Hadley Centre
for Climate Predictions and Research

Estimating signal amplitudes in optimal fingerprinting. Part II Application to general circulation models

by:

Perter A. Stott[1], Myles R. Allen[2], Gareth S. Jones[1]

[1]Hadley Centre for Climate Prediction and Research. The Met Office

[2]Rutherford Appleton Laboratory

HCTN 20 **October 2000**

Hadley Centre Technical Note

The Met.Office

Estimating signal amplitudes in optimal fingerprinting. Part II: Application to general circulation models.

Peter A. Stott[†], Myles R. Allen[‡], Gareth S. Jones[†]

[†] Hadley Centre for Climate Prediction and Research, The Met. Office

[‡] Rutherford Appleton Laboratory

Corresponding Author address :

Peter Stott

Hadley Centre for Climate Prediction and Research

The Met. Office

Bracknell, Berks, UK

RG12 2SY

Tel. +44 (0)1344 854011

Fax. +44 (0)1344 854898

e-mail pastott@meto.gov.uk

ABSTRACT

We show that there is a significant low bias in standard estimates of the amplitudes of climate change signals estimated by 4-member ensembles of the HadCM2 coupled ocean atmosphere general circulation model. This bias can be eliminated either by making larger ensembles of at least 8 members or by employing total least squares regression (tls) to take account of uncertainty in model-predicted signals.

Results using tls agree with previous work using ordinary least squares regression (ols) in showing that recent climate change is largely anthropogenic in origin. Consistent with previous results, there may be a non-zero amplitude of the solar signal in the observed record in the first half of the twentieth century. However the amplitudes of model-predicted signals in the observed record were previously underestimated by ordinary least squares regression. This implies that over the last 30 years the observations are consistent with a greater degree of greenhouse warming and sulphate cooling than previously thought and the early century warming is consistent with a much enhanced model response to solar changes with very little contribution from anthropogenic causes. However there are very large uncertainties in the weak solar signal and contributions of both anthropogenic and natural forcings to the early century warming are very poorly constrained.

1 Introduction

"Optimal detection", in which the amplitude of a model-predicted signal is estimated in the observational record, is a form of multi-variate regression. In searching for anthropogenic and natural signals in the twentieth century observational record of near-surface temperatures, Tett *et al* (1999) and Stott *et al* (2000) applied standard ordinary least squares ("ols") regression, adjusting variances by an ad-hoc factor to take approximate accout of uncertainties in model-predicted signals. Allen and Stott (2000) extended the methodology of Ripley and Thompson (1987) to explicitly take account of noise in the independent "predictor" variables in the multivariate problem of climate change detection where the noise has a complex autocorrelation structure. This is referred to as "tls" or total least squares regression.

Here we apply this methodology to patterns of near-surface temperature change from a climate model and from observations. Firstly, a "perfect model" study is carried out in which one member of a large ensemble of climate model simulations is taken to represent the observations and compared with a varying number of other model simulations. In this way, tls regression can be compared with ols regression in a controlled situation.

Secondly, we apply tls regression to the analysis of Tett *et al* (1999) and Stott *et al* (2000). Using ols regression, Tett *et al* (1999) showed that anthropogenic increases in greenhouse gases and sulphate aerosols are largely responsible for the observed temperature change since 1945. Their analysis showed that warming in the early part of the century can also only be explained when anthropogenic factors are included although they found that solar irradiance changes may have contributed. The authors used ols regression for consistency with earlier work (Hasselmann (1993); Hasselmann (1997); North *et al* (1995)), but they note that their estimates of pattern-amplitudes are biased towards zero by the sampling error in the simulated signals. This paper aims to address this limitation by recalculating the pattern amplitudes using the methodology outlined in Allen and Stott (2000).

In section 2 we discuss the climate model we have used and the analysis procedure that we have applied to process model data and produce the patterns whose amplitudes we are estimating. In section 3 we discuss the perfect model experiments and in section 4 we apply tls to the analysis of twentieth century near-surface temperature changes. Finally in section 5 we present our conclusions.

2 Model

We analyse spatio-temporal signals (or fingerprints) of climate change calculated by the second Hadley Centre coupled ocean-atmosphere general circulation model (HadCM2) (Johns *et al* (1997)). HadCM2 has a horizontal resolution of 2.5° in latitude by 3.75° in longitude with 19 atmospheric and 20 oceanic levels and uses prescribed adjustments to surface heat and water fluxes to the ocean. These flux adjustments depend only on location and time of year and are the same in control and perturbed climate simulations. The equilibrium climate sensitivity to a doubling of carbon dioxide of HadAM2 (the atmospheric component of HadCM2) coupled to a mixed-layer ocean is 4.1K (Senior and Mitchell (2000)). The effective climate sensitivity (Murphy (1995)) of HadCM2 to a doubling of CO2 changes in time, increasing from 2.6K at the time of doubling to 3.8K, 830 years after CO2 was stabilised (Senior and Mitchell (2000)). A detailed description of the model formulation is contained in Johns *et al* (1997).

In this paper, the spatio-temporal signals are calculated in the same way as Tett *et al* (1999) and Stott *et al* (2000). Decadal near-surface temperature variations are calculated for a succession of 50-year periods running from 1906-56 to 1946-96. Each 50 year period is analysed separately. Observed decadal means are calculated from a $5^o \times 5^o$ gridded dataset of near-surface temperature anomalies (Parker *et al* (1994)) by averaging annual means in decades where there were data for at least 8 out of 12 months in at least half the years. For each 50-year period, the time-mean was removed, and the data were filtered, using spherical harmonics, to retain only scales greater than 5,000 km; earlier work (Stott and Tett (1998)) had shown that we should expect recent climate change to project most strongly onto these scales.

The statistics of internal climate variability are estimated from a 1700 year control simulation (CTL) of HadCM2. Patterns with high variability estimated from CTL are given less weight by the optimal detection algorithm. For a more detailed discussion of the methodology, see Allen and Tett (1999) and Stott *et al* (2000).

HadCM2 has been used to simulate the climate's response to a number of forcing factors. In section 3 of this paper, we apply the optimal detection methodology to an ensemble of simulations with 9 members, all of which have been forced with changes in well-mixed greenhouse gases expressed as equivalent CO_2 and anthropogenic sulphate aerosols represented by changes in surface albedo as derived from the atmospheric chemistry model of Langner and Rodhe (1991) (see Mitchell *et al* (1995) and Mitchell and Johns (1997) for further details). In addition, in section 4 we consider the natural forcings due to changes in stratospheric volcanic aerosols using the series of Sato *et al* (1993) and changes in total solar irradiance using a timeseries of solar irradiance variations based on proxy data until 1991 (Hoyt and Schatten (1993) and extended to 1995 with satellite data (Willson (1997)).

3 Perfect model study

A 9-member ensemble of simulations has been carried out to simulate the response of the climate system to anthropogenic increases in greenhouse gases and sulphate aerosols. Each ensemble member has been started using initial conditions taken from CTL separated by 150 years.

We take each ensemble member in turn to represent the observations. We then compute all possible ensembles whose members do not include the simulation representing the observations; our ensembles have from 1 to 8 members. For a particular simulation representing the observations, the numbers of possible permutations are 8,28,56,70,56,28,8,1 for from 1 to 8-member ensembles respectively. Since there are 9 possible simulations to represent the observations, the total numbers of permutations are 72, 252, 504, 630, 504, 252, 72, 9 respectively.

Hypothesis testing and uncertainty intervals are calculated using independent data from that used to calculate the optimisation (Allen and Tett (1999)). When intra-ensemble variability calculated from the GS ensemble is used to calculate 5 to 95% uncertainty intervals, we would expect that 5% of permutations would have amplitudes significantly lower than 1 and 5% have amplitudes significantly greater than 1. For all the comparisons shown here, results are truncated to the first 10 eigenvectors. This was the truncation found by Tett *et al* (1999) for which all four signals passed the consistency test of Allen and Tett (1999) at the 10% level.

We now consider the case of single member ensembles selected from the 9 GS simulations, whose amplitudes we estimate in pseudo-observations represented by a different GS simulation. For each of the 9 possible observations, there are 8 possible single member ensembles and therefore a total of 72 permutations. Since each GS simulation is chosen in turn to represent the observations and is regressed against all the other GS simulations, the distribution of amplitudes should be symmetrically distributed about 1, since if $GS_i = \beta GS_j$, then $GS_j = (1/\beta)GS_i$. Using 5 to 95 percentile uncertainty intervals, Table 1 shows the number of occurrences for which ols regression gives amplitudes significantly greater and less than 1. There is a clear low bias with no instances of amplitudes significantly greater than 1 and overall 44% of cases having a signal amplitude significantly less than 1 compared to an expectation of 5%. The percentage of permutations significantly less than 1 decreases through the century from almost 60% in 1906-56 to 25% in 1946-96, as the strength of the GS signal increases. Total least squares regression shows no such low bias with identical numbers of cases significantly greater than 1 as significantly less than 1. Overall approximately 7% of permutations are significantly greater than 1 and 7% significantly less than 1, compared to an expectation of 5%, although there is considerable variation from 50-year period to 50-year period (from 2.8% to 11.1%). Since we only have available to us 9 members of the GS simulation, the overall distribution of all possible GS ensemble members is imperfectly sampled.

The bias introduced by ols regression is reduced if the number of ensemble members is increased to 4 (Table 1). Now 15% of cases have amplitudes significantly less than 1 and there are now a few cases (0.2%) whose amplitudes are significantly greater than 1. tls regression shows no systematic low or high bias with on average 7.9% of cases having amplitudes significantly less than 1 and 8.7% having amplitudes significantly greater than 1. Again with only 9 possible GS simulations to choose as pseudo-observations, imperfect sampling will affect the results.

Averaged over all possible permutations, ols regression underestimates the upper bounds and best estimates as calculated by tls regression for all periods considered and ensemble sizes from 1 to 8 (Fig 1a, b). The underestimate decreases through the century as the GS signal becomes stronger and also decreases as the number of ensemble members increases. For 8-member ensembles in the 1946-96 period ols and tls regression agree to within 10%. For 4-member ensembles upper bounds calculated using ols regression are from less than 50% to more than 80% of their value calculated using tls regression. The upper 50 to 95 percentile uncertainty ranges are underestimated more by ols than the lower 5 to 50 percentiles uncertainty ranges (Fig 1c, d). Ordinary least squares regression contains an ad-hoc correction (1+1/n where n is the ensemble size) for uncertainty in the signals which can lead to an overestimate of the uncertainties when the signal to noise is relatively high; this is the case for large ensemble sizes and later in the century (upper right part of Fig 1d).

4 Causes of twentieth-century temperature change near the Earth's surface

In this section, we address one of the caveats of Tett *et al* (1999) and Stott *et al* (2000) by taking account of the uncertainty in the model predicted signals in an analysis of twentieth century temperature change. We have carried out an analysis of decadal near-surface temperatures that is identical to that of Tett *et al* (1999) except in the single respect that we apply tls rather than ols regression in estimating the amplitudes of the model-predicted patterns in the observations.

The probabilities that signal-combinations are consistent with the observations are shown in Table 2. The results for the five 50-year periods starting in 1906, 1916, 1926, 1936 and 1946 agree with Tett *et al* (1999) in showing that neither internal variability nor any combination of entirely natural forcings can explain twentieth century near-surface temperature change. The sulphate, solar and volcanic signals in this analysis are too similar to each other for their individual pattern-amplitudes from any three-way or four-way regression to be meaningful since tests for degeneracy (Mardia *et al* (1979)) indicate we can estimate at most two quantities at a time in these diagnostics. Despite this degeneracy, the influence of greenhouse gases, either on its own or in combination with sulphate aerosols in **GS**, is detected in all combinations except the 4-signal combination **GHG**&**SUL**&**Sol**&**Vol**. In this case the greenhouse gas and sulphate signals are sufficiently degenerate with each other and with the two natural signals that neither anthropogenic signals are detected even though the anthropogenic signal **GS** is detected in combination with **Sol** and **Vol**.

We have also analysed the 90-year period from 1906-96 (Table 2). To retain the same amount of information for the 90 year analysis as was used in the 50-year analyses, the number of eigenvectors retained has been increased from 10 to 18. Both ols and tls analyses agree that internal variability alone is not consistent with the observations and greenhouse gases are detected in all one and two signal combinations in which it is included.

In the 1946-96 50-year period, both natural signals on their own fail the consistency test and the two anthropogenic signals can also be rejected since a solar signal is detected in combination with **GS** in 1906-56 and a sulphate signal is detected in combination with greenhouse gases in 1946-96. (Unlike the ols analysis, a sulphate signal is also detected in combination with greenhouse gases in 1916-56.) The same reasoning used by Tett *et al* (1999) to decide which signal combinations provided the most parsimonious and adequate explanations lead to the same conclusions. Three signals (**GHG SUL** and **Sol**) are detected during the century in combinations that are consistent with the observed record. This leaves three two-signal combinations; however in 1946-96 combinations including the influence of sulphate aerosols (**SUL**) fit the observations better than other explanations. We therefore consider further the same two two-signal combinations considered by Tett *et al* (1999), **GHG**&**SUL** and **GS**&**Sol**.

Although ols and tls analyses agree in ruling out an entirely natural explanation of twentieth century temperature change and in their choice of which signal combinations to consider, there are some major differences between the amplitudes and their uncertainties (Fig 2). Uncertainties are much larger using tls than ols with possible ranges for pattern amplitudes including infinite values, consistent with degeneracy between signal patterns. Both greenhouse gases and sulphate aerosols are detected in 1946-96, in agreement with results using ols regression. Also, like Tett *et al* (1999), a zero amplitude of solar is not consistent with the observations in 1906-56 (Fig 2), although the uncertainties are sufficiently large that they include unbounded amplitudes of the solar signal.

In the early part of the century, a reconstruction of the contributions to the observed temperature change (Fig 3 and Fig 4) shows that our two signal combinations lead to two contrasting explanations. Either the observations result from greenhouse warming somewhat greater than in the model with a slightly reduced sulphate cooling or, according to the tls analysis, the observed change results almost entirely from solar changes. In the latter case the solar signal is amplified by a factor of over five in order to agree with the observations. In contrast, the smaller solar amplitude calculated from ols regression leads to a much reduced solar contribution to the observed warming. In the latter half of the century both signal combinations show that the observed warming results from a combination of greenhouse warming and sulphate cooling, both rather greater with tls than ols regression, but balancing to give the observed warming.

Like Tett *et al* (1999) we have reconstructed 50-year trends of global mean near-surface temperature changes for the 1906-56 and 1946-96 periods. In the latter period, uncertainties in the reconstructed greenhouse gas and sulphate trends are considerably larger for tls than for ols (Fig 5a) with upper limits generally affected more than lower limits (eg the 90 and 95 percentiles of the greenhouse gas trends for 1946-96 (Fig 5a)). Nevertheless the same picture results from both analyses, of warming due to greenhouse gases being balanced by cooling due to sulphate aerosols, with both being detectable in 1946-1996. Degeneracy between these two signals means that it is hard to quantify the amplitude of each signal individually, even if some linear combination (such as **GHG+SUL**) may be well constrained (see Allen *et al* (2000)).

Reconstructed 50-year trends of global mean near-surface temperatures for GS&SOL are shown in (Fig 5b). In 1906-56, a zero amplitude of the solar signal is not consistent with the observations for either the ols or tls analysis. However uncertainties in the solar contribution to the warming during this period are much larger according to tls; the 10 to 90 percentile uncertainty range is entirely positive and includes trends as high as 0.4K/decade whereas the 5 to 95 percentile range is unconstrained. In the latter case amplitudes that span unbounded positive and negative values combine with a range of noise-reduced 50-year trends (before multiplying by their amplitudes) that includes a zero trend. In the most recent 50-year period tls regression gives a significantly positive warming of 0.07-0.21 K/decade due to greenhouse gases and sulphate aerosols, a larger range than the 0.05 - 0.13 K/decade from ols.

Stott *et al* (2000) detected a volcanic signal when they considered 10 years of annual mean data around the eruption of Mount Pinatubo in 1991. The volcanic signal was detected on its own and in all combinations involving greenhouses gases, sulphate aerosols and solar changes. No other signal was detected in any combination using this diagnostic. Using ordinary regression thay found that the model overestimated the volcanic signal when regressed on its own against the observations, giving an estimate of 0.1-0.68 (5 to 95 percentiles) with a best estimate of 0.39. They therefore concluded that the model appears to have overestimated the impact of this volcanic eruption on near-surface temperatures. However, it may be that ordinary regression is subject to a low bias and so has underestimated the true volcanic amplitude. We have therefore repeated the calculation using tls regression, obtaining a range of 0.21-0.72 with a best estimate of 0.46 (Table 3) and a similar range if **VOL** is combined with **GS**. Our results show that the volcanic signal in this diagnostic is not highly noise-contaminated since the volcanic amplitude is little changed, being significantly less than 1 with a best estimate of less than 0.5 using both estimation methods. The overestimate of the response of near-surface temperatures to the Pinatubo eruption may result from the forcing imposed in the model being too large or may indicate that the model is too sensitive to this particular forcing.

5 Conclusions

We have compared two regression schemes that have been applied in an optimal detection framework to fingerprints of twentieth century near-surface temperature change calculated by HadCM2. Ordinary least squares regression is the technique that is currently applied in standard optimal detection methodology (eg Hasselmann (1993), North and Kim (1995), Tett *et al* (1999)) whereas total least squares regression (unlike ols) takes account of noise in model-predicted signals. By means of a perfect model study, we have shown that ordinary least squares regression exhibits a significant bias towards zero at the signal to noise levels appropriate to 4-member ensembles of climate change signals in the twentieth century. Total least squares regression shows no high or low bias and produces results much closer to the expected distribution than ols regression, although at low signal to noise levels our results indicate that tls may underestimate uncertainty ranges.

We have examined the robustness of the conclusions of Tett *et al* (1999) concerning the causes of twentieth-century temperature change. With tls regression the main conclusions of Tett *et al* (1999) remain. We rule out explanations involving natural internal and externally forced natural variability. The same reasoning as applied by Tett *et al* (1999) yields the same two explanations for twentieth-century temperature change involving either greenhouse gases and sulphate aerosols in a combination which allows their relative amplitudes to vary or anthropogenic changes due to both greenhouse gases and sulphate aerosols in combination with solar changes. In the first combination, the observed warming results from a greater greenhouse warming and a smaller sulphate cooling than seen in the model and in the second combination the observed warming results from an enhanced response to solar changes. In the 1946-96 period we still detect greenhouse gases in all combinations except the degenerate four-signal combination **GHG**&**SUL**&**Sol**&**Vol**. However the anthropogenic signal, **GS**, is detected in combination with solar and volcanic forcings.

The principal attribution results of T99 are therefore robust to taking account of uncertainties in model signals. However, uncertainty ranges are generally much larger calculated by tls regression and for some signal amplitudes include unbounded values corresponding to noisy signals where the true underlying signal is near zero. Reconstructions of the contributions of these signals to observed changes shows a greater degree of greenhouse warming and sulphate cooling in the latter part of the century than shown by Tett *et al* (1999), while the observed changes in the early part of the century might be entirely consistent with warming solely from a much amplified solar signal.

The ideal solution is to make large ensembles of model runs to minimise noise contamination of the signal. Our perfect model studies indicate that at least 8 ensemble members are needed to avoid an appreciable low bias in the estimates of signal amplitudes. Until such large ensembles become available it is important to take account of uncertainty in model predicted signals. With 4-member ensembles it is therefore necessary to use total least squares regression for estimates of future uncertainty in climate prediction (Allen *et al*, 2000) which depend on accurate estimates of the strength of model-predicted signals in past climate.

References

Allen, M. R., and P. A. Stott, 2000: Estimating signal amplitudes in optimal fingerprinting. part I : theory. *Cli. Dyn.* In preparation.

——, and S. F. B. Tett, 1999: Checking for model consistency in optimal fingerprinting. *Cli. Dyn.*, **15**, 419–434.

——, P. A. Stott, R. Schnur, T. Delworth, and J. F. B. Mitchell, 2000: Uncertainty in forecasts of anthropogenic climate change. *Nature*. In Press.

Hasselmann, K., 1993: Optimal fingerprints for the detection of time-dependent climate change. *J. Clim.*, **6**, 1957–1971.

——, 1997: Multi-pattern fingerprint method for detection and attribution of climate change. *Cli. Dyn.*, **13**, 601–611.

Hoyt, D. V., and K. H. Schatten, 1993: A discussion of plausible solar irradiance variations, 1700-1992. *J. Geophys. Res.*, **98**, 18895–18906.

Johns, T. C., R. E. Carnell, J. F. Crossley, J. M. Gregory, J. F. B. Mitchell, C. A. Senior, S. F. B. Tett, and R. A. Wood, 1997: The second Hadley Centre coupled ocean-atmosphere GCM : model description, spinup and validation. *Cimate Dynamics*, **13**, 103–134.

Langner, J., and H. Rodhe, 1991: A global three-dimensional model of the tropospheric sulfur cycle. *J. Atmos. Chem*, **13**, 225–263.

Mardia, K. V., J. T. Kent, and J. M. Bibby, 1979: *Multivariate Analysis*. Academic Press.

Mitchell, J. F. B., and T. C. Johns, 1997: On modification of global warming by sulfate aerosols. *J. Climate*, **10**, 245–267.

——, ——, J. M. Gregory, and S. F. B. Tett, 1995: Climate response to increasing levels of greenhouse gases and sulphate aerosols. *Nature*, **376**, 501–504.

Murphy, J. M., 1995: Transient response of the Hadley Centre Coupled Ocean-Atmosphere model to increasing carbon dioxide. Part III : Analysis of global-mean response using simple models. *J. Climate*, **8**, 496–514.

North, G. R., and K.-Y. Kim, 1995: Detection of forced climate signals. Part II: simulation results. *J. Climate*, **8**, 409–417.

——, ——, S. S. P. Shen, and J. W. Hardin, 1995: Detection of forced climate signals. Part I: filter theory. *J. Climate*, **8**, 401–408.

Parker, D. E., P. D. Jones, C. K. Folland, and A. Bevan, 1994: Interdecadal changes of surface temperature since the late nineteenth century. *J. Geophys. Res*, **99**, 14373–14399.

Ripley, B. D., and M. Thompson, 1987: Regression techniques for the detection of an analytical bias. *Analyst*, **112**, 377–383.

Sato, M., J. E. Hansen, M. P. McCormick, and J. B. Pollack, 1993: Stratospheric aerosol optical depths (1850-1990). *J. Geophys. Res*, **98**, 22987–22994.

Senior, C. A., and J. F. B. Mitchell, 2000: The time dependence of climate sensitivity. *Geophys Res Lett*, **27**, 2686–2688.

Stott, P. A., and S. F. B. Tett, 1998: Scale-dependent detection of climate change. *J. Climate*, **11**, 3282–3294.

——, ——, G. S. Jones, M. R. Allen, W. J. Ingram, and J. F. B. Mitchell, 2000: Attribution of twentieth century temperature change to natural and anthropogenic causes. *Cli. Dyn.* In press.

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.

Willson, R. C., 1997: Total solar irradiance trend during solar cycles 21 and 22. *Science*, **277**, 1963–1965.

	1 member		4 members	
ols	< 1	> 1	< 1	> 1
1906-56	59.7	0.0	21.7	0.0
1916-66	51.4	0.0	24.8	0.5
1926-76	48.6	0.0	10.0	0.0
1936-86	36.1	0.0	5.2	0.5
1946-96	25.0	0.0	13.3	0.0
tls	< 1	> 1	< 1	> 1
1906-56	11.1	11.1	10.2	12.9
1916-66	11.1	11.1	10.8	16.8
1926-76	4.2	4.2	6.0	0.5
1936-86	2.8	2.8	2.1	11.0
1946-96	5.6	5.6	10.5	2.5

Table 1: Percentage of occurrences when ols and tls regression gives amplitudes significantly greater than and less than 1 for 1-member and 4-member ensembles.

	1906-56	1916-66	1946-96	1906-96	1906-96(ols)
Int. Var. †	**0.02**	**0.04**	**0.01**	**0.01**	**0.01**
GHG	0.26 $^{\mathbf{GHG}}$	0.09 $^{\mathbf{GHG}}$	0.07 $^{\mathbf{GHG}}$	0.29 $^{\mathbf{GHG}}$	0.32 $^{\mathbf{GHG}}$
GS	0.23 $^{\mathbf{GS}}$	0.56 $^{\mathbf{GS}}$	0.31 $^{\mathbf{GS}}$	0.37 $^{\mathbf{GS}}$	0.32 $^{\mathbf{GS}}$
Sol †	0.76 $^{\mathbf{Sol}}$	0.07	**0.05** $^{\mathbf{Sol}}$	0.11 $^{\mathbf{Sol}}$	**0.03**$^{\mathbf{Sol}}$
Vol †	0.1	0.11	**0.03**	0.07	**0.03**
GHG SUL	0.36 $^{\mathbf{GHG}}$	0.6 $^{\mathbf{GHG,SUL}}$	0.69 $^{\mathbf{GHG,SUL}}$	0.33 $^{\mathbf{GHG}}$	0.36 $^{\mathbf{GHG}}$
G Sol	0.7 $^{\mathbf{Sol}}$	0.68 $^{\mathbf{G,Sol}}$	0.06 $^{\mathbf{G}}$	0.25 $^{\mathbf{G}}$	0.29 $^{\mathbf{G}}$
G Vol	0.76 $^{\mathbf{G}}$	0.09	0.18 $^{\mathbf{G,Vol}}$	0.29 $^{\mathbf{G}}$	0.29 $^{\mathbf{G}}$
GS Sol	0.7 $^{\mathbf{Sol}}$	0.5 $^{\mathbf{GS}}$	0.34 $^{\mathbf{GS}}$	0.34 $^{\mathbf{GS}}$	0.28 $^{\mathbf{GS}}$
GS Vol	0.26 $^{\mathbf{GS}}$	0.52 $^{\mathbf{GS}}$	0.29 $^{\mathbf{GS}}$	0.38 $^{\mathbf{GS}}$	0.28 $^{\mathbf{GS}}$
Sol Vol †	0.75 $^{\mathbf{Sol}}$	0.51 $^{\mathbf{Sol,Vol}}$	**0.05** $^{\mathbf{Sol}}$	0.21 $^{\mathbf{Sol}}$	0.07 $^{\mathbf{Sol}}$
GHG SUL Sol	0.62 $^{\mathit{Sol}}$	0.79	0.61 $^{\mathit{GHG,SUL}}$	0.3 $^{\mathit{GHG}}$	0.32 $^{\mathit{GHG}}$
GHG SUL Vol	0.68 $^{\mathit{Vol}}$	0.58 $^{\mathit{GHG,SUL}}$	0.86 $^{\mathit{SUL}}$	0.37 $^{\mathit{GHG}}$	0.35 $^{\mathit{GHG}}$
GS Sol Vol	0.66 $^{\mathit{Sol}}$	0.66 $^{\mathit{Sol,Vol}}$	0.43 $^{\mathit{GS}}$	0.53 $^{\mathit{GS,Sol,Vol}}$	0.25 $^{\mathit{GS}}$
G Sol Vol	0.9	0.93 $^{\mathit{G,Sol}}$	0.77 $^{\mathit{G,Sol,Vol}}$	0.27	0.25 $^{\mathit{G}}$
GHG SUL Sol Vol	0.82 $^{\mathit{Vol}}$	0.91	0.93	0.52 $^{\mathit{GHG,SUL,Vol}}$	0.31 $^{\mathit{GHG,SUL}}$

Table 2: Consistency test using decadal-mean annual-mean data. Likelihood that best-fit signal combination is consistent with observations using an F-test with 21 d.o.f. Values in bold are inconsistent with the observations at the 5% level. For one and two signal combinations, a signal name is shown in bold superscript next to the P-value when its amplitude has an entirely positive range, equivalent to detecting that signal if that signal combination is consistent with the observations. Standard tests (see text) indicate that signal amplitudes in combinations including more than two signals are likely to be affected by degeneracy and in this case a signal name is shown in italics superscript when its amplitude has an entirely positive range. 1926-76 and 1936-86 are not shown since no signals are detected in these periods.

× Signal combination is an inadequate explanation of twentieth century temperature change as the F-test fails at least once.

ols	5%	50%	95%
1985-95 Vol on its own	0.10	0.39	0.68
1985-95 Vol in GS&Vol	0.07	0.40	0.72
tls	5%	50%	95%
1985-95 Vol on its own	0.21	0.46	0.72
1985-95 Vol in GS&Vol	0.15	0.48	0.87

Table 3: 5, 50 and 95 percentiles of distributions of signal amplitudes according to ols and tls regression.

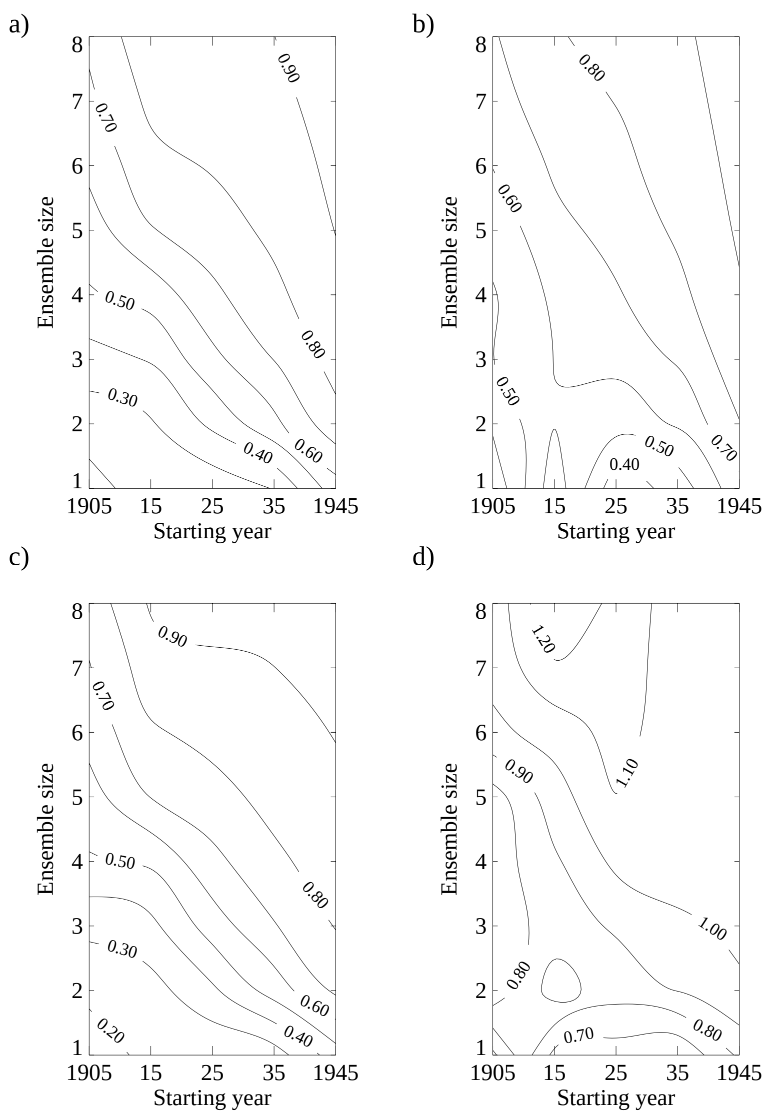

Figure 1: Fractional error in a) the 95 percentile of the distribution of signal amplitudes, b) the best estimate, c) the upper part of the 90% uncertainty range (50 to 95 percentile), d) the lower part of the 90% uncertainty range (5 to 50 percentile)

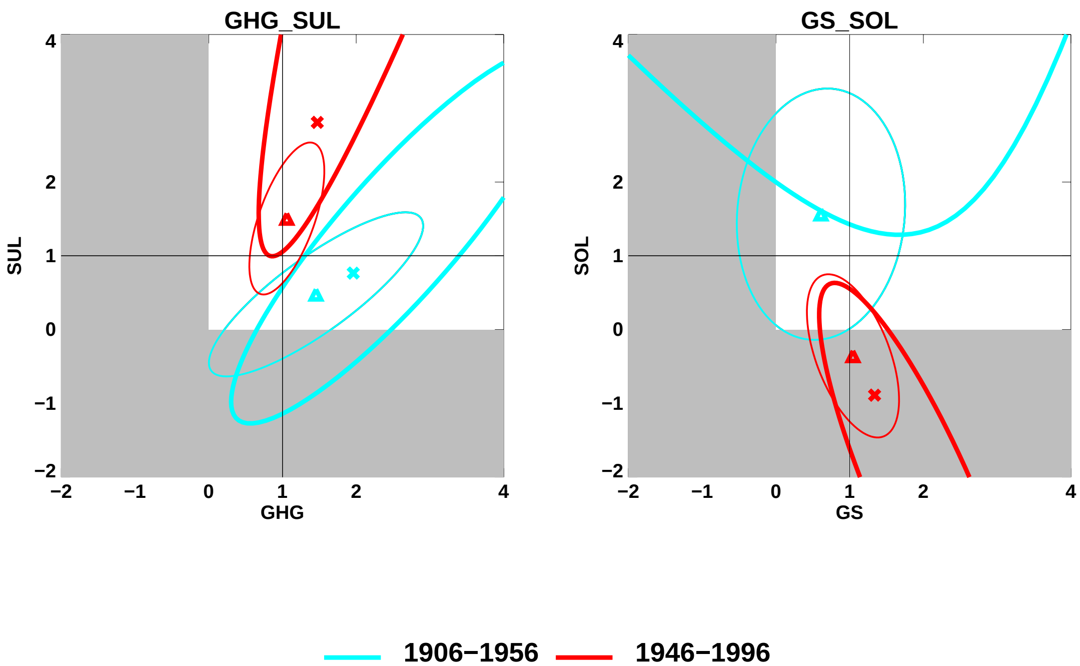

Figure 2: Ellipses containing 90% of the estimated joint distribution of the amplitudes of (a) GHG and SUL signals and (b) GS and SOL signals calculated using ols regression (thin lines) and tls regression (thick lines)

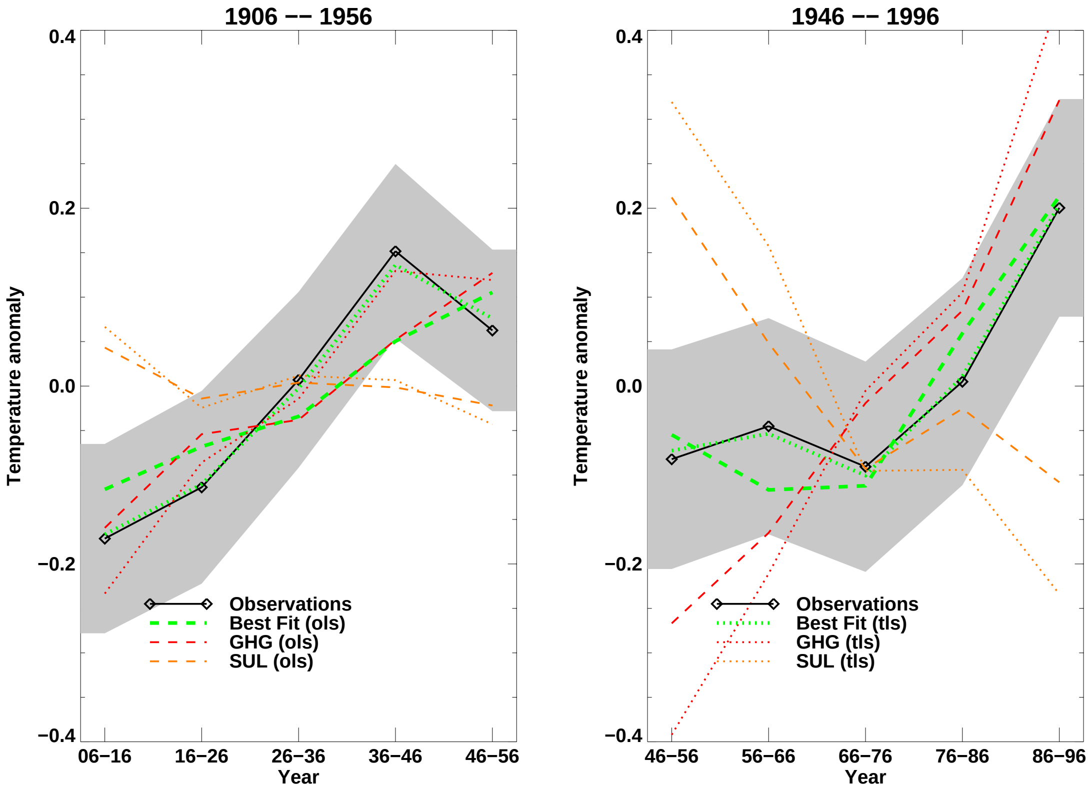

Figure 3: Best-estimate contributions to global-mean temperature change from GHG&SUL for a) 1906-56 and b) 1946-96. Black : observations. Green dashed (dotted) line : best fit from ols (tls). Red dashed (dotted) line : greenhouse gas contribution from ols (tls). Orange dashed (dotted) line : sulphate contribution from ols (tls).

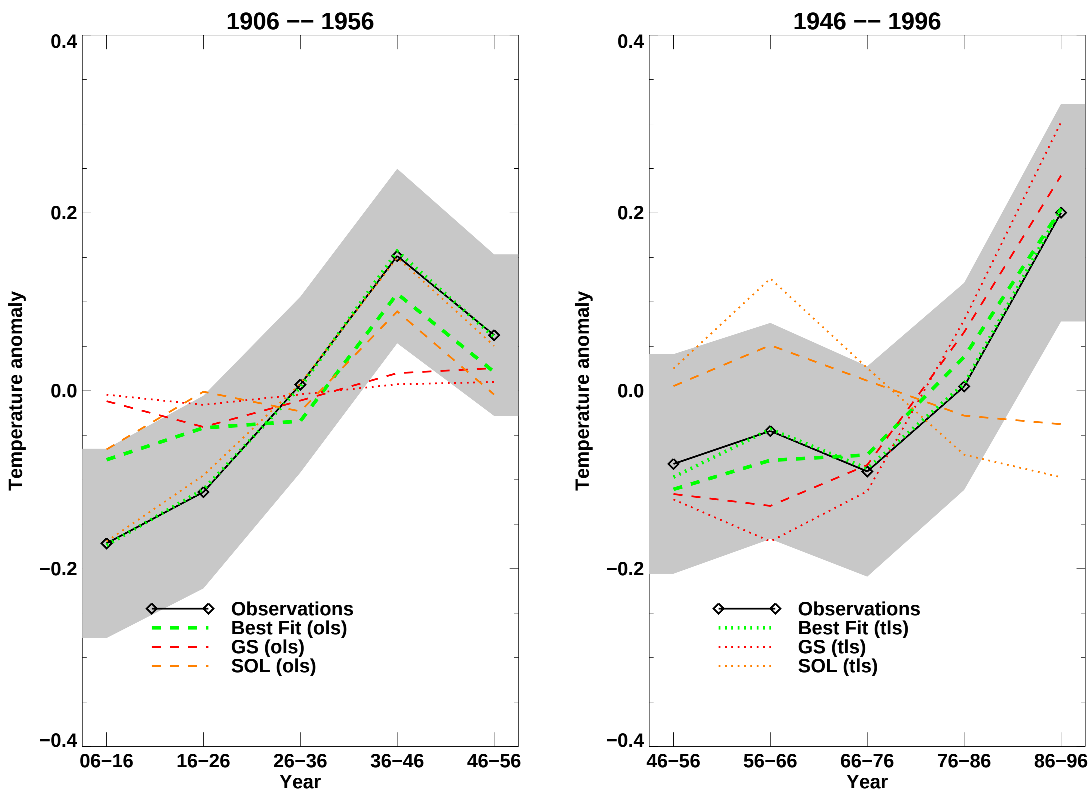

Figure 4: Best-estimate contributions to global-mean temperature change from GS&SOL for a) 1906-56 and b) 1946-96. Black : observations. Green dashed (dotted) line : best fit from ols (tls). Red dashed (dotted) line : GS contribution from ols (tls). Orange dashed (dotted) line : solar contribution from ols (tls).

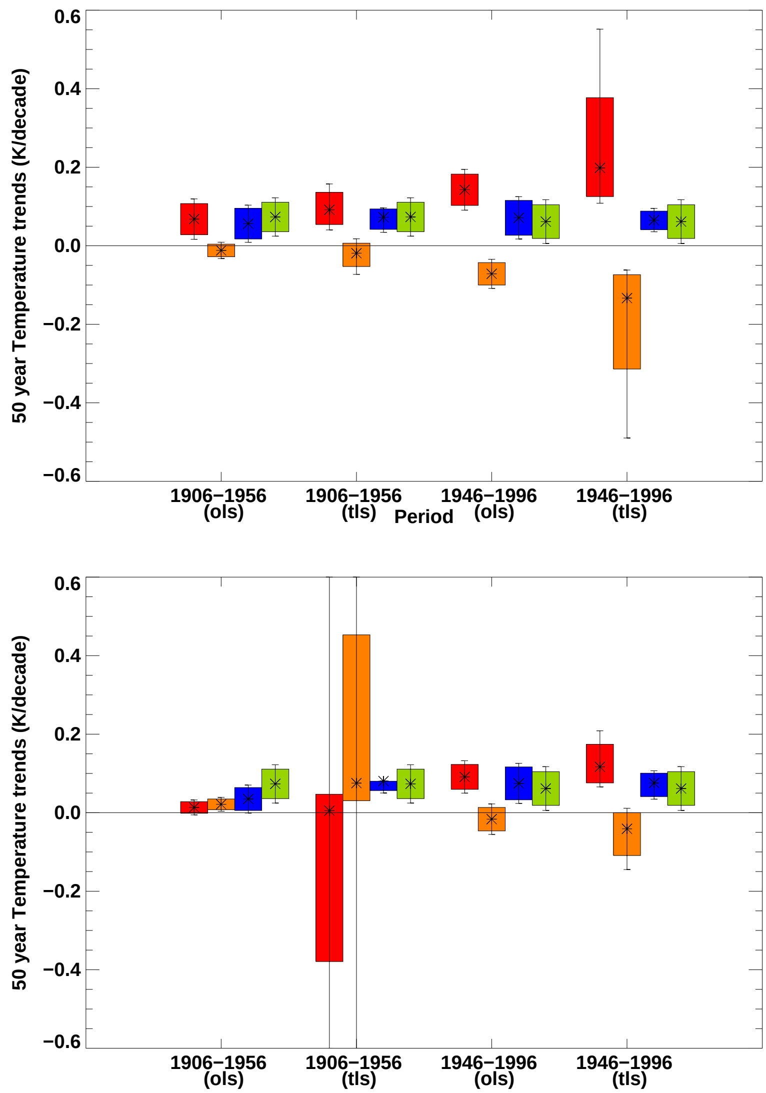

Figure 5: a)50-year global mean temperature trends and uncertainties reconstructed from GHG (red bars), SUL (orange bars), total (blue bars) and the observations (green bars), b) 50-year global mean temperature trends and uncertainties reconstructed from GS (red), SOL (orange), total (blue) and the observations (green). Best estimates shown as stars, 10-90 percentiles of distributions as coloured bars and 5-95 percentiles as lines.