Supplemental Information

SI-1 - Forecasting Sulfur Emissions

We obtain the share of 2000 global sulfur emissions for six categories of sulfur emitting activities. These activities and their share of 2000 emissions as calculated by Smith *et al* (1) are given in table S1.

These values are used to generate the quantity of sulfur emissions by activity in 2000 as calculated by Stern (2). This series is used because it is the one used to estimate the original model (3).

SI-1.1 Coal Consumption We obtain annual data on global and Chinese coal consumption from the Energy Information Administration (4). We calculate an index for sulfur emissions per unit coal consumption (*IntensitySO2*) as follows:

$$\mathrm{IntensitySO2_t} = (1 - \%\mathrm{China_t}) + ((\%\mathrm{China_t}) \times \mathrm{ESO2_t})$$

in which $\%China_t$ is the fraction of global coal consumption burned in China and $ESO2_t$ is the ratio of SO_2 emissions per kWh of electricity generated by coal fired stations in China relative to the rest of the world, both in year *t*. The value for the rest of the world is proxied by the SO_2 emission rate per coal fired kWh in the US. Observations for China are available from 1998 - 2007 from Xu (5, 6). Data for the US are calculated by dividing SO_2 emissions from the US electric power sector (7) by the quantity of electricity generated from coal-fired stations (8).

This index, along with information about global coal consumption (4) are used to calculate sulfur emissions from coal consumption (*CoalSO2*) as follows:

$$\mathrm{CoalSO2}_t = \mathrm{CoalSO2}_{t-1} \times \frac{\mathrm{IntensitySO2}_t}{\mathrm{IntensitySO2}_{t-1}} \times \frac{\mathrm{WorldCoal}_t}{\mathrm{WorldCoal}_{t-1}}$$

in which Worldcoal is global coal consumption.

SI-1.2 Smelting We obtain annual data for the global production of copper, zinc, lead, and nickel (9). Lefohn *et al.* (10) report information on the tons of sulfur emitted per ton of copper (1.2), zinc (0.5), lead (0.14), and nickel (1.2) produced. These data are used to create an estimate for the quantity of sulfur emitted by the production of these four metals (*Metal*) as follows:

$$Metal_t = \sum_{i=1}^{4} Q_{it} E_i$$

in which Q_{it} is the quantity of metal i produced in year t and E_i is the quantity of sulfur emitted per ton of metal i produced (values in parentheses above). The index *Metal* is used to forecast sulfur emissions due to smelting as follows:

$$SmeltS_t = SmeltS_{t-1} * (Metal_t / Metal_{t-1}) * (1 - Eff)$$

in which $SmeltS_t$ is the quantity of sulfur emitted by smelting in year t and *Eff* is a measure for the annual increase in the fraction of sulfur scrubbed from the waste stream (or removed from the waste stream by pre-processing) per unit of economic activity . A value for *Eff* of 0.08 is chosen based on a methodology that is described in the next section. This value for *Eff* represents an 8 percent annual increase in the fraction of sulfur removed per unit of economic activity.

SI-1.3 Bunker fuels The quantity of sulfur emitted from the burning of marine bunker fuels (*BunkerSO2*) is calculated as follows:

$$\mathrm{BunkerSO2}_t = \mathrm{BunkerSO2}_{t-1} \times \frac{\mathrm{Bunker}_t}{\mathrm{Bunker}_{t-1}} \times (1-\mathrm{Eff})$$

in which $Bunker_t$ is the quantity of marine bunker fuel consumed globally (11).

SI-1.4 Petroleum The quantity of sulfur emitted from the combustion of petroleum (*PetrolSO2*) is calculated as follows:

$$\mathrm{PetrolSO2}_t = \mathrm{PetrolSO2}_{t-1} \times \frac{\mathrm{Petrol}_t}{\mathrm{Petrol}_{t-1}} \times (1-\mathrm{Eff})$$

in which *Petrol* is the quantity of petroleum consumed globally (12).

SI-1.5 Natural Gas The quantity of sulfur emitted from the combustion of natural gas (*NGasSO2*) is calculated as follows:

$$\mathrm{NGasSO2}_t = \mathrm{NGasSO2}_{t-1} \times \frac{\mathrm{NGas}_t}{\mathrm{NGas}_{t-1}} \times (1-\mathrm{Eff})$$

in which $NGas_t$ is the quantity of natural gas consumed globally (13).

SI-1.6 Other The quantity of sulfur emitted from other activities includes land-use change, other industrial processes, and traditional biomass. The quantity of sulfur emitted from these activities (*OtherSO2*) is calculated as follows:

$$OtherSO2_t = OtherSO2_{t-1}$$

We leave these emissions constant at the 2000 level based on the notion that any increase related to an increase in economic activity or lower levels of emissions per unit activity as measured by *Eff* will offset population growth. This estimate is likely to understate sulfur emissions.

SI-1.7 Total Emissions of Sulfur is calculated from the following product:

$$TotSO2_t = TotSO2_{t-1} \times \left(\frac{CoalSO2_t + PetSO2_t + SmeltSO2_t + BunkSO2_t + NgasSO2_t + OthSO2_t}{CoalSO2_{t-1} + PetSO2_{t-1} + SmeltSO2_{t-1} + BunkSO2_{t-1} + NgasSO2_{t-1} + OthSO2_{t-1}} \right)$$

Global emissions (*TotSO2*) are converted to radiative forcing using formulae from Kattenburg (14), which include both direct and indirect effects.

SI-1.8 Validating the Methodology To validate the methodology for calculating sulfur emissions, we assemble the time series described above for 1990- 2000, and use the methodology to calculate sulfur emissions from 1991- 2000. These values are compared to values reported by Stern (2).

Efforts to abate sulfur emissions per unit of economic activity, as measured by *Eff*, vary among nations and activities. Values are available for some nations in some sectors. For example, in the US, the amount of sulfur emitted per unit of coal burned in the electric power sector declines about 5.5 percent per year between 1996 and 2007.

To choose a value for *Eff*, we use a range of values for *Eff* to generate in-sample simulations for global sulfur emissions between 1991 and 2000. To choose a value, we regress the in-sample simulation against the observed value as follows:

$$S_t = \gamma \hat{S}_t + \varepsilon_t$$

in which S_t is the value for anthropogenic sulfur emissions calculated by Stern (2) for year t and $\hat{S}$ is the value calculated using the methodology described above. We estimate the above regression using values of *Eff* that range between 0 and 0.10. We choose the value of *Eff* under which the OLS estimate for γ is closest to 1, which would be the expected point value of γ for the value of *Eff* that generates changes in sulfur emissions that most closely match observed changes between 1991 and 2000 (ie. a one-to-one correspondence) (Table S2).

Based on these results, we choose a value of 0.08. Sulfur emissions (post 2000) calculated using this range of values for *Eff* are given in Figure S-1, in which SOX (Black line) are the values of sulfur emissions calculated by Stern (2) and Eff is equal to 0.08 (Red line). The increase in sulfur emissions after 2003 that is generated by all scenarios is consistent with the results generated in a detailed country-by-country analysis through 2005 (15).

SI-2 Statistical Results

We follow the statistical methodology for estimating global mean surface temperature described by Kaufmann *et al*. (3). The long-run cointegrating relationship between the aggregate of radiative forcing (greenhouse gases, sulfur emissions, and solar insolation) (*RFAGG*) and global surface temperature (*Temp*) is estimated from the following equation using dynamic ordinary least squares (16).

$$Temp_t = \alpha + \beta_1 \; RFAGG_t + u_t \qquad \text{(S-1)}$$

The rate at which temperature adjusts to changes in radiative forcing and the short run responses to changes in the Southern Oscillation Index (SOI) and the forcing of volcanic sulfates (RFVol) is estimated using the following error correction model.

$$\Delta Temp_t = \theta + \beta_2 \hat{u}_{t-1} + \sum_{k=1}^{s} \delta_k \Delta Temp_{t-i} + \sum_{k=1}^{s} \phi_k \Delta RFAGG_{t-k} + \sum_{k=0}^{s} \pi_k SOI_{t-k} + \sum_{k=0}^{s} \zeta_k RFVol_{t-k} + \varepsilon_t \quad \text{(S-2)}$$

in which $\hat{u}_t$ is the estimated disequilibrium between observed temperature and the equilibrium implied by the long-run cointegrating relationship ($\hat{u}_t$ = $Temp_t$ – ($\hat{\alpha}$ + $\hat{\beta}_1$ $RFAGG_t$). The appropriate lag length (s) for equation (S-2) is chosen using the Akaike Information criterion (AIC) (17) and the equation is estimated using ordinary least squares (OLS). Results for three sample periods are given in Table S3. Previous efforts (3) indicate that the effect of the North Atlantic Oscillation on global surface temperature is not statistically different from zero and so is not included.

SI-2.1 In Sample vs. Out-of-Sample Figure S-2 represents the model's ability to simulate global surface temperature (Table S3). The model is initialized with data that start in 1866; no additional information about global surface temperature is given to the model after 1870, which is the first year of the in-sample simulation. Notice that simulated temperature (orange line) captures short and long term movements in observed temperature (black line) and there is no noticeable difference in the in-sample [1870-1998] simulation (orange), and out-of-sample [1999-2008] simulation (purple). A statistical test for differences between these periods is described below.

By the end of the forecast, there is little difference between the simulations that starts in 1870 and 1999. For example, the simulation that starts in 1870 (orange) simulates a global surface temperature of 0.462^{o} C in 2008; the simulation (based on the same estimation period) that starts in 1999 (purple) is 0.484^{o} C. This convergence is caused by the model's tendency to move towards the equilibrium that is implied by the long-run cointegrating relationship and the transitory effects of stationary variables like ENSO events or volcanic sulfates.

SI-2.2 Error Estimates The simulated 95% error bars for the temperature forecasts (Figures 2 & 3) represent estimation uncertainty of the regression coefficients. Measurements of greenhouse gas concentrations, solar insolation, and volcanic forcings are relatively certain compared to anthropogenic sulfur emissions. Uncertainty about this forcing is examined separately in section SI –2.4.

To simulate temperature error bars for each forecast, we augment model coefficients (both the cointegrating relationship and the error correction model) with errors drawn from the asymptotic normal distributions of the corresponding estimators. Disturbances to the equations are set to zero for the simulations so that the uncertainty bands represents estimation (sampling) uncertainty only, not random disturbance uncertainty. In other words, the error bars represent sampling uncertainty of the conditional mean of temperature, given the radiative forcing components. These bound the 2.5th and 97.5th percentiles of 1000 simulations.

SI-2.3 Cointegration Breakdown We use a methodology developed by Andrews and Kim (19) to test whether the long-run cointegrating relationship between global temperature and radiative forcing (greenhouse gas concentrations, sulfur emissions, and solar insolation) changes during the 1999-2005 (the last period for which the coinegrating relation can be estimated) period relative to the relationship estimated with data from the sample period that ends in 1998. To do so, we estimate the model with data from sample periods that start in 1864, 1920, and 1960, and end in 2005, and use the results to calculate the residual from the cointegrating relationship ($\hat{u}_t$). This residual is used to calculate the *R* test statistic as follows:

$$R = \sum_{t=T+1}^{T+m} \left(\sum_{s=t}^{T+m} \hat{u}_t\right)^2 \qquad \text{(S-3)}$$

in which *T* = 1998 and *m* is seven (representing 1999-2005). The *R* statistic tests for a breakdown in the cointegrating relationship during the 1999-2005 period. The test statistic is evaluated against an asymptotic null distribution that is generated from seven-year sub-samples, the first of which ends seven years after the start date and the last ends in 1998. These values are ranked by size and the value at the 95 percentile is used as the critical value. The *R* statistic fails to exceed the critical value for sample periods that start in 1864 and 1920 (Table S3), which indicates that we cannot reject the null hypothesis that the cointegrating relationship is stable throughout against the alternative that the cointegrating relationship breaks down during the 1999-2005 period. Conversely, we reject the null hypothesis for the sample period that starts in 1960. Because the power of the test depends on the size of *m* and *T*, the failure to reject the null hypothesis should not be interpreted as strong evidence in favor of stable cointegration (19).

SI-2.4 Uncertainty About Anthropogenic Sulfur forcing Uncertainty about anthropogenic sulfur forcing arises from two sources; uncertainty about emissions and uncertainty about the formula used to translate emissions to radiative forcing. One way to evaluate this uncertainty is to compare our estimate for the radiative forcing with published estimates. For example, our method for forecasting sulfur emission in 2005 and converting to those emissions to radiative forcing generate values of the direct (-.26

W m^{-2}) and indirect (-0.73 W m^{-2}) effect. Although generated using a very different methodology, our values are close to mean values for the direct (-.4 $^{\pm}$0.2 W m^{-2}) and indirect (-0.7 with a 5 to 95% range of –0.3 to –1.8 W m^{-2}) effects published by (21).

In addition to this similarity, we investigate how both sources of uncertainty may affect the estimate for the statistical model and the out-of-sample forecast. To evaluate the effect of uncertainty about emissions, we add a normally distributed random error to each year's point estimate for anthropogenic sulfur emissions, recalculate total radiative forcing, and use the new time series to re-estimate the model. Increasing uncertainty will diminish the statistical model's ability to match the stochastic trend in global surface temperature to radiative forcing. This diminution is evaluated by testing for cointegration between the time series for temperature and radiative forcing, which includes uncertainty about anthropogenic sulfur forcing. To quantify this effect, we generate 1,000 experimental data sets for each of three sample periods (1860-1998; 1920-1998, 1960-1998) and seven magnitudes of uncertainty ($\pm$10%, $\pm$15%, $\pm$20%, $\pm$25% $\pm$33%, $\pm$40%, $\pm$50%). A finding of cointegration for 950 of the 1000 experimental data sets indicates that a given level of uncertainty about anthropogenic sulfur forcing does not have a statistically measurable effect on the model's ability to detect a relationship between global surface temperature and radiative forcing (Table S4).

Results in Table S4 indicate that uncertainty about anthropogenic sulfur emissions has a relatively small effect on the model's ability to detect a statistically meaningful relationship between global surface temperature and radiative forcing. For the full sample period, errors ±25% do not diminish the model's ability to detect cointegration. Only when the errors become large (>±33%) or the sample size becomes small (38) does uncertainty about anthropogenic sulfur emissions interfere with the statistical model's ability to detect cointegration. Table S-4 reports rates at which the Augmented Dickey-Fuller test rejects the null hypothesis of non-cointegration. As the sample size decreases, the power of this test also decreases and detecting cointegration is less likely; this explains the pattern in Table S-4 in which cointegration is found less frequently the shorter the sample for a given level of uncertainty about anthropogenic sulfur forcing. In fact, if sulfur forcings include a substantial estimation error, our statistical analysis would be unlikely to detect cointegration.

To evaluate the degree to which uncertainty about the method used to calculate anthropogenic sulfur emissions for 2001-2008 affects the forecast for global surface temperature, we focus on uncertainty about *Eff*, which is the most uncertain determinant of our measure for emissions. To do so, we generate out-of-sample forecasts for surface temperature (1999-2008) with time series for anthropogenic sulfur emissions that are generated using values of *Eff* equal to 0.05 (green), 0.06 (light blue), 0.7 (purple), 0.08 (orange), 0.09 (red), and 0.10 (grey) (Figure S-3). By 2008, these different assumptions about *Eff* generate a 6 percent difference in radiative forcing due to anthropogenic sulfur emissions between the high and low removal scenarios (Figure S-1).

This range of values generates temperature values that fall within the 95% confidence interval (Figure S-3) of simulations generated using a value of *Eff* equal to 0.08 (orange). This result implies that uncertainty about the rate at which sulfur emissions are removed from the emission stream has little effect on conclusions regarding the model's ability to simulate the post 1998 pattern of observed global surface temperature.

To evaluate the degree to which uncertainty about the formulae used to convert anthropogenic sulfur emissions to radiative forcing affects the statistical results, we scale the annual point estimate for radiative forcing by values between 0.0 and 2.4 ($RFSOX_{Modified} = \lambda RFSOX$). These values imply a range of uncertainty about our calculation for total aerosol forcing in 2005 (-0.99) that is larger than the 95 percent confidence interval (-0.5 Wm^{-2} - -2.4 Wm^{-2}) for 2005 that is described by (21). These altered values for forcing are added to the other forcings and the statistical model is restimated through 1998. The effect on the statistical results (Table S5) is evaluated using three metrics; (1) an ADF statistic that tests for cointegration between the surface temperature and radiative forcing, (2) the statistical significance of the estimate for the long-run relationship between radiative forcing and surface temperature (β_1 in equation eq. (S-1)) and (3) the statistical significance of the error correction mechanism (β_2 in equation (S-2)) that represents how surface temperature adjusts to disequilibrium in the long-run cointegrating relation. Failure to reject the null hypothesis associated with any one of these three metrics would indicate that uncertainty about the formulae used to translate anthropogenic sulfur emissions to radiative forcing disrupts the statistical model's ability to quantify the relationship between radiative forcing and global surface temperature.

For the 1864-1998 and 1920-1998 sample periods, both the ADF statistic and the statistical significance of β_2 in equation (S-2) suggest that the statistical model would be disrupted if the actual forcing is 20 to 30 percent greater ($\lambda > 1.2 - 1.3$) than that indicated by the formulae. For the more recent period, in which the data are more accurate, the model performs well even if the actual forcing is 110 percent greater than indicated by the formula. Conversely, both the ADF statistic and the statistical significance of β_2 in equation S-2 suggest that if the statistical model would be disrupted if the actual forcing is 40 percent ($\lambda < 0.6$) or more less than indicated by the formulae.

To evaluate the degree to which uncertainty about the formulae used to convert anthropogenic sulfur emissions to radiative forcing, we take the extreme values for λ which generate statistically 'acceptable' models for the 1960-1998 period from table S-5 ($\lambda = 0.6$; $\lambda = 2.1$), estimate the model, and use the regression results to simulate temperature for 1999-2008. Simulations replicate the general pattern of observed temperature and fall well within the 95 confidence interval (Figure S-4). Together, these results suggest that uncertainty about the formulae used to translate anthropogenic sulfur emissions to radiative forcing has a relatively small effect on the statistical model's ability to quantify the relationship between radiative forcing and global surface temperature.

To evaluate the degree to which the estimation period within the period for which direct measurements of greenhouse gases are available (or changes in the reliability of the temperature data) affects the simulation of global temperatures, we generate out-of-sample forecasts for three additional sample periods: 1960-1990, 1960-1995, and 1960-2000 (Figure S-5). As expected the accuracy of the out-of-sample estimations improves with greater sample size. Nonetheless, the choice of sample period has relatively little effect on the models' out-of-sample temperature forecast for 1999-2008.

Figure S-6 illustrates the effect of the 'spin-up' date on the out-of-sample forecast. To evaluate this source of uncertainty, we start model simulations in 1900, 1925, 1950, 1975, 1985, and 1995. Regardless of year in which the simulation starts, the forecasts for 1999-2008 are similar, which indicates that the date at which the model starts has little effect on the out-of-sample forecast for 1999-2008.

SI-2.5 Alternative measure of surface temperature To evaluate the degree to which the results are sensitive to the measure of temperature, we repeat the analysis with GISS temperature data (22), which start in 1880. As indicated by Table S6, the results are essentially unchanged; (1) temperature cointegrates with radiative forcing, (2) the long run relation between temperature and radiative forcing can be measured with a high degree of statistical precision and, (3) disequilibrium in the long-run relation between temperature and radiative forcing moves temperature towards the long-run value that is implied by radiative forcing. Furthermore, the point estimates (β_1 and β_2) for these effects are similar (See Table S3). Together, these results suggest that the time series used to measure global surface temperature has little effect on the results.

A reviewer suggests that we repeat the analysis with the GISS data because the measures of temperature differ after 1998. Specifically, the CRU temperature data peak in 1998 while the GISS temperature data peak in 2005. To evaluate the degree to which these differences are meaningful, we test whether the two temperature series cointegrate and whether the cointegrating relationship between the two temperature series breaks down after 1998 (using the R statistic—equation (S-3)).

For the three sample periods described in Tables S3 and S6, the Cru and GISS measures of temperature cointegrate and we fail to reject the null hypothesis that the cointegrating relationship is stable throughout over the entire sample period, against the alternative alternative hypothesis that the cointegrating relationship breaks down during the 1999-2005 period (See Table S7). This implies that the differences between the two temperature series are not statistically meaningful and therefore the temperature series used should not have a significant effect on the statistical results, a hypothesis that is consistent with the similarity of results in Tables S3 and S6.

SI-2.6 Cointegration & Omitted Variable We recognize that our measure of radiative forcing does not include some important components, such as black carbon. But the results of the statistical model do not depend on energy balance, as described by (23). Rather, the statistical model focuses on the non-stationary changes in the independent and dependent variables and it seeks to determine whether non-stationary changes in radiative forcing match non-stationary changes in global surface temperature. These nonstationary changes constitute a 'fingerprint' that can be used to determine whether the relationship between temperature and radiative forcing is statistically meaningful. The degree to which non-stationary changes in temperature match non-stationary changes in radiative forcing is evaluated by the statistical notion of cointegration.

Cointegration between surface temperature and radiative forcing indicates that the omission of a forcing(s) (e.g. black carbon) does not diminish the statistical model. This finding implies that either; (1) the forcing(s) omitted from the statistical model is small, (2) the forcing(s) omitted from the statistical model is stationary, (3) the forcing(s) omitted from the statistical model shares the same stochastic trend as forcings that are included in the model. In the case of black carbon, the first hypothesis is unlikely—black carbon has forcing, 0.9 W/m^2, with a range of 0.4 to 1.2 W/m^2 (24). The second hypothesis cannot be tested directly because there is no annual time series for black carbon that overlaps the sample period. Data are available at ten-year intervals between 1850 and 2000, but the spacing interferes with tests designed to detect stochastic trends. Despite this limit, we use the decadal time series to evaluate the degree to which the omission of black carbon affects the results—see next section. The third hypothesis is consistent with observations that up to one third of black carbon emissions are associated with the combustion of fossil fuels, which also emit sulfur, and are therefore correlated with sulfate aerosols (25).

SI-2.7 Stratospheric Water Vapour, Black Carbon, & Omitted Variable Bias To evaluate the degree to which omitted variables, such as stratospheric water vapour and black carbon, affect the statistical results, we test whether errors from the statistical model are related to stratospheric water vapour or black carbon. If the statistical model omits an important explanatory variable, the lack of explanatory power will appear in the error term and temporal changes in the error term will be related to temporal changes in the omitted forcing variable.

To test this hypothesis, we create annual values for stratospheric water vapour and black carbon and test whether they are related to statistical estimates of the regression error from the cointegrating relation, the error correction models, or the simulation model. Annual estimates for stratospheric water vapour are created by interpolating (and then averaging) monthly observations (26). Annual estimates for black carbon are created by interpolating decadal values (25).

To test whether the omission of stratospheric water vapour or black carbon biases the estimate of the long-run cointegrating relation between surface temperature and radaitive forcing, we estimate equation (S-4):

$$u_t = \delta_0 + \delta_1 OV_t + v_t \qquad \text{(S-4)}$$

in which u_t is the error term from the cointegrating relation (S-1), OV_t is the omitted variable (stratospheric water vapour or black carbon), δ_0 and δ_1 are regression coefficients, and v_t is the regression error. If the omission of water vapour or black carbon affects the estimate for the long run relation between radiative forcing and temperature, the error from the long-run relationship will be related to the omitted variable (i.e. $\delta_1 \neq 0$). This hypothesis is evaluated by testing whether δ_1 in (S-4) is statistically significant (the unobserved error u_t is replaced by the estimate $\hat{u}_t$).

To check whether the omission of stratospheric water vapour or black carbon biases the estimate for the dynamics by which surface temperature adjusts to our measure of radiative forcings, we estimate equation (S-5 and S-6):

$$\varepsilon_t = \gamma_0 + \sum_{k=1}^{s} \gamma_k OV_{t-k} + \eta_t \qquad \text{(S-5)}$$

$$\varepsilon_t = \alpha + \sum_{k=1}^{s} \gamma_k \Delta OV_{k-i} + \eta_t \qquad \text{(S-6)}$$

in which ε_t is the error term from the error correction model (eq. S-2), $\gamma_j, j = 1,...,s$ are regression coefficients, and η_t is the regression error. The lag length (s) is chosen using AIC (17). Both a level (eq. S-5) and a first different (eq. S-6) specification are used because we cannot determine whether the time series for the omitted forcing is stationary (eq. S-5) or contains a stochastic trend (eq. S-6)—the time series for stratospheric water vapour is too short to generate a statistically meaningful conclusion and interpolating decadal values smoothes the time series in a way that distorts tests designed to detect stochastic trends. Again, if the omission of stratospheric water vapour or black carbon affects the statistical estimate of the error correction model, we expect that the null hypothesis $\gamma_1 = ... = \gamma_s = 0$ is rejected

To check whether the omission of stratospheric water vapour or black carbon affects the simulation for surface temperature that is generated by the statistical model, we estimate equation S-7:

$$Temp_t - \hat{T}emp_t = \rho_0 + \rho_1 OV_t + \upsilon_t \qquad \text{(S-7)}$$

in which Temp$_t$ is the observed value for global surface temperature, $\hat{T}emp_t$ is the value for surface temperature simulated by the model (red line, Figure S-1), ρ_0 and ρ_1 are regression coefficients estimated using OLS, and υ_t is the regression error. Again, if the omission of stratospheric water vapour or black carbon biases the forecast, we expect to reject the null hypothesis $\rho_1 = 0$ for equation S-7.

As indicated in Table S8, the results reject the null hypothesis that stratospheric water vapour is not related to the errors from the cointegrating relation (eq. S-4). But the OLS estimate of δ_1 has a negative (wrong) sign. Stratospheric water vapour has a positive effect on temperature such that the omission of this variable should cause the model to under-predict observed temperature, in which case δ_1 would be positive. We fail to reject the null hypothesis that stratospheric water vapour is not related to the short run dynamics by which surface temperature adjusts to our measure of radiative forcing (eq. (S-5) and (S-6)) or the simulation error (eq. (S-7)). Together these results suggest that the omission of stratospheric water vapour does not have a statistically meaningful effect on our results.

For black carbon we fail to reject the null hypothesis that black carbon is not related to the error from the cointegrating relation or the short run dynamics by which surface temperature adjusts to our measure of radiative forcing. We reject the null hypothesis that black carbon is not related the forecast error at the ten percent level. Although the estimate for δ_1 has the correct sign (positive), the estimated regression has little explanatory power—the R^2 is 0.017. Together, these results suggest that the omission of black carbon has little effect on the analysis.

SI-2.8 Relation to Satellite Measures of Top of Atmosphere Net Forcing. If our measure for total radiative forcing is a reasonable representation, it should be consistent with satellite measures for top of the atmosphere net energy flux (TOA). Measurements (W/m^2) are available starting 2000:Q2 from the Terra instrument and 2002:Q3 from the Aqua instrument. Because our estimates are annual, there are not enough observations to estimate statistically meaningful relation between our measure of radiative forcing and satellite measures of TOA.

Instead, we determine whether our measure of radiative forcing and satellite measures of TOA 'move' in the same direction by fitting a time trend to these variables with the following equation:

$$Y_t = \alpha + \theta Time_t + \eta_t \qquad \text{(S-8)}$$

in which Y is either our measure of radiative forcing or quarterly anomalies for a satellite measure of TOA, α and θ are regression coefficients estimated using OLS, and η is a regression error. The change in Y over the sample period is given by the sign and statistical significance of θ as evaluated by a t test of the null hypothesis $\theta = 0$.

The results in Table S-9 indicate that neither satellite measure of TOA shows a statistically measurable change between the start date and 2008:Q4. Similarly, our measure of radiative forcing does not show a statistically measurable change during comparable sample periods. Together, these results suggest that our measure of radiative forcing is consistent with satellite measures for top of the atmosphere net energy flux.

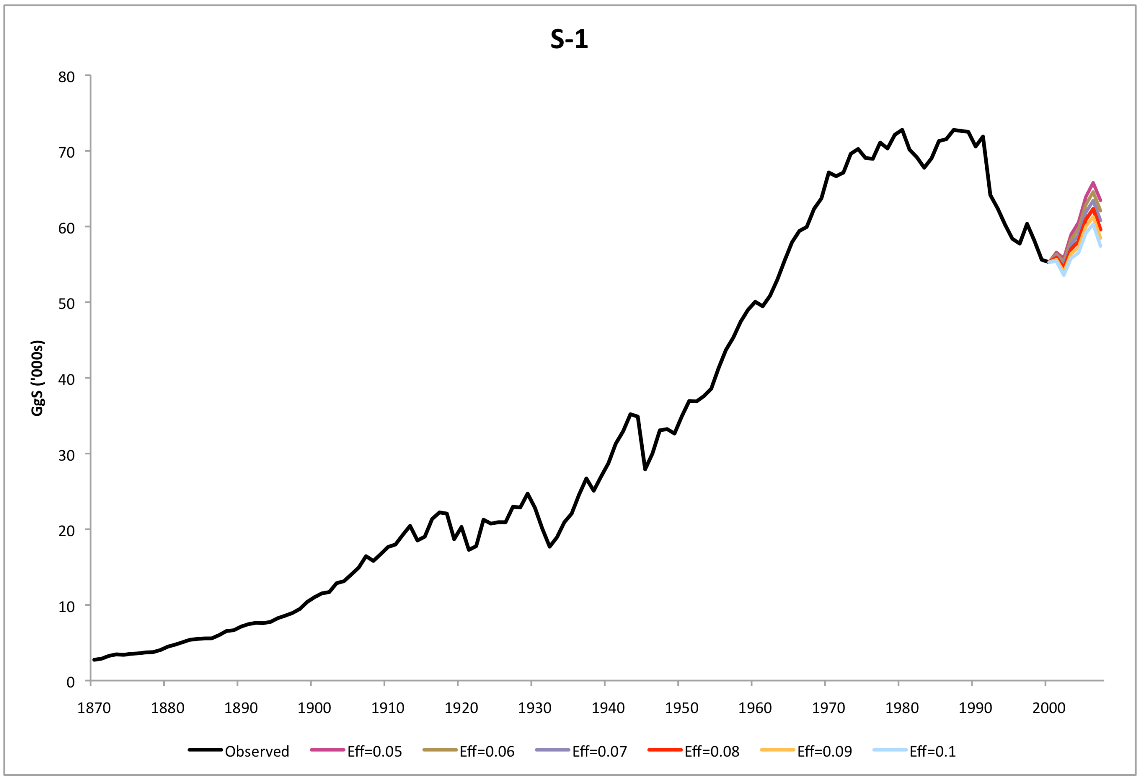

Figure S-1 Observed sulphur emissions GgS 1870-2007 (black line). Sulphur emission forecasts eff = 0.5 (purple line), eff = 0.6 (brown line), eff = 0.7 (grey line), eff = 0.8 (red line), eff = 0.9 (orange line), eff = 0.1 (light blue line).

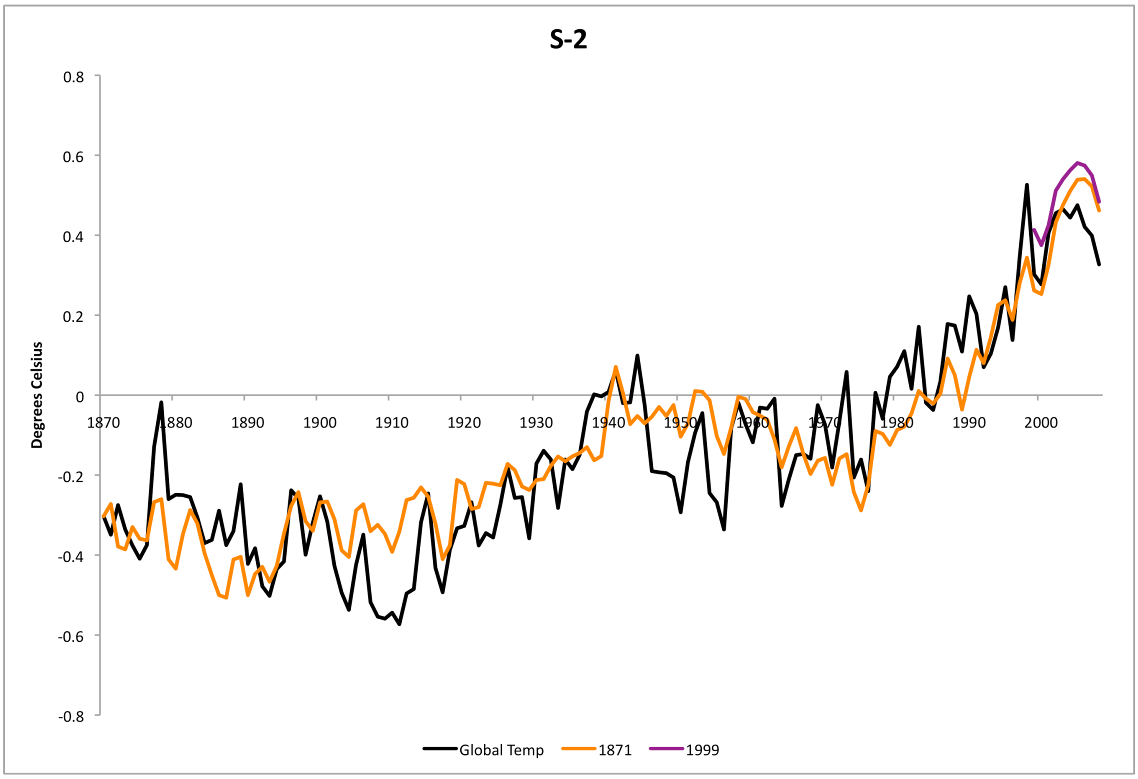

Figure S-2 Observed global surface temperature degrees Celsius (black line). Out-of-sample forecast with no additional temperature data after 1870 (orange line), and 1999 (purple line).

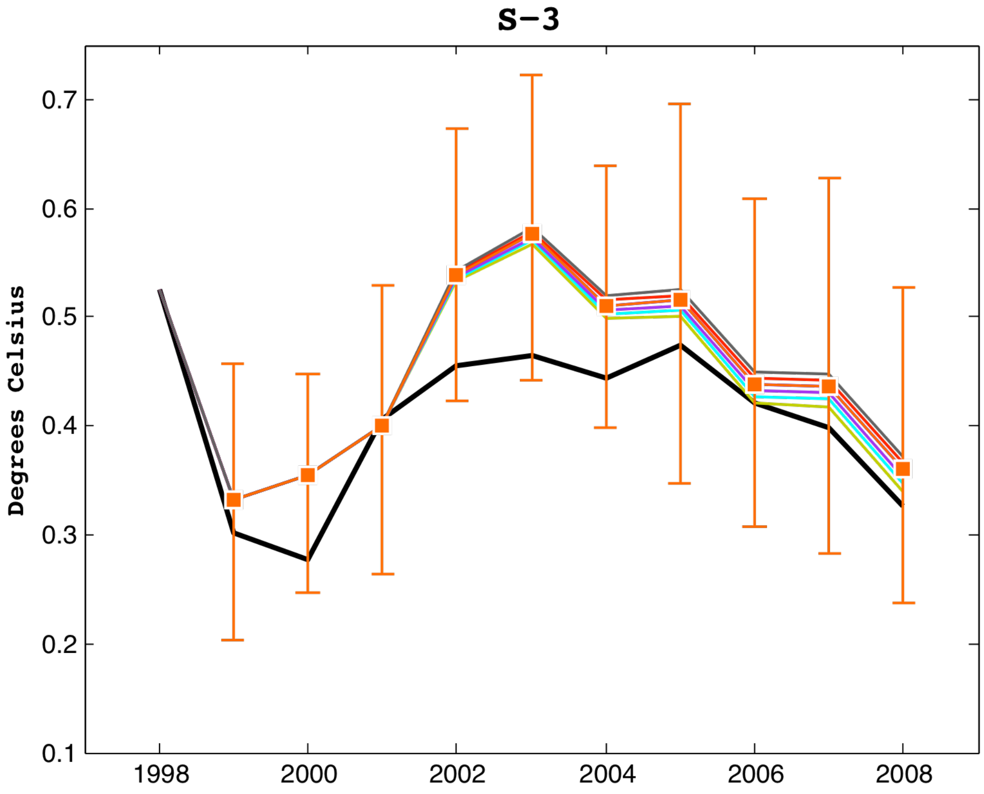

Figure S-3 Out of sample forecasts of observed surface temperature (black line) with *Eff*=0.05 (green line), *Eff*=0.06 (light blue line), *Eff*=0.07 (purple line), *Eff*=0.08 (orange line), 95% confidence intervals (orange bars), *Eff*=0.09 (red line), *Eff*=0.10 (grey line).

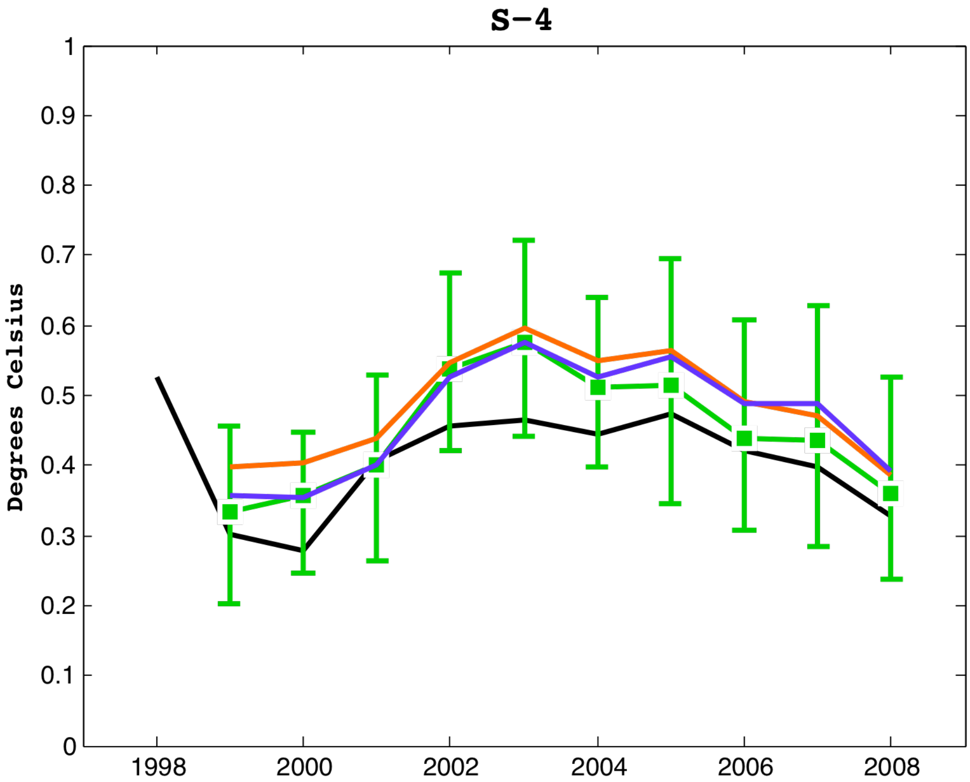

Figure S-4 Out of sample forecasts of observed surface temperature (black line) with sulphur emissions to radiative forcing conversion parameter, λ=1.0 (green line), with 95% confidence intervals (green bars), λ=0.6 (blue line), and λ=2.1 (orange line).

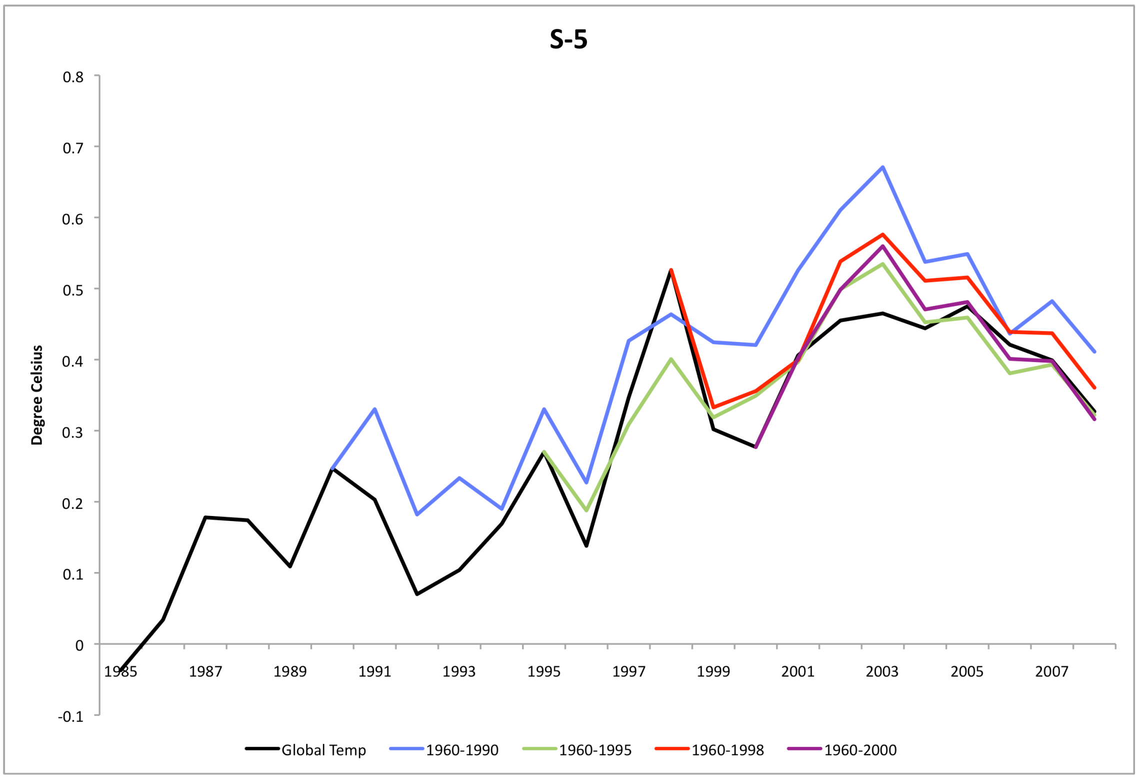

Figure S-5 Out-of-sample forecasts of observed surface temperature (black line) based on in-sample estimations from periods 1960-1990 (light blue line), 1960-1995 (green line), 1960-1998 (red line), and 1960-2000 (purple line).

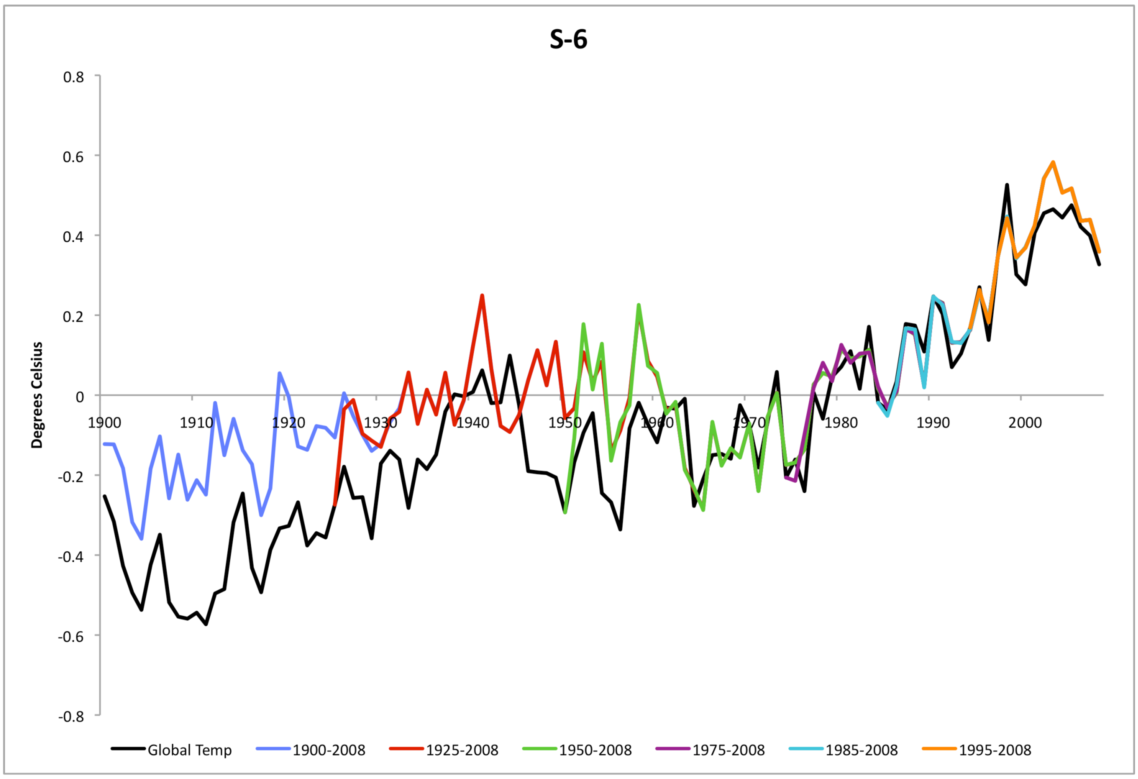

Figure S-6 Observed surface temperature (black line) forecast with spin-up in 1900 (blue line), 1925 (red line), 1950 (green line), 1975 (purple line), 1985 (light blue line), and 1995 (orange line).

Table S1 – Components of 2000 Emissions

Activity	Share of 2000 emissions
Coal	51.3%
Petroleum	23.1%
Smelting	10.0%
Marine bunkers	7.0%
Natural gas	2.1%
Other	6.4%

Table S2 – Estimating Efficiency Gains

Eff	B
0.05	0.944
0.055	0.954
0.060	0.964
0.065	0.974
0.070	0.983
0.075	0.993
0.080	1.00
0.085	1.012
0.090	1.021
0.096	1.030
0.10	1.039

Table S3 – DOLS and ECM results for three sample periods

	Estimation Sample Period		
Variable	1864-1998	1920-1998	1960-1998
Long-run relation			
α	-0.34^{**}	-0.32^{**}	-0.24^{**}
β	0.57^{**}	0.54^{**}	0.41^{**}
ADF	-5.02^{**}	-3.89^{**}	-4.39^{**}
R statistic	0.36	0.14	1.00^{*}
Error correction model			
α	0.022^{*}	0.034^{*}	$8.92\text{x}10^{-2**}$
β_2	$-.241^{**}$	-0.387^{**}	-0.816^{**}
δ_1	-0.154^{+}	$-5.94\text{x}10^{-2}$	-0.140
δ_2	$4.09\text{x}10^{-3}$	–	$-.246^{+}$
ϕ_1	$-1.48\text{x}10^{-1}$	-0.201	0.263^{**}
ϕ_2	$-4.26\text{x}10^{-2}$	–	$7.62\text{x}10^{-2}$
π_0	$-5.62\text{x}10^{-3\,**}$	$-7.20\text{x}10^{-3**}$	$-6.55\text{x}10^{-3**}$
π_1	$-2.35\text{x}10^{-3}$	$-1.00\text{x}10^{-3}$	$-8.60\text{x}10^{-3**}$
π_2	$3.18\text{x}10^{-3\,*}$	–	$-6.75\text{x}10^{-3**}$
ζ_0	$4.62\text{x}10^{-2\,**}$	$4.86\text{x}10^{-2*}$	$8.52\text{x}10^{-2**}$
ζ_1	$-8.76\text{x}10^{-3}$	$2.10\text{x}10^{-2}$	$3.91\text{x}10^{-2**}$
ζ_2	$1.38\text{x}10^{-4}$	–	$2.11\text{x}10^{-2}$

Coefficients are statistically significantly different from zero at the: **1%, *5%, +10% level as determined by standard errors that are calculated using the procedure described by Newey and West (18) with a lag length of six—conclusions about the significance level of regression coefficients do not change if a lag length of three is used.

Table S-4 Finding of cointegration* between surface temperature and radiative forcing out of 1,000 experimental data sets

	Sample Period		
Error	1860-1998	1920-1998	1960-1998
±10%	1000	950	966
±15%	1000	901	904
±20%	1000	824	707
±25%	994	709	430
±33%	932	469	123
±40%	770	299	41
±50%	506	136	8

* The critical value (-2.83) for the Augmented Dickey Fuller statistic (constant, no trend) is calculated using values from (20).

Table S5 – Effect on statistical results (ADF cointegration test, estimated cointegrating coefficient β_1, and error correction term β_2) of scaling anthropogenic sulfur by a factor of λ

	1864-1998			**1920-1998**			**1960-1998**		
λ	**ADF**	β	β_2	**ADF**	β	β_2	**ADF**	β	β_2
0.0	-2.64^{+}	0.26^{**}	-0.23^{**}	-2.49	0.27^{**}	-0.44^{**}	-5.01^{**}	0.41^{**}	-0.09
0.1	-2.71^{+}	0.28^{**}	-0.24^{**}	-2.62	0.28^{**}	-0.45^{**}	-5.12^{**}	0.41^{**}	-0.14
0.2	-2.79^{*}	0.30^{**}	-0.24^{**}	-2.77	0.30^{**}	-0.50^{**}	-5.18^{**}	0.42^{**}	-0.21
0.3	-2.87^{*}	0.32^{**}	-0.25^{**}	-2.94^{*}	0.33^{**}	-0.52^{**}	-5.21^{**}	0.42^{**}	-0.38
0.4	-2.96^{*}	0.34^{**}	-0.25^{**}	-3.13^{*}	0.35^{**}	-0.53^{**}	-5.18^{**}	0.42^{**}	-0.48
0.5	-3.06^{*}	0.37^{**}	-0.26^{**}	-3.35^{*}	0.38^{**}	-0.53^{**}	-5.12^{**}	0.43^{**}	-0.57
0.6	-3.16^{*}	0.40^{**}	-0.26^{**}	-3.59^{*}	0.41^{**}	-0.53^{**}	-5.02^{**}	0.43^{**}	-0.67^{*}
0.7	-3.25^{*}	0.43^{**}	-0.27^{**}	-3.81^{*}	0.45^{**}	-0.52^{**}	-4.88^{**}	0.42^{**}	-0.75^{*}
0.8	-5.05^{**}	0.47^{**}	-0.26^{**}	-3.98^{**}	0.48^{**}	-0.50^{**}	-4.73^{**}	0.42^{**}	-0.80^{**}
0.9	-5.07^{**}	0.52^{**}	-0.26^{**}	-4.02^{**}	0.52^{**}	-0.45^{**}	-4.56^{**}	0.42^{**}	-0.82^{**}
1.0	-5.02^{**}	0.57^{**}	-0.24^{**}	-3.89^{**}	0.54^{**}	-0.39^{**}	-4.39^{**}	0.41^{**}	-0.82^{**}
1.1	-3.91^{**}	0.62^{**}	-0.21^{**}	-3.61^{*}	0.55^{**}	-0.31^{**}	-4.22^{**}	0.40^{**}	-0.79^{**}
1.2	-3.67^{*}	0.66^{**}	-0.15^{*}	-3.24^{*}	0.54^{**}	-0.16^{**}	-4.05^{**}	0.39^{**}	-0.75^{**}
1.3	-3.34^{*}	0.70^{**}	-0.11^{+}	-2.87^{*}	0.49^{**}	-0.12^{*}	-3.88^{*}	0.38^{**}	-0.71^{**}
1.4	-2.97^{*}	0.70^{**}	-0.08	-2.51	0.42^{**}	-0.09^{+}	-3.73^{*}	0.37^{**}	-0.67^{**}
1.5	-2.16	0.64^{**}	-0.05	-2.17	0.33^{*}	-0.07	-3.58^{*}	0.36^{**}	-0.63^{**}
1.6	-1.73	0.51^{*}	-0.04	-1.84	0.24	-0.06	-3.45^{*}	0.35^{**}	-0.59^{**}
1.7	-1.23	0.33	-0.03	-1.54	0.15	-0.06	-3.32^{*}	0.34^{**}	-0.55^{**}
1.8	-0.72	0.15	-0.03	-1.26	0.08	-0.07	-3.20^{*}	0.33^{**}	-0.52^{**}
1.9	-0.28	0.00	-0.03	-1.03	0.03	-0.07	-3.09^{*}	0.32^{**}	-0.49^{*}
2.0	0.01	-0.10	-0.04	-0.84	-0.01	-0.08	-2.98^{*}	0.30^{**}	-0.46^{*}
2.1	0.16	-0.17	-0.04	-0.71	-0.04	-0.08	-2.89^{*}	0.29^{**}	-0.43^{*}
2.2	0.20	-0.21	-0.05	-0.61	-0.06	-0.09	-2.79^{+}	0.28^{**}	-0.41^{*}
2.3	0.19	-0.23	-0.05	-0.54	-0.07	-0.10	-2.71^{+}	0.27^{**}	-0.39^{*}
2.4	0.14	-0.23	-0.06	-0.49	-0.08	-0.10	-2.62	0.26^{**}	-0.38^{*}

Test statistics reject the null hypothesis at the: **1%, and *5% level

Values in bold highlight the range of λ for which the statistical results are not sensitive to uncertainty about the formulae for radiative forcing.

Table S6 – DOLS and ECM results for three sample periods estimated with the GISS temperature time series

	Estimation Sample Period		
Variable	1885-1998	1920-1998	1960-1998
Long-run relation (eq S-1)			
α	-0.21**	-0.18**	-0.11**
β	0.58**	0.52**	0.41**
ADF	-5.18**	-4.12**	-4.41**
R statistic (eq. S-3)	3.12*	0.61	0.18
Error correction model (eq. S-2)			
α	0.022+	0.034**	0.083**
β_2	-.276**	-0.417**	-0.98**
δ_1	-0.177	-0.110	-4.58×10^{-2}
δ_2	-0.072	--	0.264*
δ_3	-0.171	--	
ϕ_1	-0.145	-0.202	0.221
ϕ_2	-2.95×10^{-2}	--	0.100
ϕ_3	-6.12×10^{-2}	--	
π_0	-5.80×10^{-3} **	-6.43×10^{-3}**	-5.01×10^{-3}**
π_1	-2.06×10^{-3}	-1.30×10^{-3}	-7.91×10^{-3}**
π_2	2.94×10^{-3}	--	6.29×10^{-3}**
π_3	-2.74×10^{-3}+	--	
ζ_0	4.75×10^{-2} **	4.64×10^{-2}**	7.43×10^{-2}**
ζ_1	-2.00×10^{-3}	2.55×10^{-2}	3.80×10^{-2}+
ζ_2	2.98×10^{-2}	--	2.15×10^{-2}
ζ_3	2.21×10^{-2}+	2.18×10^{-2}	

Table S-7 - Cointergation Tests of the GISS and CRU temperature data

	ADF Test of Cointegration	R Statistic breakdown 1999-2008
1880-2008	-2.83*	8.89E-04
1920-2008	-3.32*	2.86E-03
1960-2008	-4.11**	6.76E-03

Test statistics reject the null hypothesis at the: **1%, and *5% level

Table S-8 - Regression results for omitted variable bias

	Stratospheric Water Vapour Sample period 1981-2008	Black Carbon Sample period 1866-2000
β = 0 (eq. S-4)	$\beta = -0.111$, $p < 0.01$	$\beta = -2.96E-3$, $p > 0.82$
β = 0 (eq. S-5)	$\beta = 2.44E-3$, $p > 0.95$	$\beta = 3.04E-3$, $p > 0.047$
β = 0 (eq. S-6)	$\beta = 6.06E-2$, $p > 0.10$	$\beta = 0.179$, $p > 0.06$
β = 0 (eq. S-7)	$\beta = -2.29E-2$, $p > 0.62$	$\beta = 1.57E-2$, $p > 0.37$

Table S-9 - Statistical estimates for changes in Satellite measures of TOA or our measure of radiative forcing through 2008.

Dependent Variable (Y)	Sample period	β	t statistic
Aqua	2002:Q3-2008:Q4	4.09E-03	.33
Terra	2000:Q2-2008:Q4	6.85E-03	.93
RFAGG	2000-2008	-4.57E-03	.91
RFAGG	2002-2008	-7.75E-3	.83

References

1. Smith S-J, Conception E, Andres R, Lurz J (2004) *Historical Sulfur Dioxide Emissions 1850-2000: Methods and Results*, U.S. Department of Energy, Springfield, VA.

2. Stern D (2005) Global sulfur emissions from 1850 to 2000. *Chemosphere* 58:163–175.

3. Kaufmann R-K, Kauppi H, Stock J-H (2006) Emissions, concentrations & temperature: A time series analysis. *Clim Change* 77:249-278.

4. EIA (2008) Coal Consumption - Selected Countries, Most Recent Annual Estimates, 1980-2007
http://www.eia.doe.gov/emeu/international/RecentCoalConsumptionMST.xls

5. Xu Y (in press) Revising official data of SO2 emissions from China's coal power sector and revealing the impact of SO2 scrubbers. *Atmos Environ*.

6. Xu, Y, Williams R-H, Socolow R-H (2009) China's rapid deployment of SO2 scrubbers. *Energy & Environmental Science*. 2:459-465.

7. EIA (2009) Emissions from Energy Consumption at Conventional Power Plants and Combined-Heat-and-Power Plants
http://www.eia.doe.gov/cneaf/electricity/epa/epaxlfile5_1.xls.

8. EIA (2009) Net Generation by Energy Source: Total (All Sectors)
http://www.eia.doe.gov/cneaf/electricity/epm/table1_1.html

9. USGS (2009) in *U.S Geological Survey Minerals Information - Minerals Yearbook: Metals and Minerals http://minerals.usgs.gov/minerals/pubs/commodity/myb/*

10. Lefohn A-S, Husar J-D, Husar R-B (1999) Estimating historical anthropogenic global sulfur emission patterns for the period 1850-1990. *Atmos Environ* 33:3435-3444.

11. IMO (2009) *Prevention of air pollution from ships: Second IMO GHG Stud*y, International Maritime Organization.

12. EIA (2009) World Petroleum Consumption, Annual Estimates, 1980-2008
http://www.eia.doe.gov/emeu/international/RecentPetroleumConsumptionBarrelsperDay.xls

13. EIA (2008) World Dry Natural Gas Consumption, 1980-2006
http://www.eia.doe.gov/pub/international/iealf/table13.xls

14. Kattenburg A-F. *et a*l. in *The Science of Climate Change* (eds Houghton, J. T. et al.) (Cambridge University Press, Cambridge, UK, 1996), pp 285-357.

15. Smith S-J, Van Aardenne J, Klimont Z, Andres R, Volke A-C, Delgado Arias S (2010) Anthropogenic Sulfur Dioxide Emissions: 1850-2005. *Atmos Chem Phys* 10:16111-16151.

16. Stock J-H, Watson M-W (1993) A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica* 61:783-820.

17. Akaike H *Information theory and an extension of the maximum likelihood principl*e (In 2nd International Symposium on Information Theory), Akademiai Kiado.

18. Newey W, West K (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55:703-708.

19. Andrews D-W-K, Kim J-Y (2006) Tests for cointegration breakdown over a short time period. *J Bus Econ Stat* 24:379-394.

20. MacKinnon J-G (1994) Approximate asymptotic distribution functions for unit root and cointegration tests. *J Bus Econ Stat* 12:167-176.

21. Forster P, Ramaswamy V, Artaxo P, Berntsen T, Betts R, Fahey D-W, Haywood J, Lean J, Lowe D-C, Myhre G, Nganga J, Prinn R, Raga G, Schulz M, Van Dorland R, 2007: Changes in Atmospheric Constituents and in Radiative Forcing. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S, Qin D, Manning M, Chen Z, Marquis M, Averyt K-B, Tignor M, Miller H-L (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA

22 GISS (2010) Temperature data ftp://ftp.ncdc.noaa.gov/pub/data/anomalies/annual.land_ocean.90S.90N.df_1901-2000mean.dat

23. Murphy D.M. S. Solomon, R.W. Portman, K.H. Rosenlof, P.M. Forster, and T. Wong (2009) An observationally based energy balance for the Earth since 1950. *J Geophys Res* 114:D17107.

24 Ramanathan V, Carmichael G (2008) Global and regional climate changes due to black carbon, *Nat Geosci* 1:221-227.

25. Solomon S, Rosenlof K-H, Portman R-W, Daniel J-S, Davis S-M, Sanford T-J, Plattner G-K (2010) Contributions of stratospheric water vapour to decadal changes in the rate of global warming. *Science,* B327:1219-1223.

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