Final warming of the Southern Hemisphere polar vortex in high- and low-top CMIP5 models

Article

Accepted Version

Author final version

Wilcox, L. ORCID: https://orcid.org/0000-0001-5691-1493 and Charlton-Perez, A. J. ORCID: https://orcid.org/0000-0001-8179-6220 (2013) Final warming of the Southern Hemisphere polar vortex in high- and low-top CMIP5 models. Journal of Geophysical Research - Atmospheres, 118 (6). pp. 2535-2546. ISSN 0148-0227 doi: https://doi.org/10.1002/jgrd.50254 Available at https://centaur.reading.ac.uk/30939/

Final warming of the Southern Hemisphere polar vortex in high- and low-top CMIP5 models

L. J. Wilcox and A. J. Charlton-Perez

Department of Meteorology, University of Reading, Earley Gate, PO Box 243, Reading, RG6 6BB, UK

Abstract

The final warming date of the polar vortex is a key component of Southern Hemisphere stratospheric and tropospheric variability in spring and summer. We examine the effect of external forcings on Southern Hemisphere final warming date, and the sensitivity of any projected changes to model representation of the stratosphere. Final warming date is calculated using a temperature-based diagnostic for ensembles of high- and low-top CMIP5 models, under the CMIP5 historical, RCP4.5, and RCP8.5 forcing scenarios. The final warming date in the models is generally too late in comparison with those from reanalyses: around two weeks too late in the low-top ensemble, and around one week too late in the high-top ensemble. Ensemble Empirical Mode Decomposition (EEMD) is used to analyse past and future change in final warming date. Both the low- and high-top ensemble show characteristic behaviour expected in response to changes in greenhouse gas and stratospheric ozone concentrations. In both ensembles, under both scenarios, an increase in final warming date is seen between 1850 and 2100, with the latest dates occurring in the early twenty-first century, associated with the minimum in stratospheric ozone concentrations in this period. However, this response is more pronounced in the high-top ensemble. The high-top models show a delay in final warming date in RCP8.5 that is not produced by the low-top models, which are shown to be less responsive to greenhouse gas forcing. This suggests that it may be necessary to use stratosphere resolving models to accurately predict Southern Hemisphere surface climate change.

1 Introduction

The Southern Hemisphere (SH) stratosphere and troposphere have been shown to be coupled, with wave driving from the upward propagation of tropospheric Rossby waves influencing the stratospheric zonal wind, and anomalies in the stratospheric polar vortex having an impact down to the surface. This coupling predominantly occurs in the late spring, or summer, when the final warming of the polar vortex strongly influences both the stratospheric and tropospheric circulation (Black et al., 2006), resulting in the stratospheric and tropospheric annular mode having its largest variance in this season (Baldwin et al., 2003). Changes in the strength of the polar vortex are associated with persistent circulation anomalies in the lower stratosphere, with weaker flow resulting in negative Southern Annular Mode (SAM) anomalies. Thompson et al. (2005) showed that final warming events are also associated with tropospheric circulation anomalies of the same sign, which can persist for in excess of two months. They found that significant increases in tropospheric geopotential height over the pole and decreases in the midlatitudes, with a similar structure to the negative phase of the SAM, followed major weakenings in the SH polar vortex. Coherent changes in Antarctic surface temperature, with positive temperature anomalies over much of the continent outside the Peninsula region, were also identified in association with these changes.

Climate forcings have been shown to change the final warming date of the SH polar vortex. In recent years, changes have been found to be strongly determined by decreases in stratospheric ozone concentrations, with final warming dates observed to be later in the 1990s compared to the 1980s (Waugh et al. (1999); Zhou et al. (2000); Karpetchko et al. (2005); Langematz and Kunze (2006); Haigh and Roscoe (2009)). Ozone depletion causes local cooling over the pole, resulting in an increased temperature gradient and a stronger vortex, and hence, later final warming dates.

Several studies have suggested that, in SH spring, the effects on surface climate of ozone recovery and increasing greenhouse gases will be equal and opposite, leading to a near cancellation, or even a reversal, in current trends in the early twenty-first century (Arblaster et al. (2011); McLandress et al. (2011); Polvani et al. (2011); Thompson et al. (2011); Wilcox et al. (2012)). Ozone depletion causes a larger local decrease in temperature compared to greenhouse gas increases, and has been shown to be the primary driver of recent changes in final warming date (Langematz and Kunze, 2006). It is expected that ozone recovery will similarly be the primary driver of near-term changes in final warming date, and that the vortex breakdown will become earlier. A return to later dates towards the end of the twenty-first century is possible as lower stratospheric temperature trends become dominated by well-mixed greenhouse gas forcing, which has been shown to result in an increased temperature gradient near 100 hPa (Shindell et al. (1998); Wilcox et al. (2012)). If these changes are coupled to the surface then changes in springtime Antarctic surface temperature trends would be likely to occur in conjunction with these changes in the vortex. Therefore, one important facet of the stratospheric impact on tropospheric climate is how external forcings may change the final warming date.

The significant tropospheric circulation anomalies associated with final warming events demonstrate that changes in the timing of this phenomenon will play a key role in future SH tropospheric circulation change (Black and McDaniel, 2007). Hence, understanding potential changes in final warming date, and their drivers, is an important part of SH climate prediction. Several studies have shown that the final warming signature in the SH propagates downwards (e.g. Baldwin et al. (2003); Thompson et al. (2005)). Hardiman et al. (2010) recently showed that this propagation begins at 1 hPa. As such, the representation of changes in final warming date may be sensitive to the position of the model top, which is often located near or below 1 hPa in models. Here, we attempt to quantify the effect of external forcings on SH final warming date, and the sensitivity of any projected changes to the position of the model top.

2 Data and Methods

The aim of this study is to identify robust changes in SH final warming date, their drivers, and their potential sensitivity to the position of the model top. The fifth Coupled Model Intercomparison Project (CMIP5) provides a unique opportunity to analyse the response of a large number of models to the same future greenhouse gas scenarios. CMIP5 also includes a substantial number of 'high-top' models, which have an explicit representation of the stratosphere. High-top models have been defined here as those with model tops at pressures $\leq$1 hPa, or altitudes $\geq$ 45 km. In addition to having a higher model top, the high-top models used in this study typically have higher vertical resolution in the stratosphere, and a larger proportion of model levels above 200 hPa (54% of high-top model levels are in the stratosphere, compared to 36% for low-top models). The models used in this study, their classification, and vertical distribution of levels, are shown in Table 1. Only one model from each model family is included in each classification to avoid biasing the ensemble mean.

We examine monthly mean data from the historical (1850-2005), Representative Concentration Pathway (RCP) 4.5 (Thomson et al., 2011), and RCP8.5 (both 2006 to 2100) (Riahi et al., 2011) integrations. The two future pathways result in a radiative forcing of 4.5 Wm^{-2} and 8.5 Wm^{-2} respectively by 2100, with RCP4.5 carbon dioxide emissions peaking around 2040, and RCP8.5 emissions peaking in 2100. The rate of change of greenhouse gas concentrations stabilises by $\sim$2070 in RCP4.5, and continues to increase throughout RCP8.5 (Figure 1(a)). The time series analysed in this paper are concatenations of the historical and RCP experiments for consistent ensemble members of each model, and are referred to throughout by the name of the relevant future pathway.

Although a recommended ozone time series was compiled for CMIP5 (Cionni et al., 2011), only three of the models used in this study are forced with these data. Others included modified versions of the Cionni et al. (2011) data, some prescribed ozone concentrations from different data sets, and others treat ozone interactively. The different representation of ozone in the subset of CMIP5 models used in this study is shown in Table 1, following the categorisation of Eyring et al. (2012). Example time series of the September to November mean 75°-90°S mean concentration at 50 hPa for each prescribed category are shown in Figure 1(b), alongside the time series from models with interactive ozone. Comparison of the different categories reveals a range of Antarctic stratospheric ozone concentrations, with 1900 values between 2.4 ppmv and 4 ppmv. There is some spread in the rate of recovery in the twenty-first century. Ozone concentrations tend to recover faster in the time series from models with interactive ozone. The relative change in ozone concentrations prior to 2000 is similar in the interactive and Cionni timeseries, but smaller in the other prescribed categories. However, the turning points are comparable across the categories (Figure 1(b)). The aim of this study is to identify the drivers of robust projections in SH final warming date, which will depend on the forcings, and the response to them, having the same characteristics across the model ensemble. As the turning points in the ozone timeseries are comparable, it is anticipated that the qualitative response of the final warming date to ozone will have similar characteristics across the models. Hence, the quantitative differences in the ozone forcing are not anticipated to influence our result.

To date, different numbers of ensemble members have been provided for each of the CMIP5 models. Where multi-model means have been used, they include only one ensemble member for each model to avoid biasing the mean towards models with a larger number of ensemble members.

ERA-Interim (Dee et al., 2011) and the NCEP Climate Forecast System Reanalysis (CFSR) (Saha et al., 2010) were used to assess biases in the model data.

2.1 Final warming diagnostic

The definition of vortex breakdown is subjective, and several approaches have been used in earlier studies. These include potential vorticity-based spatial diagnostics (Waugh and Randel (1999), Waugh et al. (1999), Karpetchko et al. (2005), Zhou et al. (2000)), diagnostics based on wind thresholds (Black and McDaniel, 2007), and temperature based diagnostics (Haigh and Roscoe, 2009). However, regardless of the definition used, there is a consensus that the final warming date (FWD) of the SH vortex was later in the 1990s compared to the 1980s. Potential vorticity is not a standard CMIP5 output, and the coarse vertical resolution of the archived data makes it difficult reliably to calculate potential vorticity. Therefore, only temperature (Haigh and Roscoe, 2009) and wind (Black and McDaniel, 2007) based diagnostics of the FWD have been considered.

Black and McDaniel (2007) defined the FWD as the final time that the zonal-mean zonal-wind at 60°S and 50 hPa drops below 10 ms^{-1} until the following autumn. They apply the diagnostic to 5-day running averages of daily data.

Haigh and Roscoe (2009) define the FWD as the minimum in the second time derivative of polar cap mean (90-60°S) temperature at 50 hPa. They use 3-day averages of daily and bi-daily data, smoothed with a 21-day triangular filter. However, they found that interpolation of monthly mean data gave similar fields to smoothed daily data. Here, monthly mean data is used as, at this early stage in CMIP5, it facilitates the analysis of a larger number of models. The sum of the first five Fourier components of the temperature time series is used to produce interpolated daily data. Due to the smooth nature of the evolution of the seasonal cycle in polar cap mean temperature, only negligible differences were identified between FWDs calculated using this method, and those calculated using daily data (see Figure 2).

The FWD calculated using the Haigh and Roscoe method is typically a week earlier than that calculated using the Black and McDaniel diagnostic. However, there is little qualitative difference between the diagnostics (Figure 2): the time series are strongly correlated, with r=0.95 for 1950-2005 for CNRM-CM5 data. The use of the Black and McDaniel (2007) threshold-based diagnostic may be problematic if there are significant variations in the background state between models, or under strong forcing. In some models, the use of the 10 ms^{-1} threshold results in non-identification of a FWD for some years in the historical period. As scenarios with large forcing will be considered, the Haigh and Roscoe diagnostic, from monthly mean data, will be used for the remainder of this work, in order to avoid excessive non-identification of FWDs.

2.2 Empirical mode decomposition

Climate data is often non-linear and non-stationary. Deviations from monotonic change are particularly apparent in the Southern Hemisphere where change is governed by the competing effects of increased greenhouse gases and stratospheric ozone. Changes in FWD have been established as being strongly ozone driven (Zhou et al. (2000), Karpetchko et al. (2005), Haigh and Roscoe (2009)), and a better fit is found between FWD and stratospheric ozone concentrations than can be achieved with linear trends for example (Haigh and Roscoe, 2009).

To avoid fitting extrinsic functions, which may not correspond well to the non-linearity embedded in the data, or forcing data time series, which may only account for changes via one of many mechanisms, Empirical Mode Decomposition (EMD) has been used to analyse variability in FWD. EMD is an intrinsic, adaptive method for deriving the variability of a time series on various timescales. EMD has successfully been applied to climate data in several previous studies (e.g. Lee and Ouarda (2011), Franzke (2009), Huang and Wu (2008), Wu et al. (2007), McDonald et al. (2007), and Duffy (2004)). While EMD is a useful tool for analysing variability and trends in non-linear time series, it cannot be used to unambiguously attribute particular characteristics of these trends to a given forcing mechanism. Hence, EMD is used here alongside multiple linear regression analysis.

EMD is an algorithm used to decompose a time series into a set of Intrinsic Mode Functions (IMFs), with each describing a given oscillatory mode of the data. IMFs must satisfy two conditions:

- Must have a local mean of zero
- Must have a single zero crossing between two extrema

IMFs are extracted sequentially from a data series, from the highest frequency to the lowest, until no complete oscillation can be identified. The residual from this process then describes the long-term trend in the data, where the trend is defined as the instantaneous mean of the time series.

Unlike Fourier filtering, the phase and amplitude of each IMF are time dependent. The number of IMFs extracted from a time series is typically $\ln N$, where N is the number of data points (Wu et al., 2007). There is some evidence of mode mixing (signals of different timescales identified in the same IMF) amongst the IMFs of FWD from EMD. To avoid this, Ensemble Empirical Mode Decomposition (EEMD) has been used. EEMD gives an ensemble mean of the IMFs for the product of FWD and a finite white noise series (Wu and Huang, 2009). The inclusion of a noise series provides a uniformly distributed reference scale, which preserves the dyadic property of EMD that can fail when data is intermittent (Wu and Huang, 2009). The noise is cancelled out in the ensemble mean, so it can be used to facilitate the separation of different timescales, without contributing to the final IMFs. EEMD is performed here with 200 iterations and white noise with an amplitude of 0.2 times the standard deviation of the FWD series (following Wu and Huang (2009)).

Figure 3(a) shows a time series of FWD from MIROC-ESM-CHEM under RCP4.5, calculated using the Haigh and Roscoe (2009) method, alongside the IMFs from EEMD (Figure 3(b)). Most of the high-frequency variability in the time series, with a period of less than 3 years, is contained in the first two IMFs (not shown). The local maximum near 2000 is captured in the sixth IMF, and the increasing trend through the period shown is captured in the residual. The equivalent result using EMD is shown in Figure 3(d). In this example, it can be seen that the different frequencies have not been satisfactorily separated. This is particularly clear in the third IMF (top line of Figure 3(d)), where the period of the oscillation around the year 2000 is double that in the rest of the IMF.

IMFs that can be distinguished from the equivalent IMFs of a noise time series of the same length are significant, and can be taken to represent physically meaningful signals. White (Wu and Huang (2004), Wu et al. (2007)) and red (Franzke (2009)) noise have both been used in previous studies to assign significance to IMFs from climate data. There is no physical reason why the FWD in one year would be dependent on the date in another year (Black et al. (2006) also considered each event as an independent sample). Therefore, a comparison with a white noise series has been used to determine when an IMF is significant, following Wu and Huang (2004).

A significant difference from a white noise time series is identified through analysis of the period (T) and energy density (E) of each IMF. Wu and Huang (2004) show that the probability density function for each IMF of a white noise time series is well approximated by a normal distribution, and that the probability distribution of the energy of the nth IMF, NE_n, is a χ^2 distribution, with $N\overline{E}_n$ degrees of freedom, where $\overline{E}_n$ is the mean of E_n when the number of data points, N, approaches ∞. The spread of different confidence intervals as a function of the mean energy of each IMF can then be determined. Wu and Huang (2004) define $y = \ln E$ and show that for $|y - \overline{y}| \ll 1$, the distribution of the energies is Gaussian. The spread lines can then be approximated by

$$y = -x \pm k\sqrt{\frac{2}{N}}e^{x/2} \quad (1)$$

where $x = \ln \overline{T}_n$, $\overline{T}_n$ is the mean period, and k is a constant from the percentiles of the normal distribution. Example energies and periods from 1000 white noise time series of 1000 data points, and the spread lines from the 95% confidence interval, are shown in Figure 3(c). Energy densities from a data time series that lie outside the bounds of the spread lines can be assumed to be significantly different from those expected from a white noise time series, and are therefore expected to contain some information at that confidence level.

3 Past and future trends in final warming date

Mean FWDs in the individual models are shown in Figure 4 for three periods: 1870-1900, 1979-2005, and 2070-2098. In most cases, the FWD is one to two weeks later in 2070-2098 compared to 1870-1900. In the RCP4.5 experiment the delay ranges from a change of 1 day in INMCM4 to 9 days in CanESM2, CSIRO-Mk3.6.0, GISS-ES-R, and NorESM1-M (Figure 4(a)). In RCP8.5 the delay compared to 1870-1900 ranges from 2 days in INMCM4 to 15 days in CanESM2 (Figure 4(b)). With the exception of CNRM-CM5 and GISS-E2-R, all models have later FWDs in 2070-2098 in the RCP8.5 experiment than in RCP4.5. Figure 4(c) compares FWD from 1870-1900 to 2070-2098. There is some evidence of a saturation effect here, with models with a very late historical FWD appearing to show less of a change in the future.

Figure 4(d) shows the 1979-2005 mean FWD for each model, compared to ERA-Interim and CFSR. In all models except MIROC5, the FWD is too late compared to the reanalyses, with most models having an FWD that is significantly later. Such a late bias has been identified in earlier model evaluations, e.g. Butchart et al. (2011). It can also be seen in Figure 4(d) that most models underestimate the inter-annual variability in FWD compared to reanalyses.

The late bias in model FWDs is reflected in the high- and low-top ensemble means, shown in Table 2, and in Figure 5 alongside those from ERA-Interim and CFSR. The mean FWDs in the period 1979-2005 are day 312 and day 313 in ERA-Interim and CFSR respectively. The low-top mean FWD is around 2 weeks late, with a 1979-2005 mean of day 327. The high-top ensemble mean is in better agreement with the reanalysis values, but is still late on average, with a 1979-2005 mean of day 321 (Table 2). For all periods shown in Figure 4, the mean FWD from the low-top ensemble is around a week later than that from the high-top ensemble (Table 2).

The FWD from the low- and high-top ensemble is shown in Figure 6 for the historical and RCP4.5 and historical and RCP8.5 experiments. There is more inter-model spread and inter-annual variability in the low-top ensemble, although there is still a considerable amount of inter-annual variability in the FWD from the individual high-top models.

A marked delay in FWD can be seen in the high-top ensemble from the late 1970s to the late 1990s (Figure 6). This is associated with the localised, seasonal, cooling that results from ozone depletion in this period. Under RCP4.5 this increase in FWD is followed by a steady decrease to 2100, but in RCP8.5 a more modest decrease is seen, followed by a small trend towards later FWDs by 2100. The large inter-model spread amongst the low-top models makes such features difficult to distinguish in the low-top ensemble. However, there is some sense of a shift towards later FWDs in the late twentieth century.

The large interannual variability and inter-model spread in FWD makes it difficult to compare patterns of behaviour across the models, although the spread in absolute values is important to bear in mind. The FWD in all models is now adjusted to the 1860-1900 mean to assist discussion of the change in FWD across the models. In Figure 7, an 11-year running mean has also been applied, which removes high frequency inter-annual variability, without obscuring decadal variability. The ensemble means shown in Figure 7 are calculated by first finding the ensemble mean of the adjusted raw data, then calculating the 11-year running mean.

More similarities can be seen in the behaviour of the low- and high-top models in Figure 7 compared to Figure 6. A clear increase in FWD can now be seen in the low-top ensemble, although the change is not as rapid, large, or as consistent across models, as in the high-top ensemble. A return to earlier FWDs in the twenty-first century can now be seen in the low-top ensemble mean under RCP4.5, although the rate of change is still small compared to that seen in the high-top ensemble. Under RCP8.5, the FWD in the low-top ensemble mean shows very little change in the twenty-first century. In contrast, a clear decrease can be seen in the first half of the twenty-first century in the high-top ensemble, followed by an increase towards the end of the century. The large twenty-first century inter-model spread in the low-top ensemble, even after adjusting to the 1860-1900 mean, may obscure some of this behaviour in the low-top ensemble mean. However, there is no convincing evidence of such a pattern in the FWDs from individual models. Such behaviour can be seen in a number of the high-top models.

4 Drivers of past and future trends in final warming date

The primary drivers of changes in FWD are anticipated to be changes in stratospheric ozone and well-mixed greenhouse gas concentrations. These changes will occur on different timescales, and have different functional forms in the timeseries. As such, their signature can be expected to be seen in different IMFs. Increasing greenhouse gases are expected to be linked to a delay in the FWD, while the depletion and recovery of stratospheric ozone will produce a delay followed by an advance: a signature with a period in the region of 60 years. These responses are likely to be seen in the residual and the last IMF respectively. Figure 1(b) shows that the largest changes in stratospheric ozone concentrations at southern high latitudes occur in the first half of the twenty-first century. Hence, it is anticipated that changing ozone concentrations will be the primary driver of FWD changes here, with greenhouse gases becoming increasingly important in the second half of the century. Figure 1(a) shows that greenhouse gas concentration changes in RCP4.5 and RCP8.5 are very different in the latter half of the century, with almost no change in concentrations in RCP4.5 and rapid increases in RCP8.5. The potential influence of this difference on FWD was hinted at in Figure 7. It is particularly clear in the comparison of the high-top ensemble means for the two scenarios, where a negative trend from $\sim$2070 is seen in RCP4.5 and a positive trend is seen in RCP8.5.

The sum of the residual and the last IMF for each model, and the low- and high-top ensemble means, are shown in Figure 8. The ensemble mean is calculated by finding the ensemble mean of the adjusted data, then performing EEMD on this mean. All models and the ensemble means show, with the exception of MIROC5, later FWDs around the turn of the century, under both RCP4.5 and RCP8.5. Patterns of behaviour seen in the ensemble mean are similar to those seen in the running means in Figure 7: an increase then decrease in FWD under RCP4.5; and an increase then decrease then increase in the high-tops under RCP8.5. There is even a suggestion of this RCP8.5 response in the low-top models HadGEM2-ES and CSIRO-Mk3-6-0. However, the amplitude of twenty-first century changes are smaller in the the low-top ensemble than the high-top case. The larger response of high-top models to greenhouse gas forcing towards the end of the twenty-first century is consistent with the larger temperature gradient changes at the tropopause level simulated by these models (Wilcox et al., 2012).

Significance testing was carried out to determine which IMFs show patterns significantly different to those that may be identified in a white noise time series. The Wu and Huang (2004) method was used, including their assumption that the energy of the first IMF comes solely from noise and can be used to re-scale the energy density of the other IMFs. Figure 9 shows the sum of significant IMFs (at the 5% level) with periods greater than 50 years (in order to consider only inter-decadal variability) for the low- and high-top ensemble mean (Figure 9 (a) and (b) respectively). The signatures of the high- and low-top significant IMFs follow the patterns seen in the running means, and sums of the last two IMFs: a more pronounced peak at the turn of the century in the high-top ensemble, and a trend towards later FWDs at the end of the twenty-first century in RCP8.5 in the high-top ensemble only.

The spread function of the 95% and 99% confidence intervals for white noise and energies of the individual IMFs are shown in Figure 10. Here, a significant IMF is identified when it lies outside the inner pair of dotted lines, which indicate the 5^{th} and 95^{th} percentile for white noise. The outer pair of dotted lines indicate the 1^{st} and 99^{th} percentile.

Figure 10 shows that the residual is clearly significant for both ensembles and scenarios. For the high-top ensemble, the last IMF is also significant at the 1% level for both scenarios. In a reflection of the larger inter-model spread, and the resulting weaker peak in FWD around the turn of the century, the last IMF of the low-top ensemble mean is significant at the 5% level for the historical and RCP4.5 scenario, and not at all for the historical and RCP8.5 scenario (Figure 10(b)). The higher energy of the last IMF in RCP8.5 in the high-top mean compared to the low-top mean is not due only to a differing response to ozone forcing. Analysis of the structure of the IMFs shows that in the high-top

RCP8.5 case the last IMF includes some response to GHG forcing, in addition to the anticipated ozone response. The delay in FWD towards the end of the twenty-first century is incorporated in the last IMF as the timing of the trend fits with the $\sim$60 year period of the response to stratospheric ozone changes.

Multiple linear regression analysis was also performed, regressing FWD against a constant, a timeseries of September to November mean Antarctic mean ozone at 50 hPa, and ln(GHG), where GHG is represented by the CO_2 equivalent values shown in Figure 1(a). Following Roscoe and Haigh (2007), these indices are normalised to allow direct comparison of the regression coefficients. The regression slope, Pearson correlation coefficient, and significance from a two-tailed student's t-test are shown in Table 3 for the high- and low-top ensemble mean for RCP4.5 and RCP8.5. For ensemble mean calculations, ozone was taken from the Cionni et al. (2011) data.

In both scenarios, FWD has a stronger relationship with both the GHG index and the ozone index in the high-top ensemble. This can be seen in the larger regression slopes and linear correlations shown in Table 3, and in comparison of the multiple linear correlations: 0.63 (0.64) and 0.76 (0.78) for RCP4.5 (RCP8.5) for the low- and high-top ensemble mean respectively. This is a reflection of the more consistent cross-model behaviour seen in the high-top models (e.g. Figure 8).

There is little difference between RCP4.5 and RCP8.5 in the statistics relating to the ozone index (Table 3). The more influential role of GHGs in RCP8.5 is reflected in the regression slopes as well as the significance. The larger regression slope, linear correlation, and significance associated with the GHG index in RCP8.5 for the high-top ensemble compared to the low-top is likely to be a reflection of the delay in FWD in the high-top ensemble mean near the end of the twenty-first century in response to GHG forcing , which is not seen in RCP4.5, or the low-top ensemble mean. This echoes the higher energies found in the last IMF and residual of the high-top ensemble mean in RCP8.5.

In the illustrations of FWD in CMIP5 models shown in this study, MIROC5 has been a clear outlier. The model shows almost no change in FWD from 1860 to 2100 (Figure 4, Figure 7) and the structure of the timeseries from the sum of the last IMF and the residual mirrors those from other low-top models. In the high-top ensemble, there are no such striking outliers (Figure 8). However, MIROC-ESM-CHEM shows larger inter-decadal variations in FWD than other models in the group. While the behaviour of FWD in MIROC-ESM-CHEM is not especially unusual in the context of the other models, is it possible that the large changes simulated by MIROC-ESM-CHEM and the very small changes from MIROC5 have enough influence on their respective ensemble means to dominate the differences seen between the high- and low-top ensembles?

It was found that removing the MIROC models from the ensembles had no effect on our conclusions from EEMD analysis at the 5% level. As one would expect, there are small changes to the energies of the IMFs as a result of the removal, but the IMFs identified as being significantly different to those expected from white noise are the same, and their structure is qualitatively unchanged.

The results of the multiple linear regression analysis without the MIROC models is shown alongside the results for the whole ensemble in Table 3. As expected, removing MIROC5, a model that shows little change in FWD, from the low top ensemble slightly increases the correlation between the FWD and both ozone and ln(GHG) in both the RCP4.5 and RCP8.5 case, but not to such an extent that the significance level is altered. The removal of MIROC5 results in an increase in the magnitude of the regression slope for the ozone index, and for the GHG index in the RCP8.5 scenario. It also brings the regression slope for the GHG index closer to the anticipated positive value in the RCP4.5 case.

MIROC-ESM-CHEM simulates a slightly larger response to stratospheric ozone depletion compared to the rest of the high-top ensemble, but doesn't show a delay in FWD towards the end of the twenty-first century. Thus, it is anticipated that the removal of the model from the high-top ensemble will result in a decrease in the magnitude of the regression slope of the ozone index and correlation, and an increase in the regression slope and correlation for the GHG index. Such changes can be seen in both the RCP4.5 and RCP8.5 case (Table 3). These changes are marked enough to decrease the significance of the relationship between stratospheric ozone and RCP8.5 FWD, and of the relationship between RCP4.5 FWD and GHG.

As one would expect, removing the MIROC models from the analysis does change the statistics. However, the conclusions drawn from the analysis are unchanged. The importance of stratospheric ozone changes as a driver of changes in FWD is consistent across both scenarios, with a unit change in ozone concentration having more influence on the high-top ensemble mean than the low-top ensemble mean. GHG changes play more of a role in RCP8.5 than RCP4.5, and, as for ozone changes, result in a larger change in FWD in the high-top ensemble mean than the low-top mean. The larger values of the regression coefficients in the high-top case reflect the higher energies of the residual and last IMF seen in Figure 10, and the more consistent behaviour of the models seen in Figure 8.

5 Conclusions

Changes in final warming date are known to drive persistent tropospheric anomalies with a similar structure to the southern annular mode (Thompson et al. (2005), Black et al. (2006)). Such changes are sensitive to external forcing from greenhouse gases and, in particular, stratospheric ozone. This results in pronounced changes in Southern Hemisphere final warming date, with a peak around the year 2000, which can be expected to influence spring and summertime trends in high-latitude surface climate.

The Southern Hemisphere final warming date is around one week too late in CMIP5 high-top models, and two weeks too late in low-top models compared to ERA-Interim and the Climate Forecast System Reanalysis (1979-2005). The high-top models show more consistent absolute values and changes in final warming date in both the historical and future periods than low-top models

After adjustment to the 1860-1900 mean, similar behaviour can be seen in both the high- and low-top ensembles. A shift to later final warming dates is seen in the historical period as a response to stratospheric ozone depletion, and a return to earlier final warming dates occurs as ozone recovers. In the high-top ensemble, there is also a shift towards later final warming dates in the latter half of the twenty-first century in RCP8.5, which is consistent with the larger meridional temperature gradient identified in high-top models by Wilcox et al. (2012). The high-top models show a more consistent pattern of change, and larger changes, in response to forcing compared to the low-top models. This difference is apparent in both the comparison of significant IMFs, and the coefficients from multiple linear regression.

Further investigations with larger ensembles of high- and low-top models, with consistent ozone concentrations, are required. Simpson et al. (2011) showed that the late bias in final warming date contributes to too-persistent southern annular mode anomalies in summer, and may cause models to respond too strongly to anthropogenic forcing in this season. Hence, the difference between the high- and low-top ensemble mean results, the large spread in the low-top ensemble, and the more pronounced late bias in final warming date in the low-top ensemble, suggest that high-top models are likely to be required to produce accurate projections of future Southern Hemisphere surface climate.

6 Acknowledgments

This work was funded by the National Centre for Atmospheric Science (NCAS)-Climate via a CMIP5 grant.

We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modelling groups for producing and making available the model output listed in Table 1. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.

We thank the British Atmospheric Data Centre (BADC) for providing access to their CMIP5 data archive, and acknowledge the use of ERA data made available by the BADC and NCAS-Climate. NCEP/NCAR data was provided by NOAA/OAR/ESRL PSD, and CFSR data was provided by NCAR, via the Research Data Archive (RDA).

We also thank three anonymous reviewers for their helpful suggestions.

References

Arblaster, J. M., G. A. Meehl and D. J. Karoly, 2011: Future climate change in the Southern Hemisphere: Competing effects of ozone and greenhouse gases. *Geophys. Res. Lett.*, **38**, L02701, doi:10.1029/2010GL045384.

Baldwin, M. P., D. B. Stephenson, D. W. J. Thompson, T. J. Dunkerton, A. J. Charlton, and A. O'Neill, 2003: Stratospheric memory and skill of extended-range weather forecasts. *Science*, **301**, 636–640.

Black, R. X. and B. A. McDaniel, 2007: Interannual variability in the Southern Hemisphere circulation organized by stratospheric final warming events. *J. Atmos. Sci.*, **64**, 2968–2974.

Black, R. X., B. A. McDaniel, and W. A. Robinson, 2006: Stratosphere-troposphere coupling during spring onset. *J. Clim.*, **19**, 4891–4901.

Butchart, N., et al., 2011: Multimodel climate and variability of the stratosphere. *J. Geophys. Res.*, **116**, D05102.

Cionni, I., et al., 2011: Ozone database in support of CMIP5 simulations: results and corresponding radiative forcing. *Atmos. Chem. Phys. Discuss*, **11 (4)**, 10 875–10 933.

Dee, D. P., et al., 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **137 (656)**, 553–597, doi:10.1002/qj.828.

Duffy, D. G., 2004: The application of Hilbert-Huang transforms to meteorological datasets. *J. Atmos. Oceanic Technol.*, **21**, 599–611.

Eyring, V., et al., 2012: Long-term changes in tropospheric and stratospheric ozone and associated impacts in CMIP5 simulations. *J. Geophys. Res.*, submitted.

Franzke, C., 2009: Multi-scale analysis of teleconnection indices: climate noise and nonlinear trend analysis. *Nonlin. Process Geophys.*, **16**, 65–76.

Haigh, J. D. and H. K. Roscoe, 2009: The final warming date of the Antarctic polar vortex and influences on its interannual variability. *J. Clim.*, **22**, 5809–5819.

Hardiman, S. C., N. Butchart, S. M. Osprey, L. J. Gray, A. C. Bushell, and T. J. Hinton, 2010: The climatology of the middle atmosphere in a vertically extended version of the Met Office's climate model. part I: Mean state. *J. Atmos. Sci.*, **67**, 1509–1525.

Huang, N. E. and Z. Wu, 2008: A review on Hilbert-Huang transform: Method and its applications to geophysical studies. *Rev. Geophys.*, **46**, RG2006.

Karpetchko, A., E. Kyrö, and B. M. Knudsen, 2005: Arctic and Antarctic polar vortices 1957-2002 as seen from the ERA-40 reanalyses. *J. Geophys. Res.*, **110**, D21109.

Kawase, H., T. Nagashima, K. Sudo, and T. Nozawa, 2011. Future changes in tropospheric ozone under Representative Concentration Pathways (RCPs). *Geophys. Res. Lett.*, **38**, L05801.

Lamarque, J. F., et al., 2010: Historical (1850-2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: methodology and application. *Atmos. Chem. Phys.*, **10 (15)**, 7017–7039.

Lamarque, J. F., G. P. Kyle, M. Meinshausen, K. Riahi, S. J. Smith, D. P. van Vuuren, A. J. Conley, and F. Vitt, 2011: Global and regional evaluation of short-lived radiatively-active gases and aerosols in the Representative Concentration Pathways. *Climatic Change*, **109(1-2)**, 191–212.

Langematz, U. and M. Kunze, 2006: An update on dynamical changes in the Arctic and Antarctic stratospheric polar vortices. *Climate Dynamics*, **27 (6)**, 647–660.

Lee, T. and T. B. M. J. Ouarda, 2011: Prediction of climate nonstationary oscillation processes with empirical mode decomposition. *J. Geophys. Res.*, **116**, D06107.

McDonald, A. J., A. J. G. Baumgaertner, G. J. Fraser, S. E. George, and S. Marsh, 2007: Empirical mode decomposition of the atmospheric wave field. *Ann. Geophys.*, **25**, 375–384.

McLandress, C., T. G. Shepherd, J. F. Scinocca, D. A. Plummer, M. Sigmond, A. I. Jonsson, M. C. Reader, 2011: Separating the Dynamical Effects of Climate Change and Ozone Depletion. Part II: Southern Hemisphere Troposphere. *J. Climate*, **24**, 1850–1868.

Polvani, L. M., M. Previdi and C. Deser, 2011: Large cancellation, due to ozone recovery, of future Southern Hemisphere atmospheric circulation trends. *Geophys. Res. Lett.*, **38**, L04707, doi:10.1029/2011GL046712.

Riahi, K., et al., 2011: RCP 8.5—a scenario of comparatively high greenhouse gas emissions. *Climatic Change*, **109**, 33–57, 10.1007/s10584-011-0149-y.

Roscoe, H. K., and J. D. Haigh, 2007: Influences of ozone depletion, the solar cycle and the QBO on the Southern Annular Mode. *Q. J. R. Meteorol. Soc.*, **133**, 1855–1864.

Saha, S., et al., 2010: The NCEP Climate Forecast System Reanalysis. *Bulletin of the American Meteorological Society*, **91**, 1015–1057.

Shindell, D..T., D. Rind, and P. Lonergan, 1998: Increased polar stratospheric ozone losses and delayed eventual recovery owing to increasing greenhouse-gas concentrations. *Nature*, **392**, 589-592, doi:10.1038/33385.

Simpson, I. R., P. Hitchcock, T. G. Shepherd, J. F. Scinocca, 2011: Stratospheric variability and tropospheric annular mode timescales. *Geophys. Res. Lett.*, **38**, L20806.

Thompson, D. W. J., M. P. Baldwin, and S. Solomon, 2005: Stratosphere–troposphere coupling in the southern hemisphere. *J. Atmos. Sci.*, **62**, 708–715.

Thompson, D. W. J., S. Solomon, P. J. Kushner, M. H. England, K. M. Grise and D. J. Karoly, 2011: Signatures of the Antarctic ozone hole in Southern Hemisphere surface climate change. *Nature Geoscience*, **4**, 741–749, doi:10.1038/ngeo1296.

Thomson, A., et al., 2011: RCP4.5: a pathway for stabilization of radiative forcing by 2100. *Climatic Change*, **109**, 77–94, 10.1007/s10584-011-0151-4.

Waugh, D. W. and W. J. Randel, 1999: Climatology of Arctic and Antarctic polar vortices using elliptical diagnostics. *J. Atmos. Sci.*, **56**, 1594–1613.

Waugh, D. W., W. J. Randel, S. Pawson, P. A. Newman, and E. R. Nash, 1999: Persistence of the lower stratospheric polar vortices. *J. Geophys. Res.*, **104 (D22)**, 27 191–27 201.

Wilcox, L. J., A. J. Charlton-Perez, and L. J. Gray, 2012: Trends in Austral jet position in ensembles of high- and low-top CMIP5 models. *J. Geophys. Res.*, **117**, D13115.

Wu, Z. and N. E. Huang, 2004: A study of the characteristics of white noise using the empirical mode decomposition method. *Proc. R. Soc. Lond. A*, **460**, 1597–1611.

Wu, Z. and N. E. Huang, 2009: Ensemble Empirical Mode Decomposition: A noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, **1 (1)**, 1–41.

Wu, Z., N. E. Huang, S. R. Long, and C.-H. Peng, 2007: On the trend, detrending, and variability of nonlinear and nonstationary time series. *PNAS*, **104 (38)**, 14 889–14 894.

Zhou, S., M. E. Gelman, A. J. Miller, and J. P. McCormack, 2000: An inter-hemisphere comparison of the peristent stratospheric polar vortex. *Geophys. Res. Lett.*, **27 (8)**, 1123–1126.

Table 1: CMIP5 models used in this study. High top models are are denoted by *. C[1]: Cionni et al. (2011); C[2]: Modified Cionni et al. (2011), with a solar cycle added in future; C[3]: Modified Cionni et al. (2011), with zonal averages in troposphere, and future concentrations in the stratosphere determined by combining two terms in a multiple linear regression analysis; P[1]: Lamarque et al. [2010, 2011]; P[2]: Kawase et al. (2011); S[1]: Ozone concentrations from a chemistry climate model, used offline.

Model	Model top	Number of levels	Number of levels above 200 hPa	% of levels above 200 hPa	Ozone
BCC-CSM1.1	2.917 hPa	26	13	50	C[1]
CNRM-CM5	10 hPa	31	9	29	Interactive
CSIRO-Mk3.6.0	4.52 hPa	31	9	29	C[1]
HadGEM2-ES	40 km (∼2.3 hPa)	38	15	39	C[2]
INMCM4	10 hPa	21	8	38	C[1]
NorESM1-M	3.54 hPa	26	13	50	P[1]
MIROC5	3 hPa	56	17	30	P[2]
CanESM2*	1 hPa	35	10	29	C[3]
GISS-E2-R*	0.1 hPa	40	19	48	Interactive
HadGEM2-CC*	85 km (∼0.01 hPa)	60	37	62	C[2]
IPSL-CM5A-LR*	0.04 hPa	39	22	56	S[1]
MIROC-ESM-CHEM*	0.0036 hPa	80	63	79	Interactive
MPI-ESM-LR*	0.01 hPa	47	25	53	C[2]
MRI-CGCM3*	0.01 hPa	48	20	42	C[2]

Table 2: Final warming date in the high- and low-top ensemble mean, and from reanalyses

	1870-1900	1979-2005	2070-2098 (RCP4.5)	2070-2098 (RCP8.5)
High-top	310	322	317	322
Low-top	318	327	323	325
ERA-Interim/CFSR	-	312/313	-	-

Table 3: Results from multiple linear regression analysis. Significance is from a 2-tailed t-test. Values in brackets show the equivalent values when the MIROC models are excluded from the ensemble mean.

RCP 4.5			
Low-top	Regression slope	Pearson correlation coefficient	Significance
Ozone	-11.00 (-12.89)	-0.63 (-0.69)	<0.1% (<0.1%)
ln(GHG)	-2.68 (-0.38)	0.36 (0.41)	>5% (>5%)
High-top			
Ozone	-14.65 (-12.69)	-0.75 (-0.71)	<0.1% (<0.1%)
ln(GHG)	9.69 (17.12)	0.50 (0.51)	>5% (<5%)

RCP 8.5			
Low-top	Regression slope	Pearson correlation coefficient	Significance
Ozone	-9.94 (-12.39)	-0.63 (-0.69)	<0.1% (<0.1%)
ln(GHG)	11.01 (12.64)	0.39 (0.42)	<1% (<1%)
High-top			
Ozone	-14.51 (-12.75)	-0.76 (-0.72)	<0.1% (<1%)
ln(GHG)	17.27 (22.11)	0.49 (0.51)	<0.1% (<0.1%)

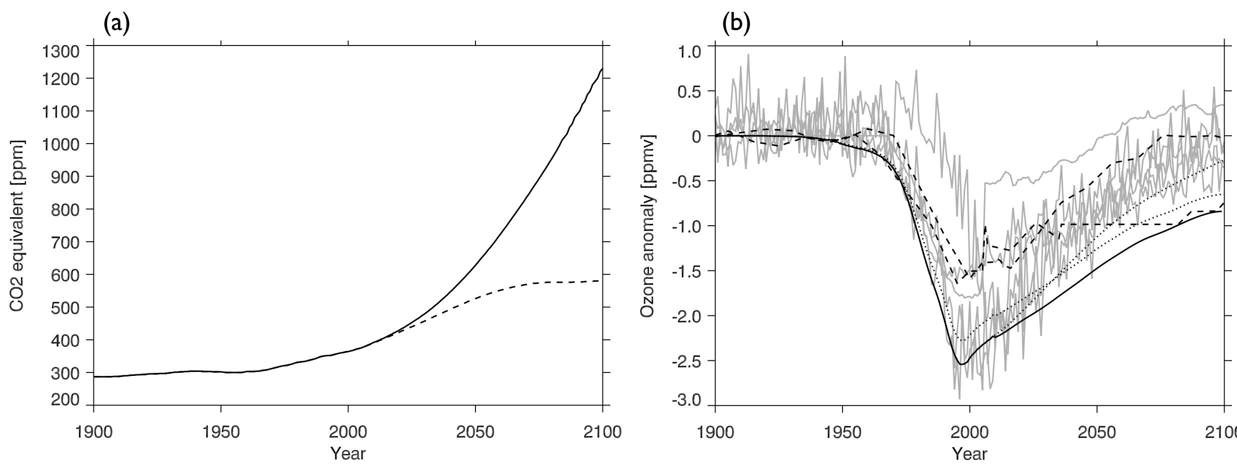

Figure 1: (a): Global-mean annual-mean greenhouse gas concentration (CO_2 equivalent) for RCP4.5 (dashed) and RCP8.5 (solid). (b): Antarctic mean (75-90°S) ozone concentrations at 50 hPa, relative to 1900 values, from Cionni et al. (2011) (black), modified versions of Cionni et al. (2011) (dotted), prescribed ozone from other sources (dashed), and from models with interactive stratospheric chemistry or those using independent chemistry climate models (grey).

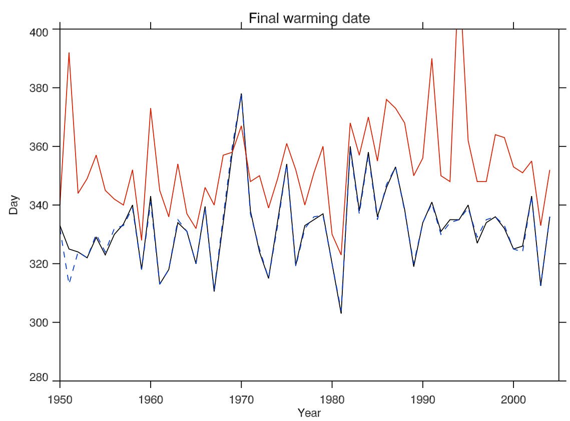

Figure 2: Historical final warming date calculated from CNRM-CM5 data using the Black and McDaniel method (red), the Haigh and Roscoe method using daily data (blue dashed), and the Haigh and Roscoe method from monthly data using interpolation (black).

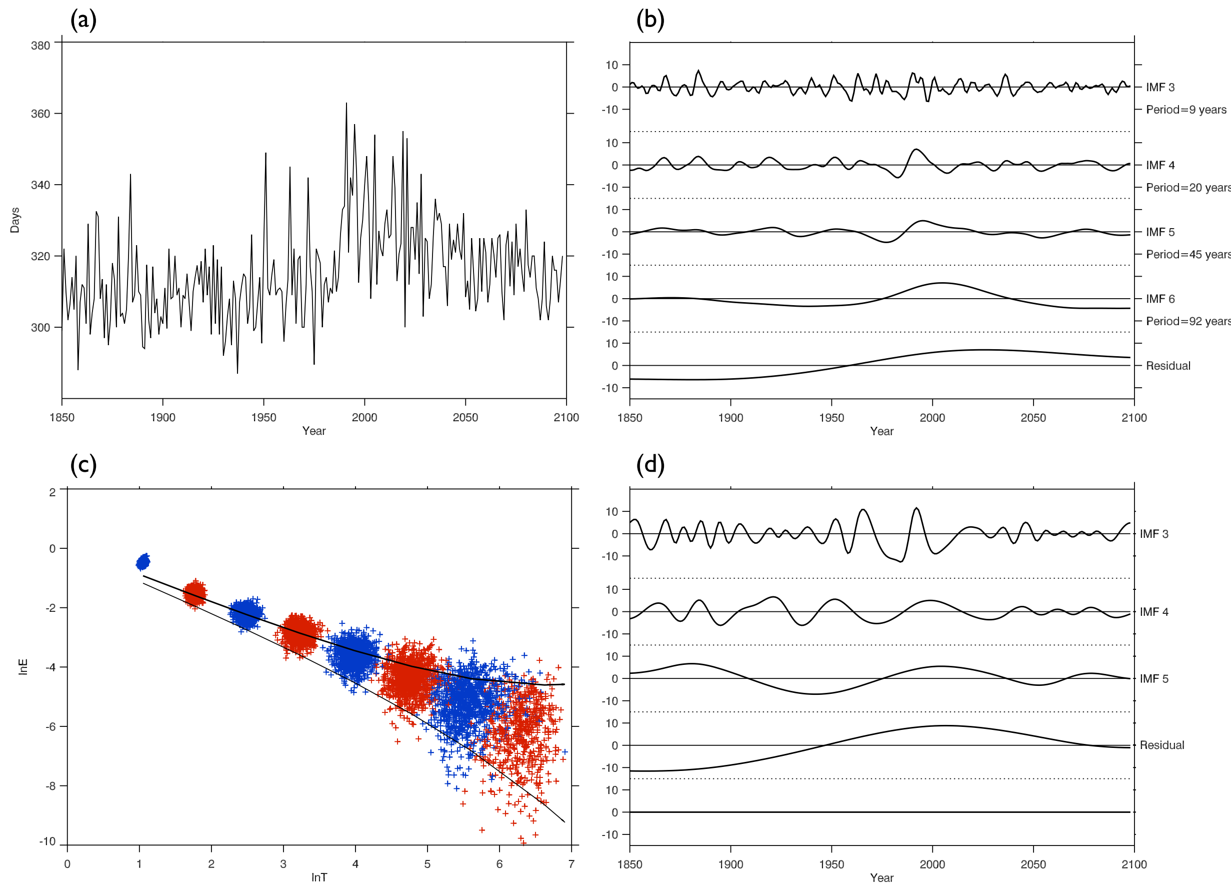

Figure 3: (a): Final warming dates from MIROC-ESM-CHEM for the historical and RCP4.5 experiments, calculated using the Haigh and Roscoe method. (b): The associated high order IMFs from EEMD. (c): The distribution of the energy and period of IMFs from 1000 white noise time series, each containing 1000 data points, and the spread function of the 95% confidence interval. (d): The associated high order IMFs from EMD, showing evidence of mode mixing.

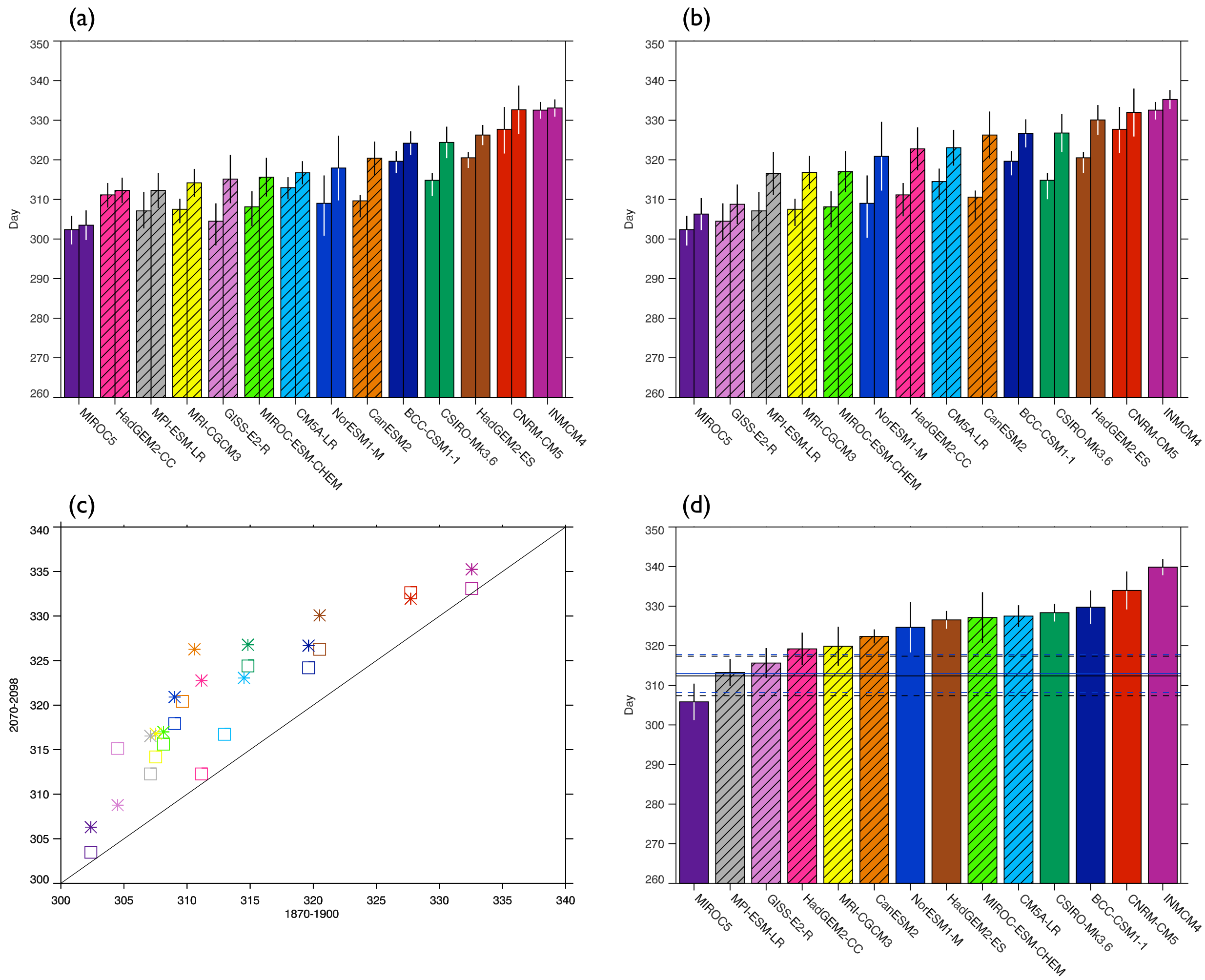

Figure 4: Mean final warming dates for each model for (a):1870-1900 (left bars) and 2070-2098 (right bars) in RCP4.5, (b): 1870-1900 (left bars) and 2070-2098 (right bars) in RCP8.5, (d): 1979-2005. Whiskers show ± 2 standard errors. High-top models are indicated by hatching. In panel (d), the horizontal solid lines show the mean final warming date from ERA-Interim (black) and CFSR (blue), with dashed lines indicating ± 2 standard errors in each case. The relationship between 1870-1900 and 2070-2098 final warming date is shown in panel (c) for RCP4.5 (squares) RCP8.5 (stars).

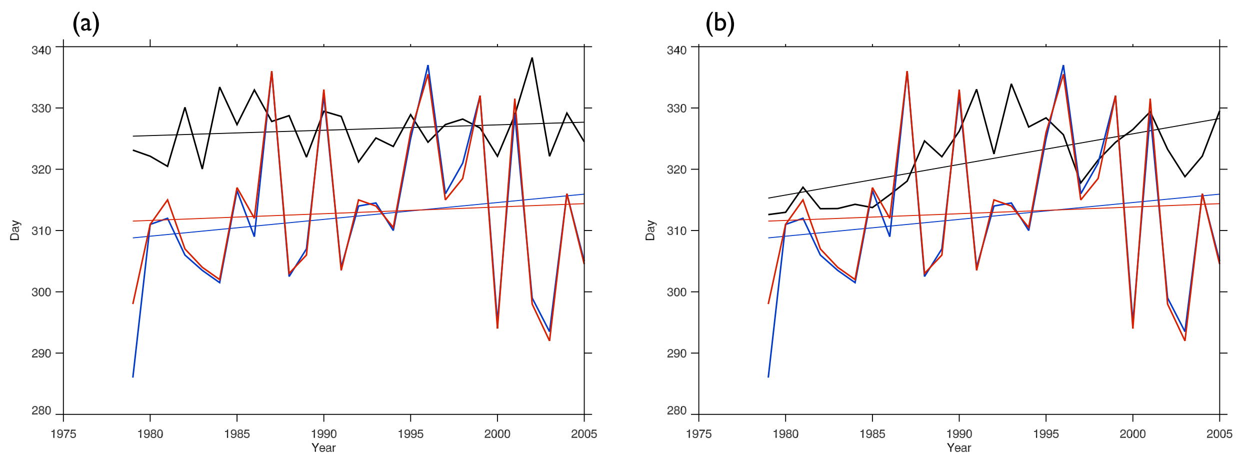

Figure 5: Final warming date from ERA-Interim (blue) and CFSR (red) with (a): the low-top ensemble mean final warming date (black), (b): the high-top ensemble mean final warming date (black).

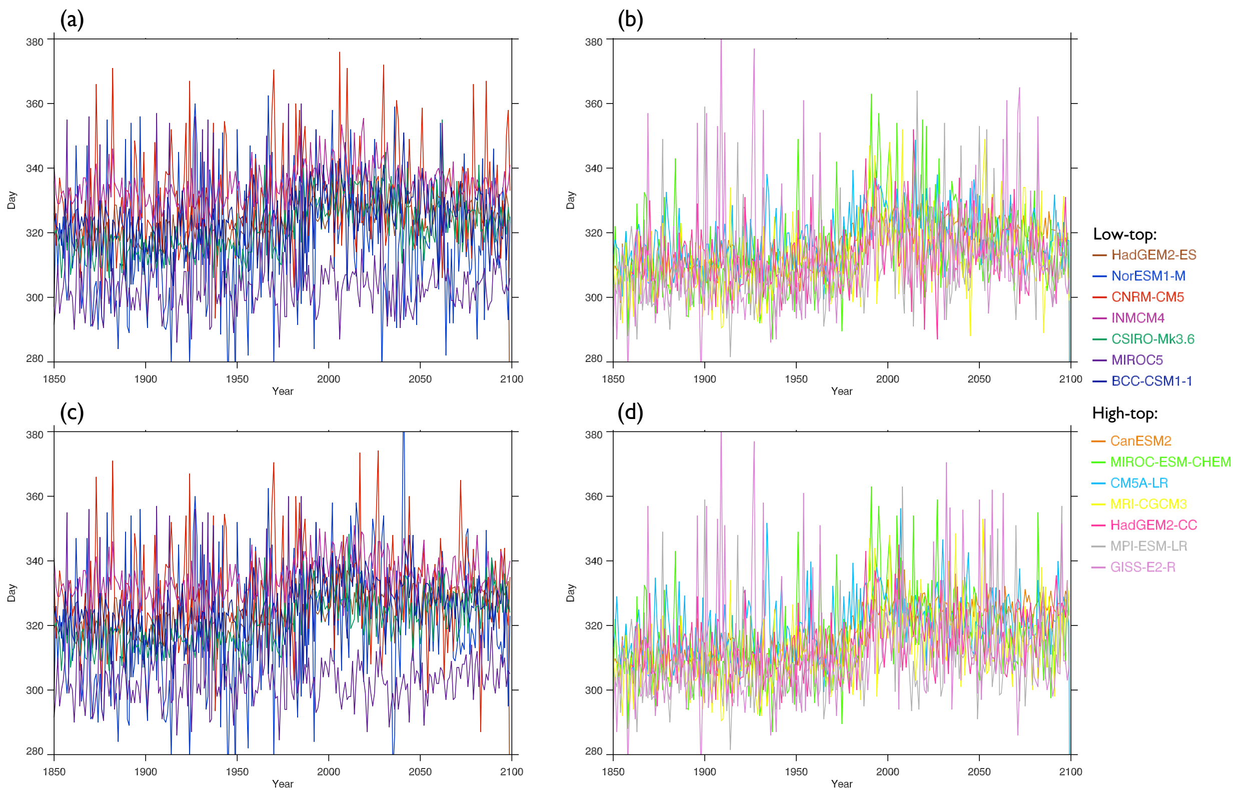

Figure 6: Final warming date for low-top (left column) and high-top (right column) models. (a,b): historical and RCP4.5, (c,d): historical and RCP8.5.

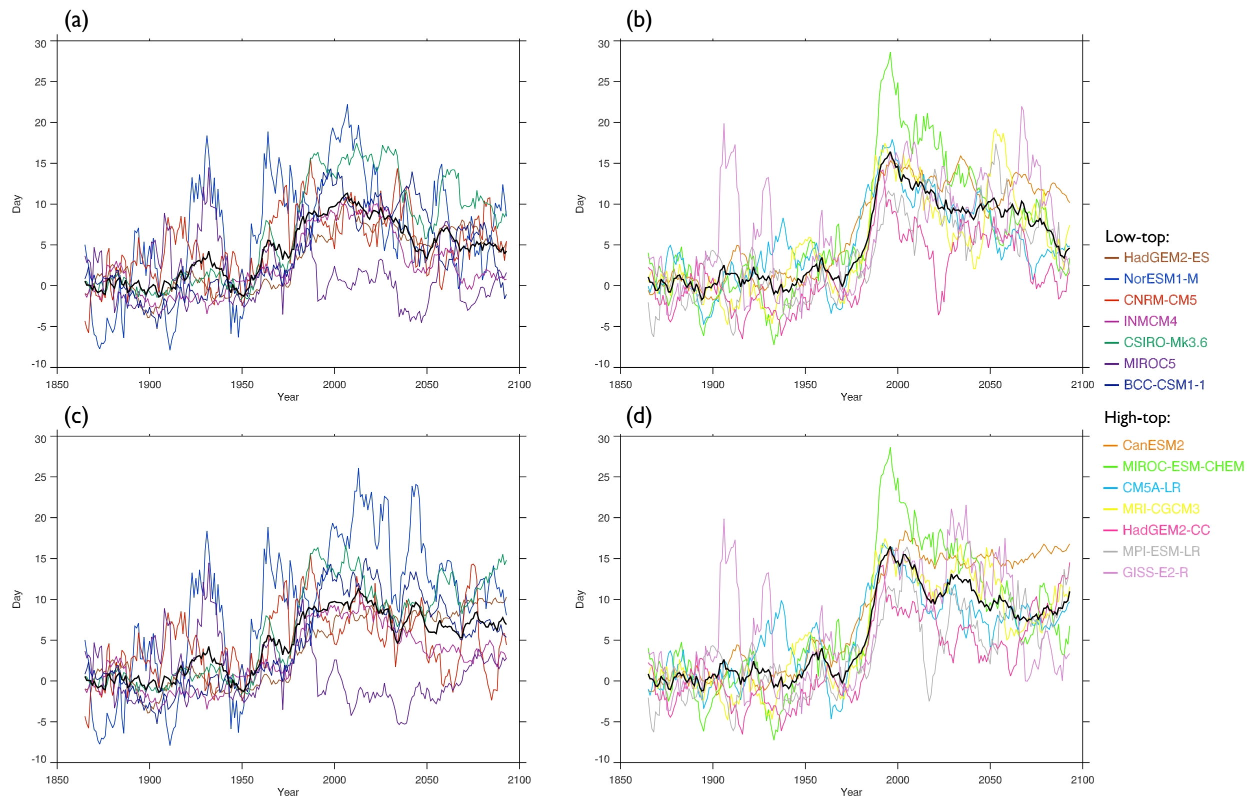

Figure 7: 11-year running mean final warming date for low-top (left column) and high-top (right column) models, with the ensemble mean (thick black line). (a,b): historical and RCP4.5, (c,d): historical and RCP8.5. Raw data is adjusted to the 1860-1900 mean.

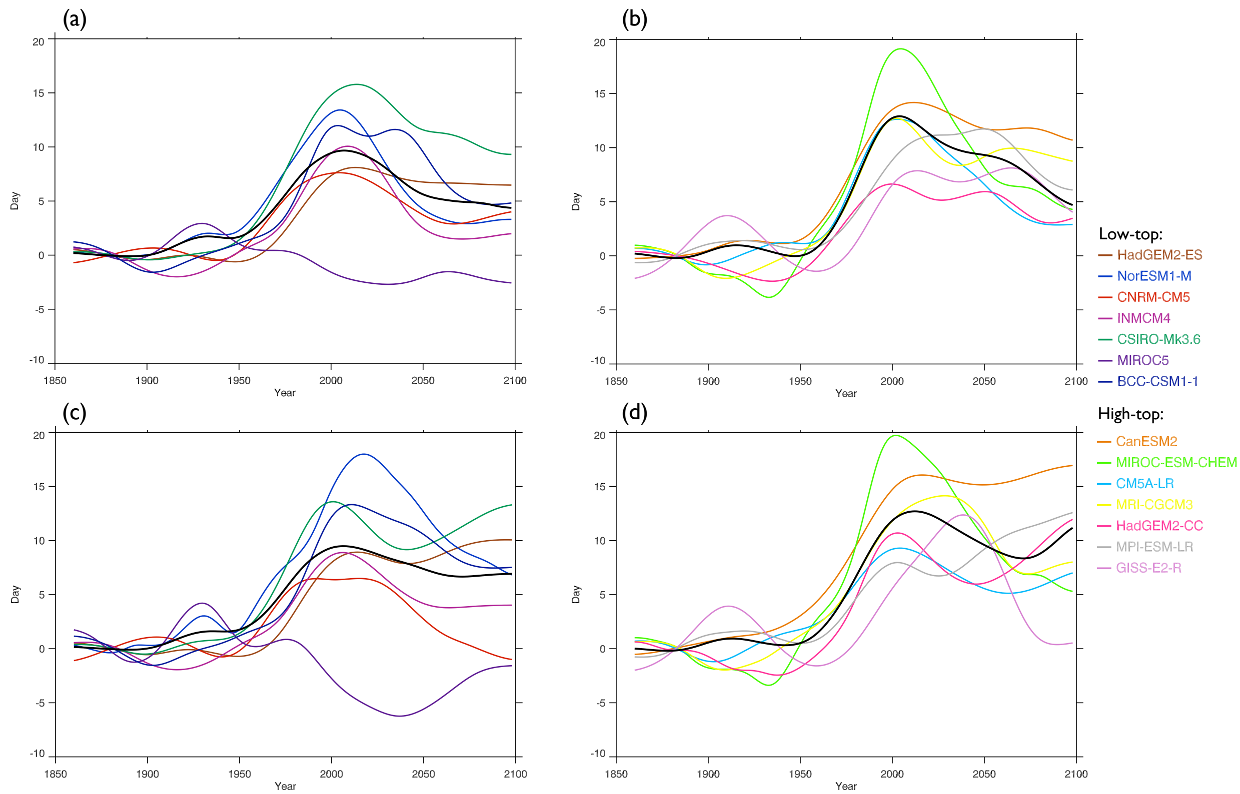

Figure 8: Sum of the residual and last IMF of final warming date for low-top (left column) and high-top (right column) models, with the ensemble mean (thick black line). (a,b): historical and RCP4.5, (c,d): historical and RCP8.5. Raw data is adjusted to the 1860-1900 mean.

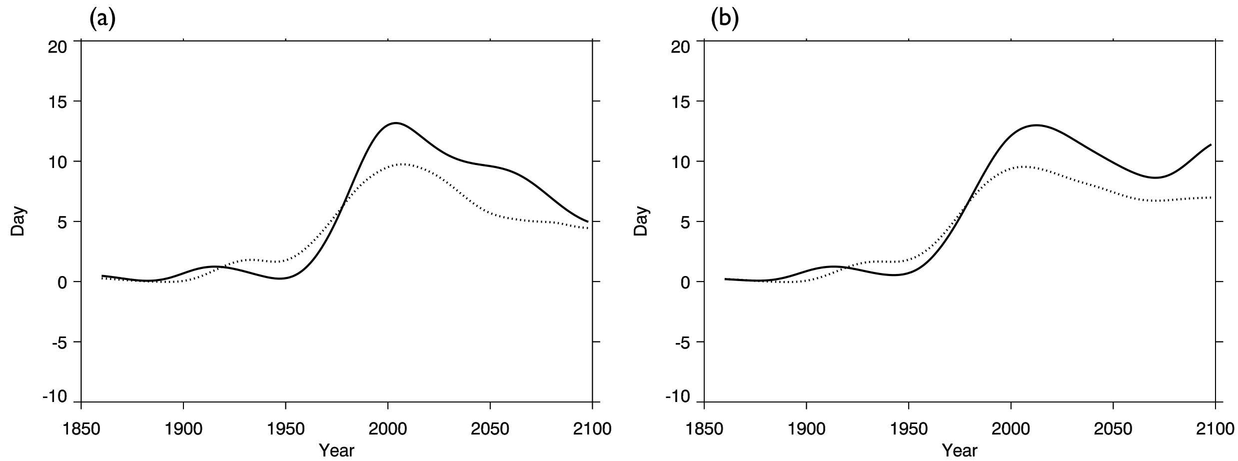

Figure 9: Sum of the significant IMFs of final warming date for low-top (dotted) and high-top (solid) ensemble means. (a): historical and RCP4.5, (b): historical and RCP8.5. Raw data is adjusted to the 1860-1900 mean.

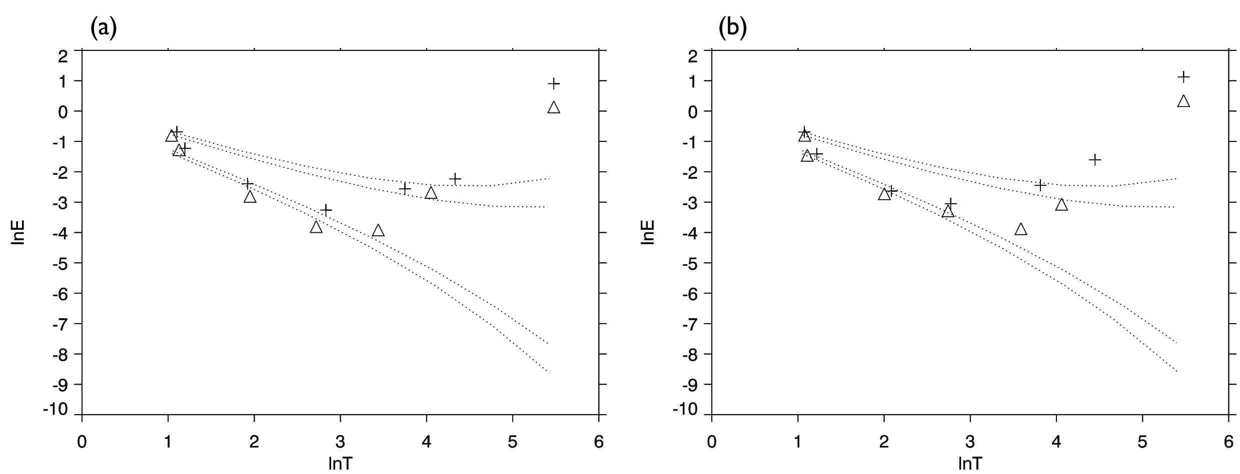

Figure 10: Spread function (dotted lines) and energies of individual IMFs for the low-top (triangles) and high-top (crosses) ensemble means. (a): historical and RCP4.5, (b): historical and RCP8.5. The inner pair of dotted lines show the 95% confidence interval, the outer pair show the 99% confidence interval.