

" At polar latitudes, we find non-statistically significant (at 90% confidence level) cooling trends for all months in NCEP, except in February-March with a characteristic abrupt enhancement of the cooling trend for the pre-1980s (Fig. 3e)."

However, it should be noted that the abrupt shift in trend from winter to early spring is a common feature in all three datasets.

8) The reviewer notes "*Figures: Apart from figure 5 all other figures are pretty tiny and very hard to read. What are the units for the y-scales in figures 1 and 2?*"

The units were added in Figure captions of Figures 1 and 2. All Figures (except Figure 5) were redrawn in order to be more comprehensible.