The DOI for this manuscript is doi: 10.1175/JCLI-D-12-00826.1

The final published version of this manuscript will replace the preliminary version at the above DOI once it is available.

If you would like to cite this EOR in a separate work, please use the following full citation:

Schurer, A., G. Hegerl, M. Mann, S. Tett, and S. Phipps, 2013: Separating forced from chaotic climate variability over the past millennium. J. Climate. doi:10.1175/JCLI-D-12-00826.1, in press.

Separating forced from chaotic climate variability over the past millennium

Andrew Schurer[1], Gabriele Hegerl[1], Michael E. Mann[2], Simon F.B. Tett[1], Steven J. Phipps [3]

1 GeoSciences, University of Edinburgh, United Kingdom

2 Dept. of Meteorology & Earth and Environmental Systems Institute, Pennsylvania State University, State College, PA

3 Climate Change Research Centre and ARC Centre of Excellence for Climate System Science, University of New South Wales, Sydney, Australia

Correspondence and request for material should be directed to Andrew Schurer, email: a.schurer@ed.ac.uk

Submitted: 23rd November 2012, resubmitted 13th February, 2013

Abstract

Reconstructions of past climate show notable temperature variability over the past millennium, with relatively warm conditions during the 'Medieval Climate Anomaly' (MCA) and a relatively cold 'Little Ice Age' (LIA). We use multi-model simulations of the past millennium together with a wide range of reconstructions of Northern Hemispheric mean annual temperature to separate climate variability from 850 to 1950CE into components attributable to external forcing and internal climate variability. We find that external forcing contributed significantly to long-term temperature variations irrespective of the proxy reconstruction, particularly from 1400 onwards. Over the MCA alone, however, the effect of forcing is only detectable in about half of the reconstructions considered, and the response to forcing in the models cannot explain the warm conditions around 1000CE seen in some reconstructions. We use the residual from the detection analysis to estimate internal variability independent from climate modelling and find that the recent observed 50-year and 100-year hemispheric temperature trends are substantially larger than any of the internally-generated trends even using the large residuals over the MCA. We find variations in solar output and explosive volcanism to be the main drivers of climate change from 1400-1900, but for the first time we are also able to detect a significant contribution from greenhouse gas variations to the cold conditions during 1600-1800. The proxy reconstructions tend to show a smaller forced response than is simulated by the models. We show that this discrepancy is likely to be, at least partly, associated with the difference in the response to large volcanic eruptions between reconstructions and model simulations.

1. Introduction

Climate variability originates from two fundamentally different mechanisms: (i) changes in the large scale (often global) energy budget of the planet due to influences external to the climate system, and (ii) chaotic interactions within and between climate system components, which generate substantial variability over a broad range of timescales (e.g. Hasselmann, 1976) and which are unrelated to this external forcing. The externally forced component can be sub-divided into that due to anthropogenic forcing (for example, due to changes in land-use and fossil fuel burning greenhouse gases and aerosols) and natural external forcings (such as solar variations and large volcanic eruptions). Changes in greenhouse gases over the last millennium have been strongly influenced by humans since the industrial revolution, while earlier changes, such as the dip over the Little Ice Age, may be at least in part due to Earth System feedbacks (see e.g. Cox and Jones 2008, Frank et al. 2010).

In order to determine the relative importance of each forcing, studies often utilise detection and attribution analysis. This first determines whether an externally forced signal can be detected in observations, given our understanding of the expected response to the forcing and internal variability, and then attempts to attribute the observed response to a particular combination of individual forcings (see Hegerl et al. 2007b for a review). Hence, detection and attribution studies require reliable estimates of internal climate variability.

Much of our understanding of the climate system originates from observations during the 20th century, a period covered by high quality instrumental data (see Trenberth et al. 2007 for a review). However, it is difficult to estimate internal climate variability from the 20th century record alone, as this period is too short to obtain well-sampled estimates of variability on multi-decadal timescales. In addition, climate over the 20th

century experienced substantial anthropogenic radiative forcing, which has to be accounted for in order to derive estimates of climate variability.

Consequently climate models are usually used to determine the characteristics of internal variability and its possible contribution to the recent warming, with the model-dependence of this estimate understood as a source of uncertainty (see e.g. Hegerl and Zwiers 2011). Reconstructions of temperature over the last millennium can provide alternative estimates of internal variability. While such estimates are prone to uncertainties (see Jansen et al. 2007, Jones et al. 2009), they nevertheless provide valuable information on the role of internal climate variability on interdecadal and longer timescales. However, to obtain these estimates we first need to separate internal variability from the externally forced component of change over the last millennium. This paper attempts to do that.

Our knowledge about the climate of the past millennium originates from two main sources: proxy reconstructions and climate modelling. Reconstructions attempt to determine past climate variability by combining information from a number of different proxies, such as tree-rings widths and/or tree-ring densities, corals, documentary evidence, ice cores, speleothems, boreholes and sedimentary deposits (see e.g. Jones 2009 for a review). Climate modelling, in contrast, aims to simulate past climate variability based on our understanding of the underlying physics. The models are driven by reconstructions of climate forcings, such as volcanic eruptions, fluctuations in solar irradiance, orbital changes, variations in CO_2, sulphate aerosols and land-use changes (see e.g. Schmidt et al. 2011, 2012 and Forster et al. 2007). Both the forcing histories and the response of the models to the forcing are sources of uncertainty. This uncertainty implies that model-based estimates of the forced component present in proxy reconstructions are incomplete, which in turn implies uncertainty in estimates of internal variability derived by removing these estimated forced components from actual reconstructions. Nevertheless, these empirically-derived estimates can provide a valuable cross-check against purely model–based estimates of internal climate variability.

Previous analyses that aimed at separating forced and internal variability over the past millennium have typically used only a limited number of climate reconstructions, few, often simple, climate models (e.g. Hegerl et al. 2007a, Weber 2005), and a very limited sample of internal climate variability. Many new reconstructions of temperature variability over the past millennium have recently become available. These reconstructions make use of an expanding body of proxy evidence in combination with improved statistical techniques aiming to better preserve variance (Ammann and Wahl 2007, Juckes et al. 2007, Mann et al. 2008; 2009, Moberg et al. 2005, D'Arrigo et al. 2006, Frank et al. 2007, Christiansen and Ljungqvist 2011, Hegerl et al. 2007a), and more thorough exploration of the sensitivity of reconstructions to the choice of proxy data and the reconstruction methods. This includes additional studies that test reconstruction methods using model output (see, for example, Hegerl et al. 2007a, Mann et al. 2007, Jones et al 2009, Smerdon 2012).

In addition, a relatively large number of simulations with fully coupled GCMs have recently been completed for the whole of the last millennium (section 3). These were predominantly performed as part of the Fifth Coupled Model Intercomparison Project (CMIP5; see Taylor et al. 2012) and Third Paleoclimate Modelling Intercomparison Project (PMIP3; Braconnot et al. 2012). Here we make use of these new model simulations and the newly expanded range of proxy reconstructions to improve our knowledge of natural variability and its potential implications for detection and attribution studies.

The reconstructions used in this paper are introduced in Section 2 and the model simulations are described in Section 3. Section 4 presents results aimed at calculating the relative importance of external forcing over the past millennium. This is done by first examining the variance explained by the forced component in the reconstructions. Then a detection and attribution analysis is carried out, followed by a discussion of results and their implication for studies of recent climate change. The relative importance of the various external forcings is analysed in Section 5, followed by a summary (Section 6).

2. Reconstructions

A list of the reconstructions used in this paper is given in table 1. These reconstructions were calibrated to three different geographical regions: 0-90°N land and sea (Ammann and Wahl 2007, Juckes et al. 2007, Mann et al. 2009, Moberg et al. 2005), 20-90°N land only (D'Arrigo et al. 2006, Frank et al. 2007) and 30-90°N land only (Christiansen and Ljungqvist 2011, Hegerl et al. 2007a). Some reconstructions are based on a fixed number of sites (Christiansen and Ljungqvist 2011, Hegerl et al. 2007a; although the sampling within sites may decline back in time), and some are based on varying numbers of proxy sites over time (e.g., D'Arrigo et al. 2006, Frank et al. 2007, Mann et al., 2009). Hence it is expected that uncertainties will increase further back in time. Some reconstructions are based on averaging across the available sites and then calibrating to the target of the reconstruction (e.g., D'Arrigo et al., 2006, Hegerl et al. 2007a; in some cases, calibrating high and low frequency bands separately e.g. Moberg et al 2005), while others are based on reconstructing the underlying spatial patterns using multilinear regression techniques (Mann et al., 2009; Ammann and Wahl 2007). Overall, the large number of reconstructions available, based on a mix of data and methods, provides a reasonable estimate of uncertainty due to varying methodological assumptions and choices of data.

The reconstructions are shown in figure 1 and generally show a warmer period around the start of the millennium from around 900-1200 (the Medieval Climate Anomaly, MCA), followed by a cooler period from around 1450-1800 (the Little Ice Age, LIA). They also show relatively abrupt periods of cooling associated with volcanic eruptions (e.g. following the eruption of Mount Tambora in 1815). Figure 1 shows the HadCRUT4 instrumental data (Morice et al. 2012) from 1850-2000 as well. All reconstructions, except Christiansen and Ljungqvist (2011), show similar trends to the HadCRUT4 data over the instrumental period. Whereas all the other reconstructions scale the proxy record in some way to the instrumental data, the Christiansen and Ljungqvist reconstruction represents an un-weighted average of a number of different proxies scaled locally. In order to ensure consistency during the modern interval with the instrumental record over the region sampled (extratropical NH land), we have rescaled that reconstruction using an inverse regression onto the instrumental temperature series (note that the inverse regression assumes that instrumental error and noise is negligible relative to that for the proxy reconstruction; see Christiansen and Ljungqvist, 2011; Hegerl et al., 2007a). Results for both the scaled and un-scaled Christiansen and Ljungqvist reconstruction will be shown throughout the paper.

We first smooth all annual reconstructions and model simulations using a 10-year Butterworth filter (see Mann; 2008, also used in Mann et al. 2009), reducing power by a half on 10 year timescales. This ensures that both simulations and data are comparable, and that the analysis focuses on the better reconstructed inter-decadal variability (see e.g. Frank et al 2007, D'Arrigo et al 2006). In our standard analysis, this is followed by an 11-yr boxcar filter in order to focus on truly interdecadal timescales. In order to determine the sensitivity to the smoothing length, our analysis has been repeated both without the additional smoothing, and using a 21-yr boxcar filter instead of an 11-yr boxcar. This tests the sensitivity to focusing the analysis on multi-decadal rather than interdecadal timescales (which e.g. Christiansen and Ljungqvist 2011 argued is more faithfully reconstructed). Results in an earlier paper (Hegerl et al., 2006) showed that calibration of a treering based reconstructions on interdecadal timescales yielded similar estimates of climate sensitivity compared to one using a multi-decadally filtered version of the same reconstruction (Cook et al., 2004), supporting the approach taken here. Extensive sensitivity tests in earlier papers (Hegerl et al., 2003; 2006; 2007) showed little sensitivity of detection results to the shape and length of the filter between the limits of 5 years (where the signal-to-noise ratios of forced vs. internal variability become increasingly low) and secular timescales (at which it becomes increasingly difficult to distinguish the effects of different external forcings). The sensitivity of our results to the choice of smoothing length is discussed later in the paper.

3. Model simulations

Table 2 contains details of all the climate model simulations which are used in the multi-model mean fingerprint used in this paper (CCSM4 – Landrum et al. 2012; MPI-ECHAM5 – Jungclaus et al. 2010; MPI-ESM-P - Giorgetta, et al., 2012; HadCM3 – Pope et al. 2000, Gordon et al. 2000; GISS-E2-R – Schmidt et al. 2006; Bcc-csm-1-1 – Wu 2012) and one additional model (CSIRO – Phipps et al. 2011, 2012) whose results contributed to the calculation of the individually forced fingerprints. The surface air temperatures (SATs) of the different models are shown in figure 2. All the model simulations are smoothed the same way as the reconstructions and are calculated as the mean over the three different geographical regions represented by the different reconstructions. Only results for 0-90°N land + sea are shown in figure 2. The GISS-E2-R simulations (figure 2a) included a significant initial model drift that was removed from the control simulation by fitting a second order polynomial to the control simulation (this is the same correction technique as applied in Tett et al. 2007).

The forcings used in the model simulations are listed in table 2. Where two forcings are given in the solar forcing column, the simulations have been driven with a combination of two solar forcings that have been spliced together, following the guidance given by Schmidt et al. (2011,2012). For the CCSM4 model and GISS-E2-R models, the land use forcing has been merged into the Hurtt et al. (2009) land-use dataset after 1850, following Schmidt et al. (2011,2012).

For the period 1850-2000 other anthropogenic forcings have been included. The CCSM4, GISS-E2-R, MPI-ESM-P and the Bcc-csm-1-1 model simulations used the CMIP5 anthropogenic historical forcings. The HadCM3 simulation followed the forcings used in Tett et al. (2007), while the MPI-ECHAM5 model simulation has been driven with aerosol concentrations following Lefohn et al. (1999; see Jungclaus et al., 2010). These differences in the treatment of the anthropogenic forcings likely explain the discrepancies in the 20th century trends seen in figure 2a.

The natural forcing datasets used by these studies are uncertain (see Schmidt et al 2011,2012). There is uncertainty in the amplitude of solar forcing (see e.g. the difference between Steinhilber et al. 2009 and Shapiro et al. 2011) and the forcing by individual volcanic eruptions (see e.g. the difference between Crowley et al 2008 and Gao et al 2008). There is also the possibility of systematic bias owing to the scaling between the observed sulphate spikes found in ice cores and the aerosol optical depth used by the models (see Hegerl et al 2006), although the different volcanic reconstructions span a range of assumptions. The level of land-use change in pre-industrial times is also debated (see Pongratz et al 2009 and Kaplan et al 2010).

There are also known model limitations in the response to these forcings. For example, it is likely that the models described in this paper may not be capable of fully capturing the dynamic response to solar forcing that has been proposed by several studies and involves an amplification of the response by ozone feedback within the stratosphere (see e.g. Shindell et al 2006, and a review by Gray et al 2011). Many of the models used here do not have a fully resolved stratosphere and contain no interactive ozone chemistry. Such dynamic responses would, however, affect the hemispheric annual mean response studied here less than regional and seasonal responses. There is also evidence that the models may not be capturing the dynamic response to volcanic forcing (see e.g. Driscoll et al. 2012), while some may be responding too strongly (see e.g. Gent et al 2011). There is therefore still considerable uncertainty in the model simulated response to climate forcing over the past millennium.

The simulations driven with all forcings are shown in figure 2a and show similar features to the reconstructions (for a comparison see figure 2b): The model simulations are slightly warmer in the MCA, although the timing of the warming is different in models and reconstructions (see Jungclaus et al. 2010). The simulations are also substantially colder than the millennial average for much of the LIA, and all show a strong increase in temperature over the 20th century.

Perhaps the most prominent features of the simulations are the pronounced cooling episodes following large volcanic eruptions (the largest of which are highlighted by grey bars in figures 2a and 2b), in particular those in 1258 (origin unknown), mid-1450s (Kuwae) and 1815 (Mount Tambora). Note there may be uncertainty in the dating of some volcanic eruptions, particularly Kuwae (Plummer et al. 2012). The volcanic cooling simulated for these large eruptions appears far larger than that seen in the reconstructions (see figure 2b). This is particularly true for the 1258 eruption, which causes a large cooling in the simulations that is hardly seen in the reconstructions. This discrepancy is further explored later in this paper.

Figure 2c shows the results from composite simulations, which include the effect of solar and volcanic forcing only. For the HadCM3 and MPI-ECHAM5 models this is the linear combination of simulations forced by volcanic and solar forcing only. For the CSIRO model this is calculated by subtracting simulations with just orbital and greenhouse gas forcings from simulations including orbital, greenhouse gas, solar and volcanic forcings. In all of these simulations the solar forcing is weak, so that the combined fingerprint is dominated by volcanic forcings. The behaviour of the combined simulations in figure 2c is similar to the all-forced simulations shown in figure 2a for the pre-industrial periods, with a correlation of +0.87 for the period 1401-1900. This suggests that, in the model world at least, these are the most important pre-industrial forcings. The composite simulations diverge from the all-forced simulations significantly from 1850 onwards as anthropogenic forcings become increasingly important.

Figure 2d shows the results from simulations forced by well-mixed greenhouse gases only. For the CSIRO model these results were calculated by subtracting simulations including just the orbital forcing from simulations with orbital and greenhouse gas forcing. The effect of the greenhouse gas forcing is clearly visible, causing a steady increase of temperature beginning around 1800. In addition to this recent warming there are also pre-industrial long-term variations in greenhouse gas only simulations, with a noticeable cooling around 1600 in response to a small dip in the abundance of CO_2 (see discussion below).

Each of the models which provided forced simulations also has an equivalent unforced control simulation of varying length (not shown). These were used to construct the internal variability samples required for the detection and attribution analysis discussed in section 4b.

4. Results: The role of external forcing

a) Explained variance

Before analysing the entire millennium or substantial parts of it in a detection and attribution analysis, changes in the role and importance of forcing are explored over 200-yr windows. This serves to test for variation in the role of external forcing vs. internal variability over the millennium in model simulations, and addresses the extent to which these variations are reflected in reconstructions.

We define the explained variance as the squared correlation between the model simulations and individual reconstructions. For this test the period encompassing the large 1258 eruption was ignored, since the large discrepancy in response to this eruption between the simulations and reconstructions (see figure 2b) is likely to dominate our results early in the millennium. Where a correlation is negative, the explained variance is set to zero, as only positive correlations are meaningful measures of the correspondence between simulations and reconstructions.

The explained variance for 200 year periods is shown in figure 3, where each coloured symbol represents the variance within a reconstruction explained by the multi-model mean. The average of all the variances calculated for each reconstruction is also shown. While there is substantial variation between reconstructions, some common features emerge: the largest explained variances are found over the most recent 200 years (1750-1950) with average values over 60% (also see Stott et al. 2000). The explained variance then decreases to an average of about 30% for 200 year periods between 1400 and 1900. Before 1300 the explained variances begin to decline further and for the periods 900-1100, 950-1150 and 1000-1200 the explained variance is negligible (note that this is robust with respect to the exclusion of the 1258 eruption). Could this decline be due to a decreasing role of external forcings back in time?

To address this question, we performed a “perfect model” test. In this analysis the explained variance was calculated from the correlations between individual model simulations and a fingerprint derived from all the other simulations, which are then averaged. If the models have a similar level of internal variability to the observations, and if the simulated response to external forcing is accurate, then this perfect model correlation should be similar to that between the multimodel mean and the reconstructions. Errors in the external forcing used in the model simulations, errors in the model physics and errors and additional noise in the reconstructions will reduce the explained variance relative to the average explained variance obtained from the simulations. Therefore, we expect the average of the explained variance obtained from the simulations to yield an approximate upper limit of explained variance given varying forcing levels over time (see dashed line in figure 3). If there was a strong divergence between the perfect model result and the explained variance in reconstructions, this would suggest an increasing role of data uncertainty, or that the true forcing uncertainty is larger than that represented in the forcings used in the simulations, or systematic biases in the responses of the models to external forcings, or any combination of these.

As expected the highest explained variance for the perfect model test is found for the most recent 200 year period, since this is when multi-decadal forcing is strongest due to anthropogenic activity. As seen in the results for the multi-model mean vs. reconstruction comparison the explained variance in the perfect model study also decreases back in time. However, if the 1258 eruption is not removed, the variance decreases by a smaller amount due to the presence of a strong cooling event common to all the model simulations (not shown). The perfect model explained variance remains within the range of results from the reconstructions from about 1200. This shows that a decrease in the importance of external forcing relative to internal variability can explain much of the observed decrease in explained variance in the simulation-reconstruction comparison.

A striking result of this perfect model study is that the explained variance during the MCA is quite low even in the perfect model study, of order 20%, suggesting that given the forcings used this period should be dominated by internal variability rather than strongly forced (with the exception of the enigmatic 1258 eruption, which was excluded). This is possibly due to the substantially reduced volcanic activity (other than the 1258 eruption) during this period. Therefore values of the explained variance as low as 20% are expected. However, the correlations between the models and the reconstructions for this period are substantially lower than the perfect model values. As discussed in Section 2, increased sampling error in the reconstructions (e.g. due to decreasing availability of proxy data) could be partly responsible for the reduction in correlations with the model simulations (e.g. Frank et al 2007 and D'Arrigo et al. 2006 caution overuse of their reconstructions prior to

1200 and 1117 respectively, see also Esper and Frank 2009). Unusually pronounced internal variability during this period may also account for the reduction in explained variance (see Goosse et al. 2012). Of course, the observed discrepancies may result from some combination of these factors.

b) **Detection and attribution analysis**

The previous results show that there is agreement between the model simulations and the reconstructions, particularly for time periods after 1200, demonstrating at least some role for external forcing in the climate of the past millennium over most 200-yr segments. Here we use detection and attribution techniques to estimate the magnitude of the forced change, separating the climate response into forced and internal variability.

The multi-model mean response, smoothed in order to focus on multi-decadal frequencies, provides a fingerprint for forced variability in the reconstructions. The contribution by the fingerprint of external forcings to reconstructed NH temperature has been estimated using a total least squares (TLS) detection and attribution technique (see Allen and Stott 2003 for details) which estimates a scaling factor β to best match the time dependent fingerprint $X_i(t)$ to the reconstructions, Y(t).

$$Y(t) = \sum_{i=1}^{m} \big(X_i(t) - v_i(t)\big)\beta_i + v_0(t) \quad (1)$$

The fingerprint of external forcing, X_i, is provided by the mean of an ensemble of climate model simulations (averaged over the same region as the reconstruction), and represents the time-fingerprint of NH mean temperature in vector form. As only a limited ensemble of forced simulations is available, each fingerprint $X_i(t)$ will still contain internal variability generated within the simulation $v_i(t)$, whose variance is reduced by averaging over the ensemble. The reconstruction is assumed to have an associated internal variability v_0. The method assumes a ratio of noise variance between reconstructions and that in the fingerprint, which we set to 1/n with n equal to 11, the number of ensemble members.

The scaling factors β_i are determined following Allen and Stott 2003 by calculating the singular value decomposition (SVD) of Z:

$$Z = U\Lambda V^T \tag{2}$$

Where Z is a matrix formed by combining the model fingerprints and reconstructions (after scaling to equal noise variance; see Allen and Stott 2003):

$$Z = [X, Y] \tag{3}$$

We can estimate the true underlying response to forcing represented in the model simulations and reconstructions, ž, where ž is calculated following equation 38 in Allen and Stott (2003):

$$\check{z}Z - Z\tilde{v}\tilde{v}^T \tag{4}$$

and $\tilde{v}$ is taken from the SVD.

The uncertainty in scaling factors can be approximated analytically (see Allen and Stott 2003), but is here calculated by superimposing 2000 random samples of internal variability taken from the control simulations onto both the noise reduced observations and model fingerprints, ž. To construct these model based samples of internal variability we use segments of control simulations of the same length as the analysis period, taken from the same models as are used to form the model-mean fingerprint. The 5-95% uncertainty range for β is based on the sampling distribution derived from these multiple samples, and should be a credible range over which to quantify uncertainty given the 12,000 years of control simulation used (see e.g. Hegerl et al. 1996; Allen and Tett 1999).

A fingerprint is detected in the reconstructions if the scaling factor β is significantly larger than zero. This means that the effect of external forcing is detected at the 5% confidence level in the reconstructions if the calculated 5-95% scaling range does not encompass zero.

To validate the consistency of the fit, the residuals of the regression were checked against the estimates of model-based internal variability. If a fit to a reconstruction yields a residual with a χ-squared value (Allen and Stott eq. 26) that is smaller than the sum-of-squares of 90% of the control samples then the amplitude of v_0 is said to be consistent with the internal variability as sampled by the control simulations.

The detection and attribution analysis was performed for several different time periods: a recent time period where reconstructions are based on a larger database (1401-1950; see Jansen et al., 2007), a short period encompassing the MCA (851-1400), the full time period including the MCA (851-1950), and the corresponding pre-industrial time periods (851-1850 and 1401-1850). The results for the full time period for three representative reconstructions are shown in figure 4a. Figure 4b shows detection results for all time periods and all 8 reconstructions plus the re-scaled version of the Christiansen and Ljungqvist (2011) reconstruction. The results show that the fingerprint for external forcing is detectable in all reconstructions for four of the time periods to a 5% significance level. While both the reconstructions of external forcing and of temperature are uncertain, the uncertainties between the two should be independent from each other, making spurious detection of the fingerprint highly unlikely. Thus, our results confirm a clear and important role of external forcing during the last millennium, even when the last 150 years are excluded. For the time period 851-1400, however, the external forcing is only detected in half of the reconstructions. This is perhaps unsurprising given the poor correlations found in the previous section over this period, and is at least partly due to the smaller role of forcing as estimated by the perfect model correlations.

Scaling factors were also calculated using an ordinary least squares (OLS) fit (see Allen and Tett 1999). The results are very similar, with the fingerprint for external forcing detectable in all reconstructions for all time periods except for the period 851-1400 (results not shown). OLS analysis places all the internal variability onto the reconstructions, so this is the limiting case where error and internal variability in the fingerprint is negligible relative to that in the reconstructions. This analysis shows that the results are insensitive to the assumed ratio of internal variability between the reconstructions and models.

Figure 4b shows the internal variability samples calculated as part of the TLS analysis, v_0. As can be seen from this figure, as well as the bars shown in figure 4c, the residual variability of several of the reconstruction derived samples are not consistent with the model's internal variability. This is especially true for the time periods containing the MCA. To test whether this potentially larger variability in certain reconstructions exerts any leverage on our detection results, the variance of the samples of internal variability taken from the control simulations used to calculate the range of scaling factors was scaled to fit the variance of the TLS generated sample of internal variability, if the latter was larger prior to repeating the detection analysis. As figure 4c shows the external forcing is still detectible testing against this inflated variability in all but one reconstruction (excluding the period 851-1400) and even for that reconstruction for all but one of four time periods.

Questions have been raised about the faithfulness of the low-frequency climate signal recorded by climate proxy data (see e.g. Jones et al. 2009 for a review). One recent study (Esper et al. 2012) for example argues, based on a comparison of tree-ring width and tree-ring density record estimates from one location in the Arctic, that tree ring width records, and therefore potentially any reconstructions using them, may underestimate millennial-scale trends such as those associated with orbital forcing. Whether or not this effect actually impacts hemispheric temperature reconstructions, which reflect a mix of proxy data and sample diverse seasonal windows and latitudinal ranges, is less clear. If a long term trend, such as that suggested by Esper et al 2012 was missing in any of the reconstructions studied here, it should lead to a positive trend in the residuals shown in figure 4b. We find, however, that for all of the multiproxy reconstructions, the residuals exhibit a negative long-term trend (ranging from -0.23 ºC/1000yr in Mann et al. 2009 to -0.03 Juckes et al. 2007 ºC/1000yr), suggesting if anything an overestimation of any potential long-term cooling trend. Interestingly the two tree-ring only reconstructions (D'Arrigo et al. 2006 and Frank et al. 2007) do exhibit a positive long-term trend, and quite a substantial one in the case of the Frank et al. 2007 reconstruction (0.17 ºC/1000yr), that is consistent with the potential bias noted by Esper et al. 2012. Attributing some of this trend, to orbital forcing is difficult, however, due to the large uncertainties in the reconstructions themselves, and given that internal climate variability (e.g. at the time of the MCA; residuals shown in figure 4b) projects onto the trend.

Figure 5 shows scatter plots for the externally forced model fingerprints plotted against the reconstructions (based on the decadally smoothed data used for the regression) and the regression lines calculated in the above analysis. This plot further highlights differences in the estimated amplitude of the forced response for different reconstructions. Several of the reconstructions have periods during the LIA that are clearly colder in the reconstructions than in the models; equally, there are several reconstructions that have periods of the MCA which are significantly warmer in the reconstructions than in the model simulations. Neither of these features is present for every reconstruction, however, indicating that there is substantial uncertainty in the level to which the MCA and LIA can be reproduced due to external forcing (see also figure 4b). Also present in many of the regressions, particularly those for the period 851-1950, are tails where the models are far cooler than the reconstructions. These tails result from volcanic cooling and highlight that the reconstructions tend to exhibit considerably less of a cooling response to the largest volcanic eruptions than is simulated by the models (figure 2b).

c) **Possible explanations for the model data mismatch in amplitude**

A striking result in figure 4b is that the multi-model fingerprint appears to have too strong a response when compared to the reconstructions, as indicated by many scaling factors being significantly less than unity. A scaling factor less than unity means that the model response needs to be reduced in amplitude to match those reconstructions. We first check the dependence of this effect upon the degree of smoothing that the model simulations and reconstructions undergo prior to the analysis. Results for when no additional smoothing is added (on top of the decadal Butterworth filter) are shown in figure 6a and results when an additional 21-year box-car filter is used (rather than the normal smoothing length of 11 years) are shown in figure 6b. These figures show that there is some dependence on the calculated scaling values with smoothing length. With less smoothing the scaling ranges are less consistent with unity, indicating a larger discrepancy in response to forcings in the model simulations compared to the reconstructions. When the smoothing is increased, however, to focus on lower frequency responses, the model response becomes more consistent with the reconstructions. Many reconstructions now yield scaling factors that are consistent with unity, at least for the more reliable more recent periods. It is worth noting that although the values of β may be sensitive to the smoothing length, the detectability of the external forcing is not.

It is also possible that the modelled response to volcanism is systematically too large. Comparisons between simulated and observed 20th century records suggest a stronger simulated response than that of the observations (see Hegerl et al., 2007b); however, the observations are within the uncertainty range, and the cooling response may have been masked by substantial El Niño events closely following several large eruptions. As the uncertainties in reconstructed forcing and model response are larger prior to the 20th century, the possibility of an excessively large model response can neither be ruled out nor confirmed based on present data However, if the response of the multi-model mean to every forcing was systematically too large, then the observed response should be smaller than the model response regardless of the choice of smoothing length. This is not the case (figure 6c).

As our previous analyses indicate, this problem seems to be linked to the high frequency response. Volcanic eruptions play a substantial role over much of the last millennium (see Hegerl et al. 2003, 2007a, 2007b, Miller et al. 2012, Weber 2005) and show the strongest response on short timescales. A visual inspection reveals that some of these seem to be excessively large in the fingerprint relative to the reconstructions. This forcing has large short term effects, therefore the low scaling factors observed in the high frequency response could plausibly result from discrepancies between the simulated and observed responses to volcanic forcing, rather than a systematic error in the model response. One possible factor may be errors in the volcanic forcing history. For example, Hegerl et al. (2006, SI) estimated a total uncertainty in the magnitude of the overall volcanic forcing timeseries of ~35% due to uncertainty relating to the scaling of sulphate measurements in ice cores to the aerosol forcing. This would therefore indicate that a scaling factor as small as about 0.7 might not be inconsistent with the data given forcing uncertainties, which would yield a multi-model mean response consistent with many more of the reconstructions, at least over the best reconstructed periods (figure 4). It is further possible that inaccuracies in the implementation or response to the volcanic forcing could play a role (see e.g. Driscoll et al 2012, Gent et al 2011, Timmreck et al 2012), especially for larger eruptions such as the 1258 eruption because of the coagulation of sulphate aerosol particles (Timmreck et al. 2009). On the other hand, Mann et al 2012a showed that a reconstruction displayed less cooling than energy balance models even when forced using the smallest published volcanic forcing estimates (Mann et al, 2012a), although an older density based record (Briffa et al., 2001) showed volcanic cooling in the past few centuries that was very similar to that simulated by an energy balance model (Hegerl et al., 2003).

Other recent work (Mann et al. 2012a) suggest that this discrepancy could arise from limitations in certain types of proxy information used in temperature reconstructions, in particular tree-ring width temperature proxies which are typically obtained from tree-line proximal environments. This finding has been challenged by Anchukaitis et al (2012), which in turn has been challenged by Mann et al (2012b).

To test whether the low scaling factors could be arising solely due to the differences in response to large volcanic eruptions, the detection analysis was repeated with the years surrounding the largest volcanic eruptions masked out. For this analysis large volcanic eruptions were defined as periods when the aerosol optical depth in the tropics within the Crowley et al. 2008 dataset (which many of the models implement, see table 2) exceeds 0.25. All these events (namely, 3 major eruptions in the 13th century, Kuwae in the mid-15th century, and Tambora in 1815), plus 5 years on either side were masked out (indicated by grey bars in figure 2) prior to the detection analysis. The results are shown in figure 6c and are similar to those calculated using 21-year smoothing (figure 6b). The majority of the scaling factors now lie around unity, indicating that the model response is consistent with the reconstructions. The uncertainty ranges have also increased. This is to be expected, as the large volcanic eruptions represent some of the strongest signals in the record. By masking out large volcanic eruptions, substantial constraints on the scaling factors are removed and the signal-to-noise ratio is reduced.

d) Implications for the detection of recent climate change

We now turn to examining internal climate variability on long timescales. We have two alternative samples of internal variability: one taken from model control simulations, and one given by the residual variability in the reconstructed temperature not explained by the fingerprint for external forcing, calculated from ž (see equation 4). For the TLS regression to be self-consistent, the variability of the residuals should be comparable to that of the control simulations. Figure 4b shows that the residual from six out of eight reconstructions (ignoring the un-scaled Christiansen and Ljungqvist 2011 reconstruction) is consistent with at least one control simulation for the period 1401-1850. The other two show a larger residual over part of the LIA, (see figure 5) and have poor correlations with the model simulations (see figure 3). In contrast, residuals from only four reconstructions are consistent for the longer time period 851-1850 because the largest residuals occur early in the millennium (figure 4b), during the MCA whose peak does not coincide with periods of strong forcing (e.g. high solar activity see Ammann et al. 2007, and Jungclaus et al. 2010) and which, if the model fingerprints are correct, would point either toward unusually pronounced internal variability (Goosse et al. 2012) or, perhaps, increased sampling uncertainty and data noise in the reconstructions and/or forcings.

If the control simulations do not adequately sample the full range of the climate's internal variability then it could have a profound impact on many detection studies carried out over the last couple of decades (see Hegerl and Zwiers, 2011), as these have mainly relied on samples of internal variability derived from models. To examine if the recent warming is detectably different from internal variability, given the estimates of residual variability calculated here, we examine the largest trends in these estimates of internal variability and compare them to the recent period. Figure 7 shows the recent 50 year trend (corresponding to 1960-2010) calculated from the HadCRUT4 data (Morice et al. 2012) for all domains considered here compared to estimates of internal climate variability from the reconstructions. For all the reconstructions investigated, this alternative sample of internal variability calculated from the residuals of the regression has 50 year trends that are much smaller than the recent instrumental trend in the domain reconstructed (this conclusion also holds for 100 year trends, not shown). Thus, reconstructed temperatures of the last millennium confirm that the contribution by internal climate variability to the recent warming is small, strengthening the claim that internal variability alone is 'extremely unlikely' to explain recent warming (Hegerl et al. 2007b).

The recent observed trends are also unusual in the context of total natural climate variability (forced and unforced) since the maximum trends calculated from all the raw reconstructions for pre-industrial periods (850-1850) are found to be significantly smaller than the recent 50 year trend (not shown). This is also true for the multi-model mean; however, several of the individual model simulations contain a small number of slightly larger 50 year trends associated with the largest volcanic eruptions.

5. Which forcings are important?

To address the question of which external forcing is most important to explain the changes observed, individually forced simulations are required. Here we use multi-model fingerprints from three different GCMs (see figure 2c and 2d and table 2) to investigate the contribution from natural external forcings (solar and volcanic forcing combined) and from changes in the concentrations of well mixed greenhouse gases, particularly the dip in CO_2 recorded over parts of the LIA (see e.g. MacFarling Meure et al. 2006). The fingerprint method is based on the period 1400-1900, after which other anthropogenic forcings, particularly anthropogenic aerosols and, to a lesser extent, land use change become increasingly important (e.g. Hegerl et al. 2007b and Tett et al 2007). This analysis used the TLS detection and attribution method (equation 1) where several scaling factors β_i were estimated to fit the fingerprints $X_i(t)$ to the reconstructions $Y(t)$. Several of the model simulations which are used to calculate the fingerprints (see figure 2c and 2d) are themselves calculated as the sum of two simulations and this was taken into account when estimating the ratio of internal variability in the fingerprints to that in the reconstructions.

The detailed results for three reconstructions and the scaling factors for a larger range of reconstructions are shown in figure 8. The combined volcanic and solar fingerprint is detectable in all the reconstructions used and causes large cooling episodes in the mid-15th, 17th and early 19th centuries. Since the volcanic signal dominates the volcanic plus solar fingerprint, at least in the models, these results suggest that volcanic forcing is the dominant driver of forced variability in pre-industrial SATs for the time period studied here. However, independently from solar and volcanic forcing, a significant temperature change has been detected in response to pre-20th century greenhouse gas variations in all but three reconstructions. This forcing caused a small but sustained cooling during much of the 16th and 17th centuries with a best estimate of up to ~0.1-0.2°C (depending on the reconstruction used) relative to the mean temperature for the period 1400-1900 (see figure 8a).

The cause of this decrease in CO_2 has not been conclusively determined. Some authors (e.g. Ruddiman 2003; Faust et al. 2006; Nevle and Bird 2008) have argued that it could be a consequence of human land-use activity, attributing the decrease in CO_2 to a decrease in agricultural usage and therefore a subsequent increase in natural vegetation following the conquest of the Americas (~1519 to ~1700). However Pongratz et al. suggest (2011) that this is unlikely. Yet other studies (Joos et al 1999; Trudinger et al 2002) attribute the drop to natural forcings, such as solar and volcanic forcing. It is also possible that internal climate variability could partly explain some of the dip (see e.g. Jungclaus et al 2010). Despite this uncertainty in the origin of the reduced greenhouse gas concentration over that period, our paper shows for the first time that this decrease in CO_2 and the subsequent slow increase caused a detectible temperature response to greenhouse gases prior to 1900, highlighting the role of greenhouse gas forcing prior to the more recent period of industrial greenhouse gas emissions.

6. Discussions and Conclusions

The work presented in this paper examines the role of external forcings on the climate of the last millennium. Consistent with earlier studies (Crowley 2000; Yoshimori et al. 2005; Hegerl et al. 2007a), we find the LIA likely to have been in large part externally forced, since a large fraction of the variance in most reconstructions can be explained by the model simulations and since the model fingerprint for forced variability is detectable at the 5% level in all the reconstructions analysed.

The variance of the residuals that is not explained by the response to external forcing as simulated in the models is, for the majority of reconstructions, consistent with the variance of control simulations if analysed over the past 600 years. There are, however, large differences between the different reconstructions. Several are only poorly correlated to the model simulations and have large residuals that cannot be explained by the estimated radiative forcing even over this shorter interval. Since the uncertainties in the model simulations and reconstructions are independent of each other, the high correlation between the models and some reconstructions is unlikely to be due to chance alone. From attribution analysis using fingerprints of natural (volcanic and solar) and anthropogenic (greenhouse gas forcing), it can be shown that explosive volcanism and changes in solar output combined are the dominant drivers of forced variability over the second half of the last millennium, although greenhouse gas variations are also likely to have significantly contributed to the cold conditions during the period 1600-1800.

The variance of the residuals calculated from the detection analysis encompassing the MCA is for many of the reconstructions larger than the variance of the control simulations during this period. This could be due to increased uncertainty in the reconstructions, for example, due to the declining number of proxies or to errors in the forcing datasets used to drive the models. It could also be due to strong and anomalous periods of internal variability, or both.

The 50 year trends in the samples of internal variability resulting from the detection analysis for the full period analysed here (850-1950) were compared to the recent 50 year temperature trend. This shows that for all the samples of internal variability calculated (even those with higher variance than the control simulations) the largest 50 and 100 year trend found in reconstructions after removing the forced component is much smaller than that found in the last 50/100 years of the instrumental record (1960-2010 and 1910-2010). This substantially strengthens the claim that internal variability alone is ‘extremely unlikely’ to explain recent warming (Hegerl et al 2007b).

For the majority of the reconstructions the detection analysis estimates scaling factors significantly less than unity, indicating that the response to external forcing in the models is stronger than that inferred from the proxy reconstructions. While we cannot rule out that this discrepancy is due to an excessively large response in the multi-model mean to all forcings, this would not explain our finding that the discrepancy between the simulated and reconstructed responses is no longer apparent when disregarding a short period immediately following the largest volcanic eruptions of the past millennium. Possible explanations for this latter observation, as noted earlier, are (a) better fidelity of the low-frequency signal in proxy reconstructions, or (b) possible loss of fidelity of certain types of proxy data (particularly tree-ring data) in resolving very large volcanic cooling episodes. Other possible factors are (c) uncertainties in the magnitude of the volcanic forcing used in the multi model ensemble used here, (d) uncertainty in the representation of volcanic forcing within the models, (e) errors in the response of the models to volcanic cooling, or some combination of all of these factors.

To conclude, this paper builds on previous studies looking at the detection and attribution of the causes of climate change in NH temperature reconstructions, such as those by Hegerl et al 2003 and 2007. This work uses an ensemble of GCM simulations, many of which have only just become available as part of the CMIP5/PMIP3 initiative, as well as many more reconstructions compared to earlier results using fewer simulations, less reliable forcing estimates and sometimes Energy Balance Models. Our analysis also pushes detection of the forced response back to 850 in many cases.

Our results have enabled us to better place the recent warming in the context of long term change, have strengthened the evidence for the importance of natural forcing in the climate of the last millennium, and have highlighted that the model-reconstruction discrepancy in the response to volcanic eruptions, as well as significant differences in the magnitude of the MCA, that cannot be fully explained by our understanding of internal variability. We also detect, for the first time, a pre-industrial greenhouse gas signal prior to 1900.

Acknowledgements

The publication was supported by NERC grant NE/G019819/1 and NCAS via a CMIP5 grant and a long term grant. M.E.M. acknowledges support from the ATM program of the National Science Foundation (Grant ATM-0902133).

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 (listed in Table 2 of this paper) for producing and making available their model output. 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 acknowledge the help of Bette Otto-Bliesner, Johann Jungclaus, Gavin Schmidt and Jie Zhang for help obtaining and using the model data and Gareth Jones for making a long HadCM3 control simulation available for our use, and thank three anonymous reviewers for the insightful comments that helped improve the manuscript.

References

Allen, M. R. and P. A. Stott, 2003, Estimating signal amplitudes in optimal fingerprinting, Part I: Theory. *Clim. Dyn*. **21**, 477-491

Allen, M. R. and S. F. B. Tett, 1999: Checking for model consistency in optimal finger printing. *Clim Dyn*, **15**, 419–434

Ammann, C. M. and E. Wahl, 2007: The importance of the geophysical context in statistical evaluations of climate reconstruction procedures. *Climatic Change*, **85**, 71-88.

Ammann, C. M., F. Joos, D. S. Schimel, B. L. Otto-Bliesner and R. A. Tomas, 2007: Solar influence on climate during the past millennium: Results from transient simulations with the NCAR Climate System Model. *Proc. Natl Acad. Sci*. **104**, 3713-3718.

Anchukaitis, K. et al: “Tree rings and volcanic cooling”, Nature Geoscience, **5**, 836–837, (2012)

Braconnot, P., et al., 2012: Evaluation of climate models using palaeoclimatic data. Nature Clim. Change **2** (6), 417–424

Briffa, K. R. et al. 2001: Low-frequency temperature variations from a northern tree ring density network. J Geophys. Res. 106, 2929-2941.

Christiansen, B. and Ljungqvist, F. C., 2011: Reconstruction of the extratropical NH mean temperature over the last millennium with a method that preserves low-frequency variability, *J. Climate*, **24**, 6013-6034.

Cook, E., J. Esper and R. D'Arrigo, 2004: Extra-tropical Northern Hemisphere land temperature variability over the past 1000 years. Quat. Sci. Rev 23, 2063-2074.

Cox, P. and C. Jones, 2008: Climate change - illuminating the modern dance of climate and CO2. Science, **321**, 1642–1644.

Crowley, T. J., 2000: Causes of climate change over the past 1000 years. Science, **289**, 270–277.

Crowley, T. J., G .Zielinski, B. Vinther, R. Udisti, K. Kreutzs, J. Cole-Dai, and E. Castellano, 2008: Volcanism and the Little Ice Age, PAGES Newsletter, **16**, 22-23

D'Arrigo, R., R.Wilson, and G. Jacoby, 2006: On the long-term context for late twentieth century warming. J. Geophys. Res., **111**, D03103, doi:10.1029/2005JD006352.

Driscoll, S., A. Bozzo, L. J. Gray, A. Robock, and G. Stenchikov (2012), Coupled Model Intercomparison Project 5 (CMIP5) simulations of climate following volcanic eruptions, *J. Geophys. Res.*, doi:10.1029/2012JD017607, in press.

Esper, J., E. R. Cook, and Schweingruber F. H. (2002): Low-frequency signals in long tree-ring chronologies for reconstructing past temperature variability, Science, 295, 2250–2253.

Esper, J., and D. C. Frank, 2009: IPCC on heterogeneous Medieval Warm Period. Climatic Change, **94**, 267–273.

Esper, J. and coauthors, 2012: Orbital forcing of tree-ring data, *Nature Climate Change* (advance publication), DOI: 10.1038/NCLIMATE1589

Faust F.X., Gnecco C., Mannstein H. and Stamm J., 2006: Evidence for the post conquest demographic collapse of the Americas in CO2 levels. *Earth Interactions* **10**: doi:10.1175/EI157.1.

Forster, P. V and Coauthors, 2007: Understanding and attributing climate change. Climate Change 2007: The Physical Science Basis, S. Solomon et al., Eds., Cambridge University Press.

Frank, D., J. Esper, and E. R. Cook, 2007: Adjustment for proxy number and coherence in a large-scale temperature reconstruction. *Geophys. Res. Lett*. **34** L16709 doi:10.1029/2007gl030571

Gao, C. C., A. Robock, and C. Ammann, 2008: Volcanic forcing of climate over the past 1500 years: An improved ice core based index for climate models, *J. Geophys. Res*., **113**, D23111, doi:10.1029/2008JD010239

Gent, Peter R., et al., 2011: The Community Climate System Model Version 4. *J. Climate*, **24**, 4973–4991.

Giorgetta, M.A., and coauthors. 2012: Climate change from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project 5, submitted to JAMES, special issue The Max Planck Institute for Meteorology Earth System Model.

Goosse, H., E Crespin, S Dubinkina, M.-F. Loutre, M. E. Mann, H. Renssen, Y. Sallaz-Damaz, D. Shindell, 2012: The role of forcing and internal dynamics in explaining the "Medieval Climate Anomaly". *Clim Dyn*, **39**, 2847-2866.

Gordon, C., and Coauthors, 2000: The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Clim Dyn* **16** :147–168

Gray L. J., et al., Reviews of Geophysics, **48** (2010)

Hasselmann, K., 1976: Stochastic climate models. Part 1. Theory. Tellus, **28**, 473–485.

Hegerl G.C., v. Storch H., Hasselmann K., Santer B. D., Cubasch U. and Jones P. D. 1996: Detecting greenhouse gas induced Climate Change with an optimal fingerprint method. J. Climate **9**, 2281-2306

Hegerl, G.C., T. J. Crowley, S. K. Baum, K.-Y. Kim, and W. T. Hyde, 2003: Detection of volcanic, solar, and greenhouse gas signals in paleo-reconstructions of Northern Hempispheric temperature. Geophys. Res. Lett., **30**, 1242, doi:10.1029/2002GL016635.

Hegerl, G. C., T. J., Crowley, W. T. Hyde, and D. J. Frame, 2006: Climate sensitivity constrained by temperature reconstructions over the past seven centuries. Nature, **440**(7087), 1029-1032.

Hegerl, G. C., T. J. Crowley, M. Allen, W. T. Hyde, and H. N. Pollack, 2007a: Detection of human influence on a new, validated 1500-year temperature reconstruction. J. Climate, **20**,650–666.

Hegerl, G. C. and Coauthors, 2007b: Understanding and attributing climate change. Climate Change 2007: The Physical Science Basis, S. Solomon et al., Eds., Cambridge University Press, 663-745.

Hegerl, G.C. and F.W. Zwiers, 2011: Use of models in detection and attribution of climate change. WIRES: Climate Change, **2**, 570-591.

Hurtt, G. C., and coauthors, 2009: Harmonization of Global Land-Use Scenarios for the Period 1500-2100 for IPCC-AR5, *Integrated Land Ecosystem-Atmosphere Processes Study (iLEAPS) Newsletter*, 6–8

Jones, P.D. and Coauthors, 2009: High-resolution palaeoclimatology of the last millennium: A review of current status and future prospects. Holocene, **19**, 3–49.

Juckes, M. N., M. R. Allen, K. R. Briffa, J. Esper, G. C. Hegerl, A. Moberg, T. J. Osborn, and S. L. Weber, 2007: Millennial temperature reconstruction intercomparison and evaluation. Climate Past, **3**, 591–609.

Jansen, E et al., 2007: Palaeoclimate. In: Climate Change 2007: The Physical Science Basis, S. Solomon et al., Eds., Cambridge University Press, 433-497.

Jungclaus, J. H., and coauthors, 2010: Climate and carbon-cycle variability over the last millennium, *Clim. Past*, **6**, 723-737, doi:10.5194/cp-6-723-2010.

Kaplan, J. O., K. M Krumhardt, and N Zimmerman 2009: The prehistoric and preindustrial deforestation of Europe, *Quat. Sci. Revs*., **28**, 3016–3034.

Kaufman, D. Sand co-authors, 2009: Recent warming reverses long-term Arctic cooling. *Science*, **325**, 1236-1339.

Landrum, L., B. L. Otto-Bliesner, E. R. Wahl, A. Conley, P. J. Lawrence, N. Rosenbloom, and H. Teng, 2011: Last millennium climate and its variability in CCSM4. (submitted)

Lefohn, A. S., J. D Husar., and R. B . Husar, 1999: Estimating historical anthropogenic global sulfur emission patterns for the period 1850–1990*, Atmos. Env*., **33**, 3435–3444.

MacFarling Meure, C., D. and co-authors. 2006: Law Dome CO2, CH4 and N2O ice core records extended to 2000 years BP, *Geophys. Res. Lett.,* **33**, L14810.

Mann ME, Bradley RS, Hughes MK (1998) Global-scale temperature patterns and climate forcing over the past six centuries. Nature 392:779–787

Mann, M. E. Smoothing of climate time series revisited, 2008: Geophys. Res. Lett. **35** doi:10.1029/2008GL034716

Mann, M.E., Zhang, Z., Hughes, M.K., Bradley, R.S., Miller, S.K., Rutherford, S., Proxy-Based Reconstructions of Hemispheric and Global Surface Temperature Variations over the Past Two Millennia, *Proc. Natl. Acad. Sci.*, **105**, 13252-13257, 2008.

Mann, M. and coauthors, 2009: Global signatures and dynamical origins of the Little Ice Age and Medieval Climate Anomaly. Science, **326**, 1256–1260.

Mann, M. E., J. D. Fuentes and S. Rutherford, 2012: Underestimation of volcanic cooling in tree-ring-based reconstructions of hemispheric temperatures, *Nat. Geosci*., **5**, 202–205, doi:10.1038/ngeo1394, 2012. 1655

Mann, M.E. and coauthors, Reply to “Tree rings and volcanic cooling”, Nature Geoscience, **5**, 837–838, (2012)

Mann, M.E., Rutherford, S., Wahl, E., Ammann, C., 2007: Robustness of Proxy-Based Climate Field Reconstruction Methods, *J. Geophys. Res.*, 112, D12109, doi: 10.1029/2006JD008272, 2007.

Miller, G. H., and co-authors, 2012, Abrupt onset of the Little Ice Age triggered by volcanism and sustained by sea-ice/ocean feedbacks, Geophys. Res. Lett., **39**, L02708, doi:10.1029/2011GL050168.

Moberg, A., D. M. Sonechkin, K. Holmgren, N. M. Datsenko, and W. Karlén, 2005: Highly variable Northern Hemisphere temperatures reconstructed from low- and high resolution proxy data. Nature, **433**, 613–617; Corrigendum, 439,1014.

Morice, C. P., J. J. Kennedy, N. A. Rayner, and P. D. Jones, 2012: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set, *J. Geophys. Res*., **117**, D08101, doi:10.1029/2011JD017187.

Nevle R.J. and Bird D.K., 2008: Effects of syn-pandemic fire reduction and reforestation in the tropical Americas on atmospheric CO2 during European conquest. Palaeogeography, Palaeoclimatology, Palaeoecology **264**: 25–38.

Phipps, S. J., L. D. Rotstayn, H. B. Gordon, J. L. Roberts, A. C. Hirst and W. F. Budd, 2011: The CSIRO Mk3L climate system model version 1.0 - Part 1: Description and evaluation, Geoscientific Model Development, **4**, 483-509, doi:10.5194/gmd-4-483-2011

Phipps, S. J., L. D. Rotstayn, H. B. Gordon, J. L. Roberts, A. C. Hirst and W. F. Budd, 2012: The CSIRO Mk3L climate system model version 1.0 - Part 2: Response to external forcings, Geoscientific Model Development, **5**, 649-682, doi:10.5194/gmd-5-649-2012

Pongratz, J., C. H . Reick, T. Raddatz and M. Claussen, 2008: A reconstruction of global agricultural areas and land cover for the last millennium, *Glob. Biogeochem. Cycles*, **22**, GB3018, doi:10.1029/2007GB003153.

J. Pongratz, K. Caldeira, C.H. Reick, and M. Claussen, 2011: Coupled climate-carbon simulations indicate minor global effects of wars and epidemics on atmospheric CO2 between AD 800 and 1850, *The Holocene,* **21**, 843-851

Pope, V.D., M. L. Galliani, P.R. Rowntree, R.A. Stratton, 2000: The impact of new physical paramaterizations in the Hadley Centre climate model – HadAM3. *Clim Dyn* **16**: 123-146.

Plummer, C. T., Curran, M. A. J., van Ommen, T. D., Rasmussen, S. O., Moy, A. D., Vance, T. R., Clausen, H. B., Vinther, B. M., and Mayewski, P. A 2012.: An independently dated 2000-yr volcanic record from Law Dome, East Antarctica, including a new perspective on the dating of the c. 1450s eruption of Kuwae, Vanuatu, Clim. Past Discuss., **8**, 1567-1590, doi:10.5194/cpd-8-1567-2012.

Schmidt, G.A and co-authors. 2006: Present day atmospheric simulations using GISS ModelE: Comparison to in-situ, satellite and reanalysis data. J. Climate **19**, 153-192.

Ruddiman W, 2003: The anthropogenic greenhouse era began thousands of years ago. Climatic Change **61**: 261–293.

Schmidt, G. A and co-authors. 2011: Climate forcing reconstructions for use in PMIP simulations of the Last Millennium (v1.0), *Geosci. Model Dev*., **4**, 33-45, doi:10.5194/gmd-4-33-2011

Schmidt, G. A and co-authors. 2012: Climate forcing reconstructions for use in PMIP simulations of the Last Millennium (v1.1), *Geosci. Model Dev*., **5**, 185-191, doi:10.5194/gmd-5-185-2012

Shapiro A. I., Schmutz W., Rozanov E., Schoell M., Haberreiter M., Shapiro A. V., and Nyeki S., 2011: A new approach to long-term reconstruction of the solar irradiance leads to large historical solar forcing, Astron Astrophys 529:A67.

Shindell, D.T., G. Faluvegi, R.L. Miller, G.A. Schmidt, J.E. Hansen, and S. Sun, 2006: Solar and anthropogenic forcing of tropical hydrology. *Geophys. Res. Lett.*, **33**, L24706, doi:10.1029/2006GL027468.

Smerdon J. E., 2012: Climate models as a test bed for climate reconstruction methods: pseudoproxy experiments. WIREs Clim Change 2012, 3:63–77. doi: 10.1002/wcc.149

Steinhilber, F., J. Beer, and C. Frohlich, 2009.: Total solar irradiance during the Holocene, *Geophys. Res. Lett.,* **36**, L19704, doi:10.1029/2009GL040142.

Stott, P. A., S. F. B. Tett, G. S. Jones, M. R. Allen, J. F. B. Mitchell, G. J. Jenkins., 2000. External control of twentieth century temperature change by natural and anthropogenic forcings. Science, **290**, 2133-2137.

Stott, P. A., J. F. B. Mitchell, M. R. Allen, T. L. Delworth, J. M. Gregory, G. A. Meehl, B. D. Santer, 2006: Observational constraints on past attributable warming and predictions of future global warming. J. Climate, **19**, 3055–3069.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, B. Am. Meteorol. Soc., **92**, 485–498, doi:10.1175/BAMS-D-11-00094.1, 2012. 1655, (2012)

Tett, S. F. B., P. A. Stott, M. R. Allen, W. J. Ingram, and J. F. B. Mitchell, 1999: Causes of twentieth century temperature change near the earth's surface. Nature, **399**, 569–572.

Tett, S. F. and co-authors. 2007:. The impact of natural and anthropogenic forcings on climate and hydrology since 1550. Climate Dynamics, **28**(1), 3-34. Springer. Retrieved from http://centaur.reading.ac.uk/16636/

Timmreck, C., S. J. Lorenz, T. J. Crowley, S. Kinne, T. J. Raddatz, M. A. Thomas, and J. H. Jungclaus, 2009: Limited temperature response to the very large AD 1258 volcanic eruption, Geophys. Res. Lett., **36**, L21708, doi:10.1029/2009GL040083.

Timmreck, C., 2012: Modelling the climatic effects of large explosive volcanic eruptions. WIREs Clim Change, **3**: 545–564.

Trenberth, K. E. and Coauthors, 2007: Understanding and attributing climate change. Climate Change 2007: The Physical Science Basis, S. Solomon et al., Eds., Cambridge University Press.

Vieira, L. E. A., S. K. Solanki, N. A. Krivova, and I. Usoskin, 2011: Evolution of the solar irradiance during the Holocene, *Astron. Astroph*., **531**, A6, doi:10.1051/0004-6361/201015843.

Wang, Y.-M., J. L. Lean, and N. R Sheeley, Jr., 2005: Modeling the Sun's Magnetic Field and Irradiance since 1713, *Astrophys. J.*, **625**, 522–538, doi:10.1086/429689.

Weber, S.L., 2005: A timescale analysis of the NH temperature response to volcanic and solar forcing in the past millennium. Climate of the Past, **1**, 9–17.

Wu T. W., 2012: A Mass-Flux Cumulus Parameterization Scheme for Large-scale Models: Description and Test with Observations, *Clim.Dyn*., **38**, 725-744.

Yokohata, T., and co-authors., 2005: Climate response to volcanic forcing: Validation of climate sensitivity of a coupled atmosphere-ocean general circulation model. Geophys. Res. Lett., **32**, L21710, doi:10.1029/2005GL023542.

Tables and Figures

Table 1 – Reconstructions used - *The table includes citation (column 1), details of the geographical region of the reconstructions (column2), the time period covered (3rd column) and lists if multiproxy or tree-ring only based (for more details see papers). The additional notes column details which reconstruction is used if the paper referenced contains more than one. The name in brackets represents the label given to the reconstruction in subsequent figures.*

Table 2 –Model simulations and their forcings for *further details see references; the references are CEA – Crowley et al. (2008), GRA – Gao et al. (2008), VSK – Viera et al. (2011), SBF – Steinhilber et al. (2009), WLS – Wang et al. (2005). SJA – Schmidt et al. (2011), PEA – Pongratz et al. (2008), Hur- Hurtt et al. (2009), KK10 – Kaplan et al. (2009). JLT – Jungclaus et al. (2010) MM -* MacFarling Meure et al. (2006). *An X indicates that the forcing is not included. The model simulations indicated by a star have been made available as part of the CMIP5 and PMIP3 projects.*

Figure 1 – Reconstructed northern hemisphere land and sea surface air temperature *a) All reconstructions that represent a) the whole NH (land and sea). b) 20-90N land only c) 30-90N land only. On all panels the HadCRUT4 instrumental data (Morice et al 2012) is plotted in black. All annual data are first smoothed with a 10 year Butterworth filter (to enable comparison to reconstructions), and are further smoothed by an 11 year boxcar filter to focus on interdecadal timescales (see text for discussion).*

Figure 2 - Model simulations for the region 0-90N land and sea*. a) Simulations forced with most complete set of external forcings, referred to as* ***ALL forced*** *simulations. The ensemble mean is shown in black. b) A comparison of the ensemble mean shown in figure 2a with the NH reconstructions shown in figure 1a, where the light orange shading shows the outer bounds for all 4 reconstructions and the solid orange line the mean of all four reconstructions. c) Simulations driven with a combination of solar and volcanic forcing. d) Simulations driven with just well mixed greenhouse gas forcing. All simulations are smoothed by a 10 year Butterworth filter and then an 11 year running box-car filter. The grey bars on the top three panels show periods of high volcanism.*

Figure 3 –Variance in reconstructions that is explained by the models.
Explained variance (R^2) using the smoothed ensemble mean for 200 year periods (thin black box: analysis period for first and last 200yr period). Symbols show explained variance for the individual reconstructions, while the thick black line shows the average R^2. The period 1250-1270 is not included in this particular analysis due to the known large discrepancy between reconstructions and model response to the

1258 eruption, which substantially drives up the correlations between model simulations. Symbols are centred on the period considered. The black dashed line shows the mean explained variance in the perfect model study.

Figure 4: Contribution by external forcing to NH mean temperatures. *a) Estimate of the contribution by the multi-model fingerprint (orange, solid; 5-95% uncertainty range for scaling only dark orange) to three of the reconstructions (blue, green, red), calculated for the period 851-1950 compared to the 5-95% uncertainty range of internal variability (light orange shading). b) Component of internal variability calculated from every reconstruction analysed (i.e. the residual between the fitted model results and reconstructions). The horizontal lines show two standard deviations of control simulation variability. c) Detection results for all reconstructions considered (see axis label). Best fit scaling factors (crosses) are shown with 5-95% ranges (vertical rectangles); results from an analysis with noise variance scaled to the residual variance are shown by a vertical line. Fingerprints are detectible if scaling factors are significantly above zero and consistent with the reconstruction if not significantly different from unity. A solid rectangle indicates that the variability of the residual is smaller than ~90% of the control samples, a dashed rectangle that the variability is smaller than at least one control sample. An open rectangle indicates that the residual is not consistent with any of the model control samples.*

Figure 5 - Regression lines (851-1950) – *plots show reconstructions on the y axis against model results on the x axis, with the calculated regression lines shown in blue for a TLS estimate and purple for the OLS estimate (best fit: solid, 95% range dotted). Red asterisks show MCA years (950-1250), green asterisks show LIA years*

(1400-1700), orange asterisks show 20th century years, any other year is shown in black.

Figure 6 – **As figure 4b, but showing detection results for sensitivity tests**. a) Results for the standard analysis but without the extra 11 year boxcar smoothing. b) Results for the standard analysis but with 21 year boxcar smoothing instead of the usual 11 year smoothing. c) Results for analysis with major volcanic eruptions masked out in both the reconstructions and model simulations.

Figure 7 Distribution and maximum 50 year trends of internal variability estimated from reconstructions – a) *Distribution of 50 year trends found in the scaled residuals covering the time period 851-1950. The distribution of the trends is shown in the form of histograms with a Gaussian fit through the points. The grey shaded Gaussian shows the distribution of the 50 year trend found in the combined control simulations. The largest positive and negative 50 year trend from each reconstruction and the control simulations is shown by a bold vertical line. The recent 50 year trend (1960-2010) in the HadCRUT4 instrumental record (Morice et al. 2012) is shown by a burgundy vertical line. a) Results for NH mean SATs; b) for extra-tropics land only 20-90N and c) for extra-tropics land only 30-90N.*

Figure 8 - Results from detection and attribution analysis using individually forced fingerprints. *- a) Individually forced fingerprints for solar and volcanic forcing combined (green) and greenhouse gas forcing (blue) scaled to fit three different reconstructions over the period 1400-1900 (white area of the plot), with the 5-95% scaling uncertainty range shown by the shaded region; b) Best fit scaling factors for*

both fingerprints for several reconstructions (cross) with 5-95% uncertainty range (vertical bar). Fingerprints are detectible if scaling factors are significantly above zero and consistent with the reconstruction if not significantly different from unity. Solid rectangle: the variability of the residual is smaller than ~90% of the control samples. Dashed rectangle: variability smaller than at least one control sample.

Reconstruction	**Geographical Region**	**Period (CE)**	**Time resolution**	**Proxy Types**	**Additional Notes**
Mann et al 2009 (*Mann_09*)	0-90N land and sea	500-2010	Decadal	Multi-Proxy	
Ammann & Wahl 2007 (*Ammann*)	0-90N land and sea	1000-1980	Annual	Multi-Proxy	Update - Mann et al. 1998
Moberg et al 2005 (*Moberg*)	0-90N land and sea	1-1979	Annual	Multi-Proxy	Tree-rings only for high frequency variability
Juckes et al 2007 (*Juckes*)	0-90N land and sea	1000-1090	Annual	Multi-Proxy	Union, CVM method
D'Arrigo et al 2006 (*D'Arrigo*)	20-90N land only	713-1960	Annual	Tree-rings only	RCS reconstruction
Frank et al 2007 (*Frank*)	20-90N land only	831-1992	Annual	Tree-rings only	Update – Esper et al. 2002
Hegerl et al 2007 (*CH_blend*)	30-90N land only	946-1960	Decadal	Multi-Proxy	CH-Blend
Christiansen &	30-90N land	1000-1975	Annual	Multi-Proxy	*Christ_ scaled*

Ljungqvist 2011 (*Christiansen*) *(Christ_scaled)*	only				– is scaled to instrumental data

Table 1 – Reconstructions used - *The table includes citation (column 1), details of the geographical region of the reconstructions (column2), the time period covered (3rd column) and lists if multiproxy or tree-ring only based (for more details see papers). The additional notes column details which reconstruction is used if the paper referenced contains more than one. The name in brackets represents the label given to the reconstruction in subsequent figures.*

Model	**No.**	**Resolution**		**Forcings**			
	Ens.	**Atmosphere**	**Ocean**	**Volc**	**Solar**	**GHG**	**Land-use**
* CCSM4	1	288x192xL26	320x384xL60	GEA	VK/WLS	SJA	PEA/Hur
MPI-COSMOS	5	96x48xL19	GR3.0xL40	CEA	JLT	Inter-active	PEA
*MPI-ESM-P	1	196x98xL47	256x220xL40	CEA	VK/WLS	SJA	PEA
HadCM3	1	96x73xL19	288x144xL20	CEA	SBF/WLS	SJA	PEA
*GISS-E2-R	1	144x90xL40	288x180xL32	CEA	VK/WLS	SJA	PEA/Hur
*GISS-E2-R	1	144x90xL40	288x180xL32	GRA	VK/WLS	SJA	KK10/Hur
*Bcc-csm1-1	1	128x64xL40	360x232xL40	GRA	VK/WLS	SJA	X
CSIRO-MK3L-1-2	-	64x56xL18	128x112xL21	GRA	SBF	MM	X

Table 2 –Model simulations and their forcings for *further details see references; the references are CEA – Crowley et al. (2008), GRA – Gao et al. (2008), VSK – Viera et al. (2011), SBF – Steinhilber et al. (2009), WLS – Wang et al. (2005). SJA – Schmidt et al. (2011), PEA – Pongratz et al. (2008), Hur- Hurtt et al. (2009), KK10 – Kaplan et al. (2009). JLT – Jungclaus et al. (2010) MM* - MacFarling Meure et al. (2006). *An X indicates that the forcing is not included. The model simulations indicated by a star have been made available as part of the CMIP5 and PMIP3 projects.*

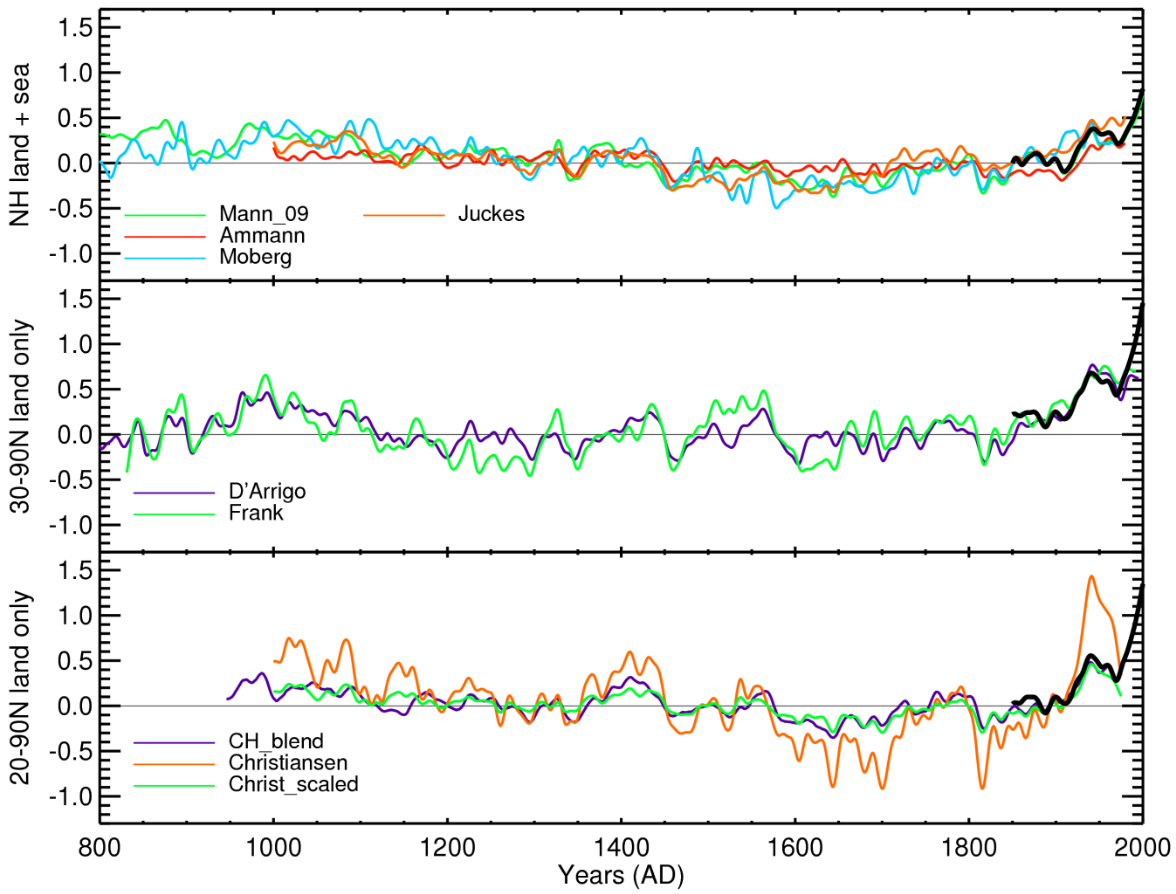

Figure 1 – Reconstructed northern hemisphere land and sea surface air temperature *a) All reconstructions that represent a) the whole NH (land and sea). b) 20-90N land only c) 30-90N land only. On all panels the HadCRUT4 instrumental*

data (Morice et al 2012) is plotted in black. All annual data are first smoothed with a 10 year Butterworth filter (to enable comparison to reconstructions), and are further smoothed by an 11 year boxcar filter to focus on interdecadal timescales (see text for discussion).

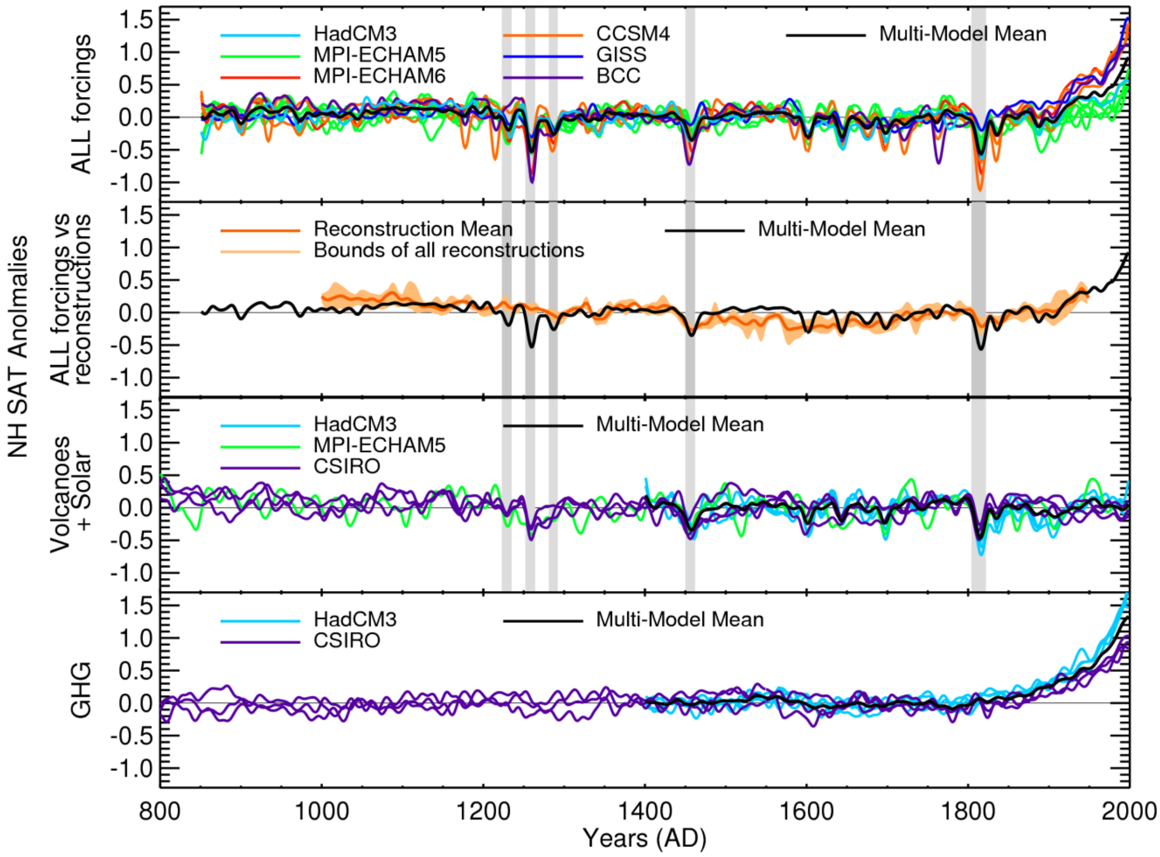

***Figure 2 - Model simulations for the region 0-90N land and sea**. a) Simulations forced with most complete set of external forcings, referred to as **ALL forced** simulations. The ensemble mean is shown in black. b) A comparison of the ensemble mean shown in figure 2a with the NH reconstructions shown in figure 1a, where the light orange shading shows the outer bounds for all 4 reconstructions and the solid orange line the mean of all four reconstructions. c) Simulations driven with a combination of solar and volcanic forcing. d) Simulations driven with just well mixed greenhouse gas forcing. All simulations are smoothed by a 10 year Butterworth filter*

and then an 11 year running box-car filter. The grey bars on the top three panels show periods of high volcanism.

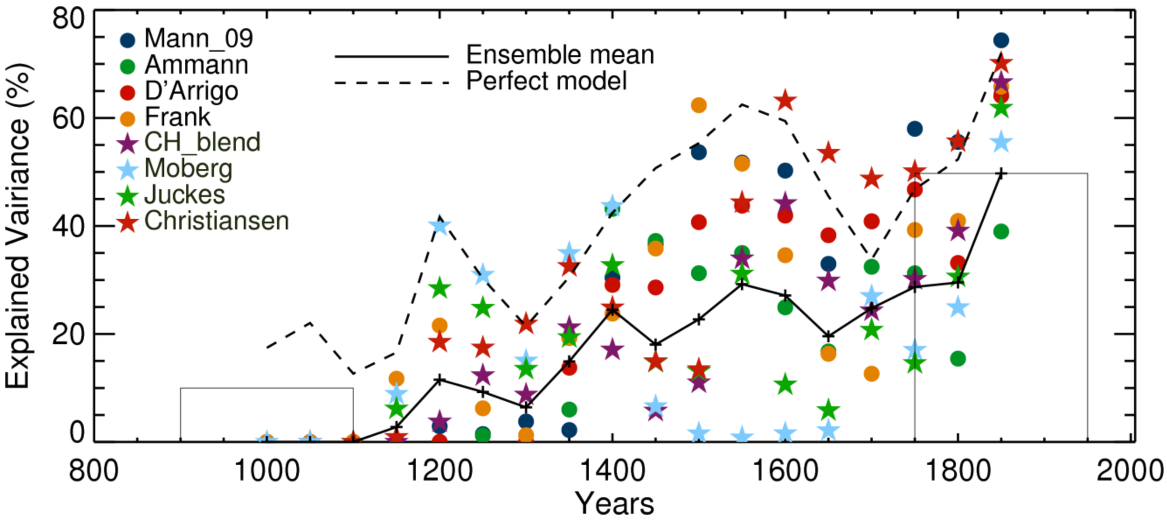

Figure 3 –Variance in reconstructions that is explained by the models*.*
Explained variance (R^2) using the smoothed ensemble mean for 200 year periods (thin black box: analysis period for first and last 200yr period). Symbols show explained variance for the individual reconstructions, while the thick black line shows the average R^2. The period 1250-1270 is not included in this particular analysis due to the known large discrepancy between reconstructions and model response to the 1258 eruption, which substantially drives up the correlations between model simulations. Symbols are centred on the period considered. The black dashed line shows the mean explained variance in the perfect model study.

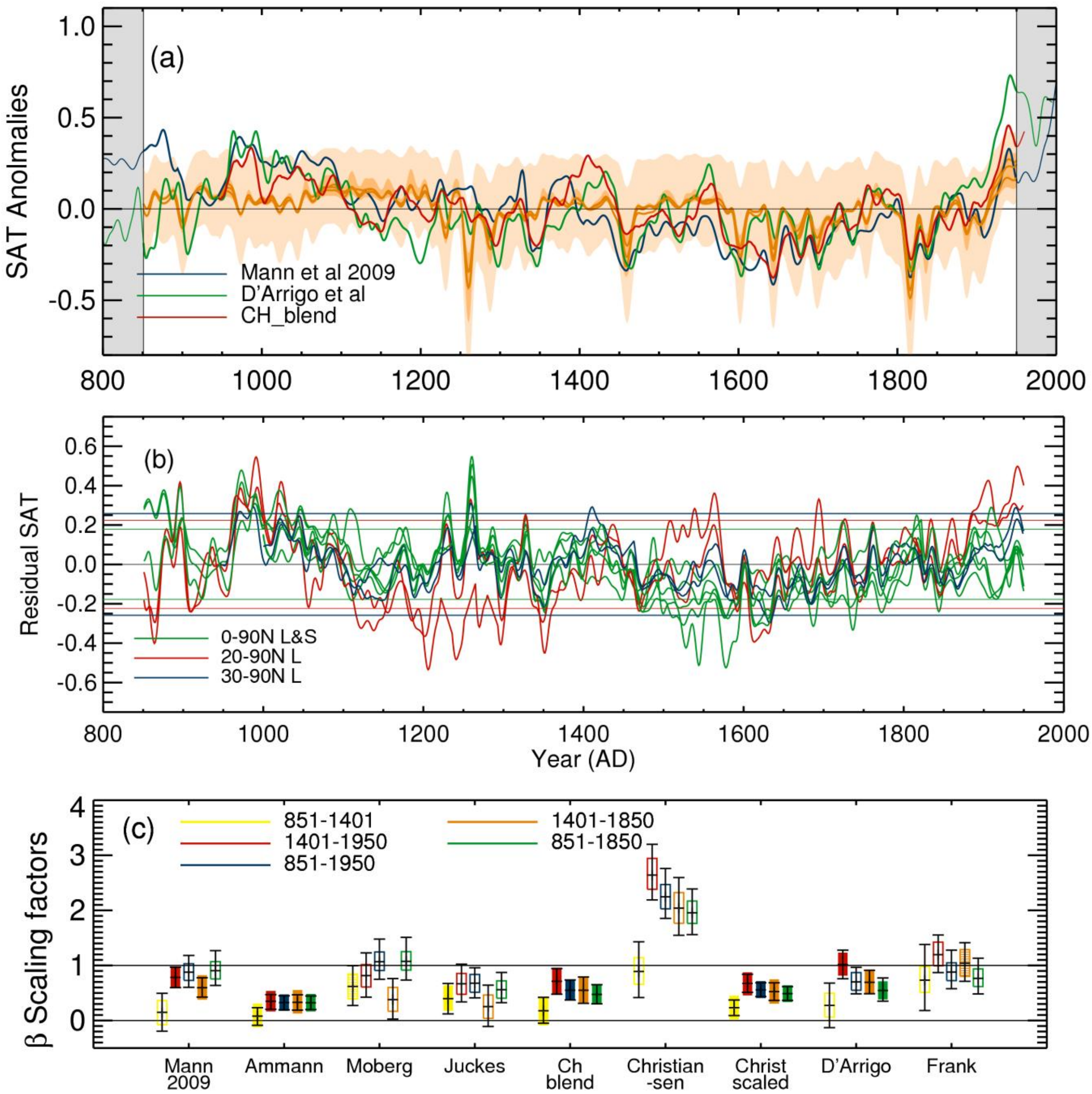

Figure 4: Contribution by external forcing to NH mean temperatures. *a) Estimate of the contribution by the multi-model fingerprint (orange, solid; 5-95% uncertainty range for scaling only dark orange) to three of the reconstructions (blue, green, red; note that they represent slightly different domains which is accounted for in the analysis), calculated for the period 851-1950 compared to the 5-95% uncertainty range of internal variability (light orange shading). b) Component of internal variability calculated from every reconstruction analysed (i.e. the residual between the fitted model results and reconstructions). The horizontal lines show two*

standard deviations of control simulation variability. c) Detection results for all reconstructions considered (see axis label). Best fit scaling factors (crosses) are shown with 5-95% ranges (vertical rectangles); results from an analysis with noise variance scaled to the residual variance are shown by a vertical line. Fingerprints are detectible if scaling factors are significantly above zero and consistent with the reconstruction if not significantly different from unity. A solid rectangle indicates that the variability of the residual is smaller than ~90% of the control samples, a dashed rectangle that the variability is smaller than at least one control sample. An open rectangle indicates that the residual is not consistent with any of the model control samples.

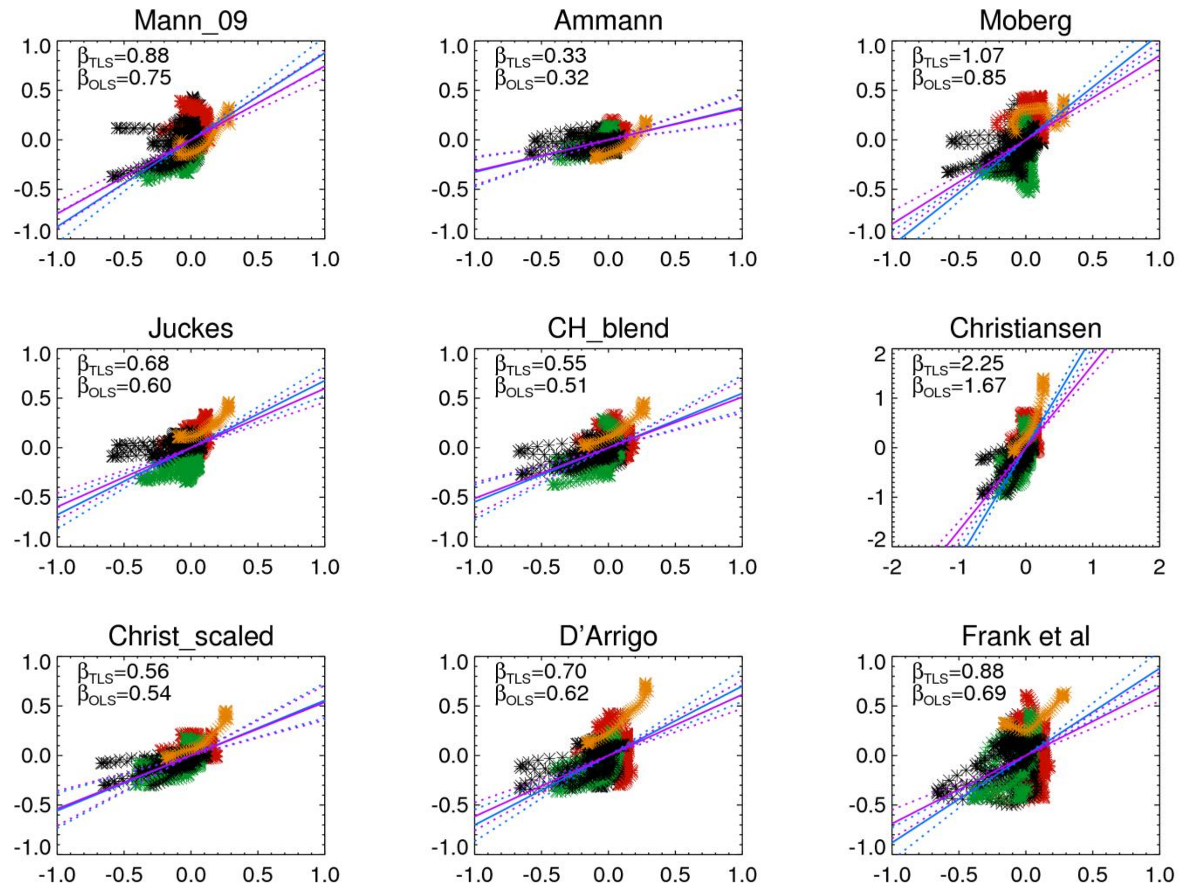

Figure 5 - Regression lines (851-1950) – *plots show reconstructions on the y axis*

against model results on the x axis, with the calculated regression lines shown in blue for a TLS estimate and purple for the OLS estimate (best fit: solid, 95% range dotted). Red asterisks show MCA years (950-1250), green asterisks show LIA years (1400-1700), orange asterisks show 20th century years, any other year is shown in black.

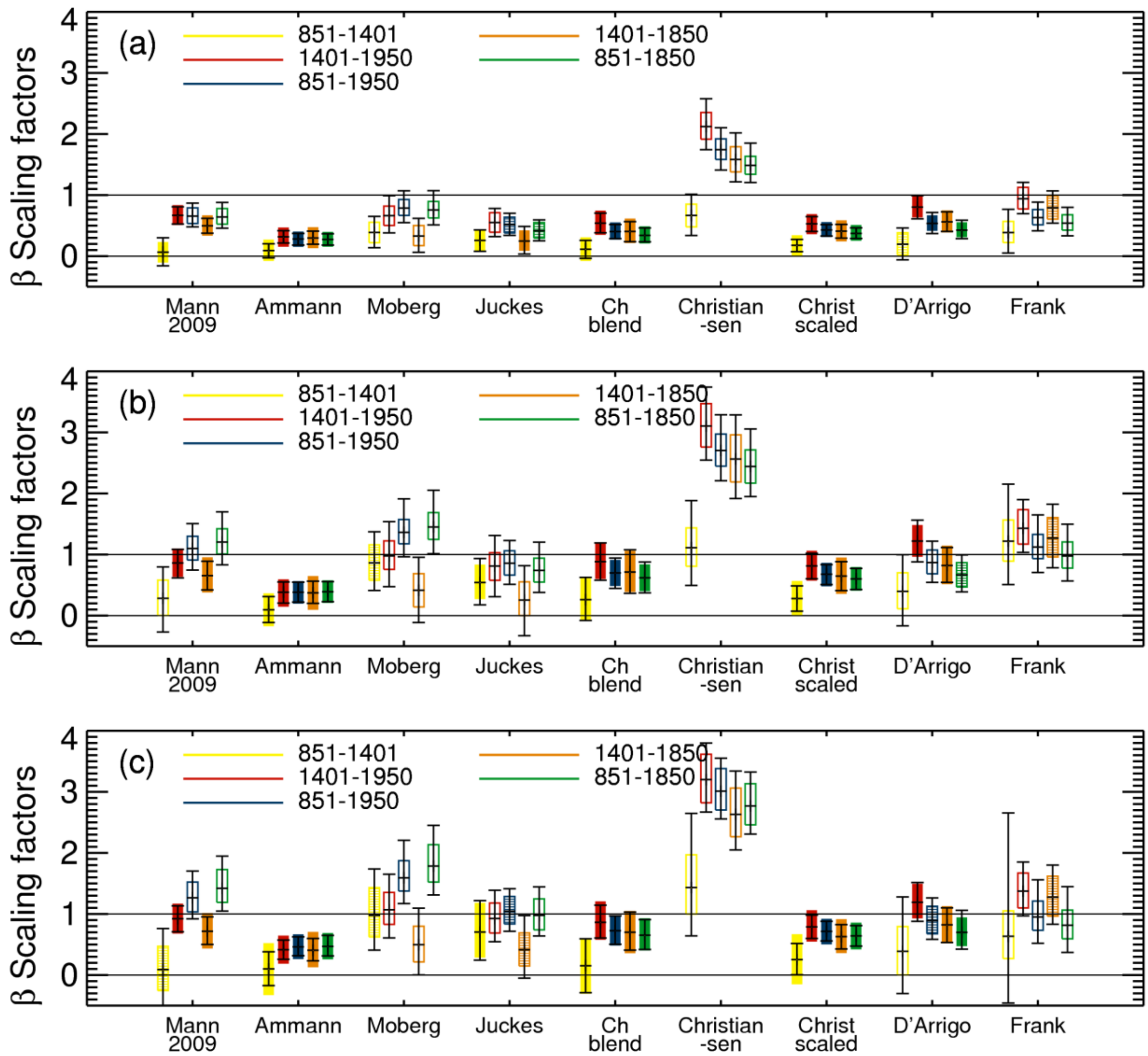

Figure 6 – **As figure 4b, but showing detection results for sensitivity tests**. a) Results for the standard analysis but without the extra 11 year boxcar smoothing. b) Results for the standard analysis but with 21 year boxcar smoothing instead of the

usual 11 year smoothing. c) Results for analysis with major volcanic eruptions masked out in both the reconstructions and model simulations.

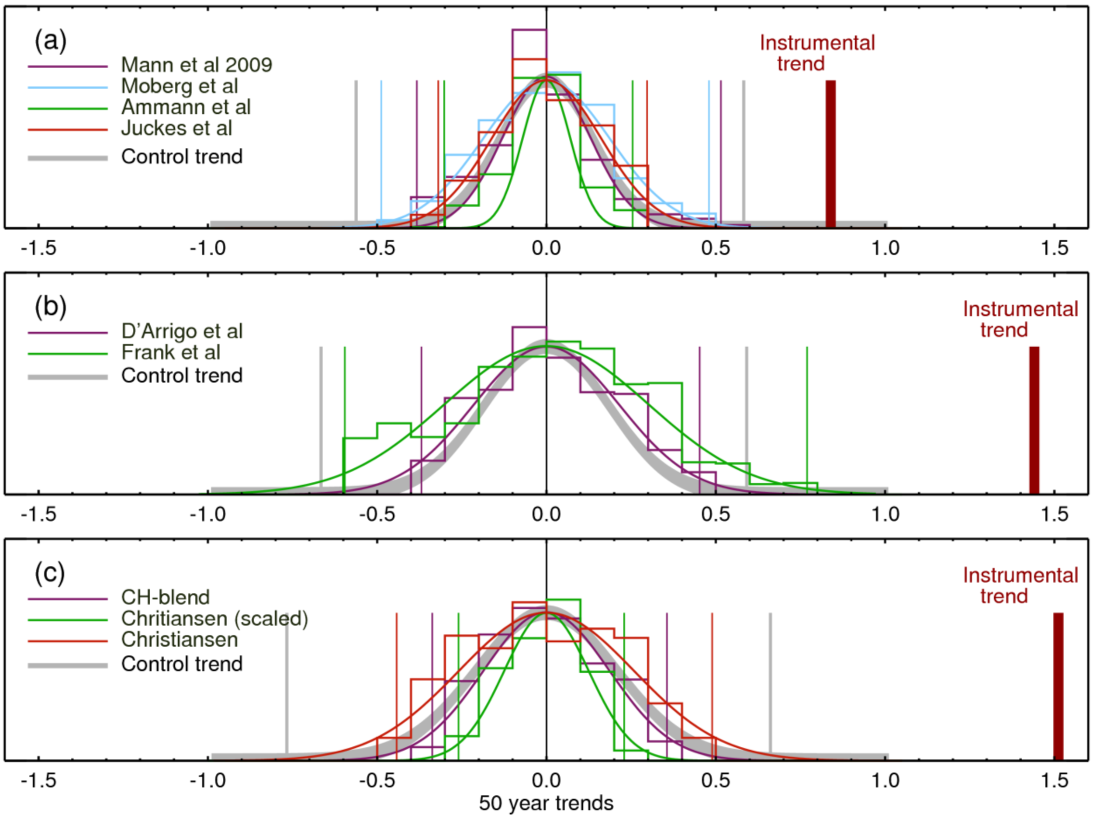

Figure 7 Distribution and maximum 50 year trends of internal variability estimated from reconstructions – a) *Distribution of 50 year trends found in the scaled residuals covering the time period 851-1950. The distribution of the trends is shown in the form of histograms with a Gaussian fit through the points. The grey shaded Gaussian shows the distribution of the 50 year trend found in the combined control simulations. The largest positive and negative 50 year trend from each reconstruction and the control simulations is shown by a bold vertical line. The recent 50 year trend (1960-2010) in the HadCRUT4 instrumental record (Morice et al. 2012) is shown by a burgundy vertical line. a) Results for NH mean SATs; b) for extra-tropics land only 20-90N and c) for extra-tropics land only 30-90N.*

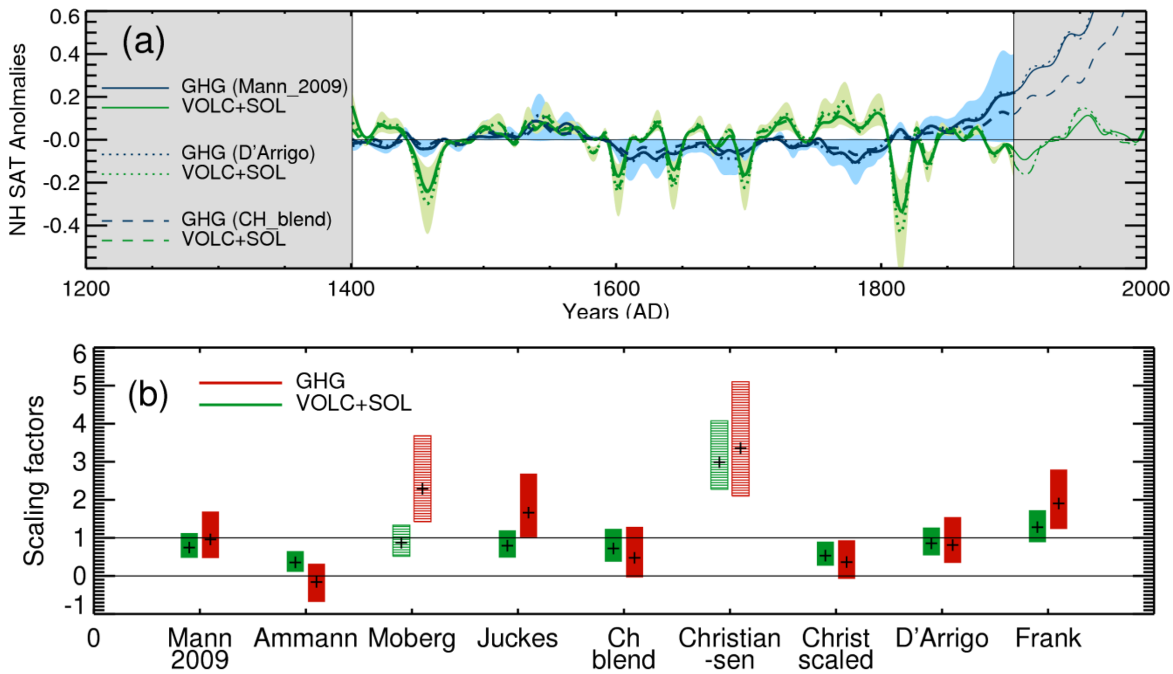

Figure 8 - Results from detection and attribution analysis using individually forced fingerprints. - *a) Individually forced fingerprints for solar and volcanic forcing combined (green) and greenhouse gas forcing (blue) scaled to fit three different reconstructions over the period 1400-1900 (white area of the plot), with the 5-95% scaling uncertainty range shown by the shaded region; b) Best fit scaling factors for both fingerprints for several reconstructions (cross) with 5-95% uncertainty range (vertical bar). Fingerprints are detectible if scaling factors are significantly above zero and consistent with the reconstruction if not significantly different from unity. Solid rectangle: the variability of the residual is smaller than ~90% of the control samples. Dashed rectangle: variability smaller than at least one control sample.*