Testing the robustness of the anthropogenic climate change detection statements using different empirical models

J. Imbers,[1] A. Lopez,[2] C. Huntingford,[3] and M. R. Allen[1]

[1] This paper aims to test the robustness of the detection and attribution of anthropogenic climate change using four different empirical models that were previously developed to explain the observed global mean temperature changes over the last few decades. These studies postulated that the main drivers of these changes included not only the usual natural forcings, such as solar and volcanic, and anthropogenic forcings, such as greenhouse gases and sulfates, but also other known Earth system oscillations such as El Niño Southern Oscillation (ENSO) or the Atlantic Multidecadal Oscillation (AMO). In this paper, we consider these signals, or forced responses, and test whether or not the anthropogenic signal can be robustly detected under different assumptions for the internal variability of the climate system. We assume that the internal variability of the global mean surface temperature can be described by simple stochastic models that explore a wide range of plausible temporal autocorrelations, ranging from short memory processes exemplified by an AR(1) model to long memory processes, represented by a fractional differenced model. In all instances, we conclude that human-induced changes to atmospheric gas composition is affecting global mean surface temperature changes.

Citation: Imbers, J., A. Lopez, C. Huntingford, and M. R. Allen (2013), Testing the robustness of the anthropogenic climate change detection statements using different empirical models, *J. Geophys. Res. Atmos.*, *118*, 3192–3199, doi:10.1002/jgrd.50296.

1. Introduction

[2] Central to the debate surrounding global warming is whether or not a trend of increasing temperatures can be detected and in particular if it can be attributed to anthropogenic causes. The methods developed to detect and attribute climate change are based on a multiregression approach that assumes that the observed climate changes can be separated into components, one of them representing the internal variability of the climate system and the others corresponding to signals that result from changes external to the system, such as changes in the Earth's radiative budget due for instance to greenhouse gas increases. The goal of detection and attribution is therefore to show that the signal of change can be attributed to anthropogenic forcings using a robust statistical methodology [*Allen et al.* 2006; *Hegerl and Zwiers*, 2011].

[3] The internal variability of the climate system is in this context defined as the variability that would occur in the absence of natural (solar and volcanic) or anthropogenic forcings, solely due to the coupling of atmosphere, ocean, biosphere, and cryosphere dynamics. The attribution of the signal to anthropogenic causes relies heavily on an accurate characterization of this variability to prevent the potential of mis-attributing a natural signal to the anthropogenic influence. In principle, information about unforced variability could be extracted from historical records. However, due to the fact that the instrumental record is relatively short and sparse and affected by the anthropogenic forcings, many uncertainties remain regarding the appropriate statistical characterization of internal variability. In most detection and attribution studies, control runs of Global Climate Models (GCMs) are used to characterize it [*Allen et al.* 2006; *Hegerl and Zwiers*, 2011]. More recently, Imbers et al. (J. Imbers et al., Sensitivity of climate change detection and attribution to the characterization of internal variability, submitted to Journal of Climate, 2013) explored the robustness of the detection and attribution statistics for the global mean surface air temperature record, when the internal variability is characterized by simple stochastic models that nevertheless span a representative range of plausible temporal autocorrelation structures, finding that the greenhouse signal remains detectable under the detection model employed.

[4] The signals considered in the majority of the surface air temperature detection and attribution studies comprise the forced temperature responses to (1) raised levels of atmospheric greenhouse gas (GHG) concentrations, (2) altered atmospheric aerosol concentrations, (3) influence of volcanoes, and (4) solar forcing variation.

[5] During the last few years, different authors [*Lean and Rind*, 2009; *Folland et al.*, 2011; *Kaufmann et al.*, 2011; *Lockwood*, 2008] have considered the possibility of

[1]Department of Atmospheric Oceanic Planetary Physics, University of Oxford, Oxford, UK.
[2]Center for the Analysis of time series and CCCEP, London School of Economics, London, UK.
[3]Centre for Ecology and Hydrology, Wallingford, UK.

Corresponding author: A. Lopez, Center for the Analysis of time series and CCCEP, London School of Economics, London, UK. (ana.lopez@univ.ox.ac.uk)

including other signals that could explain the evolution of the global mean temperature over the recent past. In their statistical models, these authors included, apart from the solar, volcanic, and anthropogenic components, signals representing the effect of El Niño (through the Southern Oscillation Index (SOI) for *Kaufmann et al.* [2011], a multivariate ENSO index for *Lean and Rind* [2009], the first high-frequency eigenvector of global SST for *Folland et al.* [2011], and the ENSO3.4 index for *Lockwood* [2008]) and the Atlantic Multidecadal Oscillation (AMO; for *Folland et al.* [2011]). The goal of these studies was not to perform a detection and attribution study of the different signals. Instead, *Lean and Rind* [2009] and *Folland et al.* [2011] aimed to build a statistical model that could be used to predict surface temperature changes in the near future. *Kaufmann et al.*'s [2011] goal was to reconcile the plateau in global surface air temperature observed between 1998 and 2008 with anthropogenic climate change and the steady increase in greenhouse gases concentrations. And *Lockwood* [2008] intended to analyze the contribution of changes in solar output to global mean air surface temperature.

[6] In this paper, the goal is to test the robustness of the anthropogenic climate change detection statements under different characterizations of internal variability, when the global mean temperature responses to internal oscillations of the climate system such as ENSO or AMO are considered as separate signals and not implicitly included in the noise term in the multiregression analysis. To this end, we assume that the different signals in the global mean surface air temperature record identified in the aforementioned studies do quantify the global mean temperature response to external forcings (anthropogenic, solar, and volcanic) and modes of variability (ENSO and AMO) and test whether or not the detection statistics changes when the residual internal variability is characterized by two simple stochastic models than span a range of plausible temporal autocorrelation structures (Imbers et al., submitted manuscript, 2013).

[7] In section 2, we first describe the input data employed in our analysis. Since these data include the signals identified by *Lean and Rind* [2008, 2009], *Kopp and Lean* [2011], *Folland et al.* [2011], *Kaufmann et al.* [2006, 2011], and *Lockwood* [2008], we summarize in this section the key features of these studies. We then briefly discuss the detection and attribution linear regression approach and how it is modified when the noise model corresponds to each of the two stochastic models used to characterize the residual internal variability. In section 3, we test the robustness of the detection statements under changes in the depiction of the residual internal variability. We also analyze how the robustness of the detection of different signals changes depending on whether the ENSO oscillation is explicitly considered as a separate signal or implicitly taken into account in the noise term of the linear regression. Finally, we consider the robustness of the detection statements when oscillations that have been argued to be important to explain recent warming [*Loehle and Scafetta*, 2011] are included in our analysis. Section 4 is devoted to the conclusions.

2. Data and Methods

[8] The multiregression approach underlying the detection and attribution of climate change requires information about the observed climate changes, the signals of interest, and a representation of the residual internal variability of the climate system. As mentioned in section 1, our aim is to test the robustness of the anthropogenic climate change detection statements under different characterizations of internal variability, when the global mean temperature responses to internal oscillations of the climate system such as ENSO or AMO are considered as separate signals and not implicitly included in the noise term in the multiregression analysis.

[9] In section 2.1, we summarize the findings of the four studies that identified the different signals in the global mean surface air temperature record including ENSO and AMO. We then discuss, in section 2.2, the methodology, including a brief description of the detection and attribution approach and the key features of the noise models used to characterize the residual internal variability .

2.1. Summary of Previous Studies and Data

[10] The four studies that our analysis builds on were not, in general, detection and attribution studies. Their aims were slightly different, as well as the signals included into the global mean temperature decomposition and length and sampling intervals of their time series. In what follows, we briefly summarize the main features of theses studies.

[11] *Lean and Rind*'s [2009] goal is to forecast global and regional climate change in the near future by decomposing the observed record of monthly mean surface air temperature in terms of its combined linear response to ENSO, solar and volcanic activity, and anthropogenic influences (see also *Lean and Rind* [2008]; *Kopp and Lean* [2011]). They use 1980–2008 monthly time series of mean surface temperature anomalies with respect to 1951–1980 and perform a multivariate linear regression against the instrumental surface temperature record (HadCRUT3v [*Brohan et al.*, 2006]) to find the optimal combination of those four signals that better explain that record. Their solar, volcanic, anthropogenic, and ENSO signals are lagged by 1, 7, 120, and 4 months, respectively, with respect to the temperature observations in order to maximize the proportion of global variability that the statistical model captures (76% of the variance observed since 1980).

[12] *Folland et al.*'s [2011] aim is to forecast annual global mean temperature anomalies using a statistical model that estimates the contributions of six physical factors to global mean temperature change and variability. These factors are anthropogenic greenhouse gases and aerosols, volcanic aerosols and changes in solar output, and two internal modes of variability: ENSO (represented by the first high-frequency eigenvector of global sea surface temperatures) and the AMO. The influence of these factors on observed surface temperatures is estimated from cross-validated multiple linear regression using annual surface temperature values from 1891 to 2010 from HadCRUT3, NOAA's National Climatic Data Center, and Goddard Institute for Space Studies. Due to the cross-validation method, an ensemble of 120 reconstructions of the observed variable is obtained. In our analysis, we use the ensemble mean time series resulting from the regression with the HadCRUT3 dataset.

[13] *Kaufmann et al.* [2011] use a statistical model derived to estimate the relation among emissions of carbon dioxide and methane, the concentrations of these gases,

and global surface temperature [*Kaufmann et al.*, 2006] to evaluate whether anthropogenic emissions of radiative active gases along with variability can account for the 1998–2008 hiatus in warming. The model is estimated with annual data from 1960 to 1998 and used to project 1998–2008 temperatures. The signals included in this model are the following: greenhouse gases, anthropogenic sulfur emissions, solar insolation, ENSO (represented by the Southern Oscillation Index(SOI)), and radiative forcing of volcanic sulfates.

[14] *Lockwood* [2008] intends to analyze the contribution of changes in solar output to global mean air surface temperature. The statistical model consists of a multivariate fit to the global monthly mean surface air temperature anomaly for the period 1953–2007. The signals included in the fit are the solar, volcanic, and anthropogenic components (the latest as a linear trend) and the ENSO3.4 index to represent the effect of El Niño.

[15] The results of these four studies are summarized in Figure 1. In all cases, the temporal resolution of the data is the same as in the original studies, and the authors have kindly provided updated time series to December 2010. The time series representing the different contributions to the global mean temperature fit in the top panel of Figure 1 constitute the individual signals that we use in section 3 for the detection sensitivity analysis.

[16] Note that in the literature there are many studies on the global mean temperature using a wide range of statistical methods, such as empirical decomposition (see, for instance, *Lee and Ouarda* [2011], *Lee and Ouarda* [2012], and *Wu et al.* [2007]). These studies would have been equally relevant in this paper; however, in our work we only describe the four studies mentioned above because the authors provided their data for our robustness tests.

2.2. Detection and Attribution and Representations of Internal Variability

[17] The standard detection and attribution methodology assumes that, given the observed temperature change T_{obs} and a set of n (model derived) responses determined for each forcing T_i , the observed change can be written as

$$T_{\text{obs}} = \mathbf{T}\boldsymbol{\beta} + \mathbf{u}, \tag{1}$$

Here T is a matrix with $n + 1$ columns corresponding to the signals T_i and a constant term that is included to remove the mean. The stochastic term $\mathbf{u}$ is called the "climate noise" and represents the residual internal variability of the climate system. The β_i are the scaling factors corresponding to each of the signals. They account for possible errors in the amplitudes of signals by scaling the signal patterns to best match observations while assuming that the temporal pattern of the response is correct [*Hegerl and Zwiers*, 2011].

[18] Equation (1) shows that the detection and attribution approach assumes that signals corresponding to different forcings can be superimposed linearly, discarding the possibility of nonlinear interactions between the response to different forcings and between forced and unforced variability as represented by $\mathbf{u}$ [*Allen et al.* 2006].

[19] Under the assumption that $\mathbf{u}$ is multivariate normal [*Allen and Tett*, 1999], it can be shown that the best (in a least square sense) linear unbiased estimator (BLUE) of the

vector of scaling factors, $\boldsymbol{\beta} = (\beta_1, \beta_2, ..\beta_{n+1})$, is given by the following:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{T}^\dagger \boldsymbol{\Omega}^{-1} \mathbf{T}\right)^{-1} \mathbf{T}^\dagger \boldsymbol{\Omega}^{-1} T_{\text{obs}}, \tag{2}$$

with variance as follows:

$$V(\hat{\boldsymbol{\beta}}) = \left(\mathbf{T}^\dagger \boldsymbol{\Omega}^{-1} \mathbf{T}\right)^{-1}, \tag{3}$$

where † is used to denote the transpose of a matrix, and $\boldsymbol{\Omega}$ is the noise lag-covariance matrix, i.e., $\boldsymbol{\Omega} = E(\mathbf{u}\mathbf{u}^\dagger)$ [*Kmenta*, 1971].

[20] According to the standard detection and attribution interpretation of the scaling factors, if a chosen uncertainty range for β_i does not include 0, then one can conclude that the signal T_i is likely present in the observations and therefore detectable according to this detection model. If the signal T_i is detectable but the uncertainty range does not include 1, that indicates that the response has to be re-scaled to match the observations. The simulated T_i scaled by its corresponding β_i provides an estimate of the range of responses that are consistent with the observations [*Allen et al.* 2006; *Hegerl and Zwiers*, 2011].

[21] In this paper, the regression in equation (1) is performed using the globally averaged temperature time series. This is a limitation of this analysis, because spatial information can be important for detecting climate response signals [*Hegerl and Zwiers*, 2011]. However, since our goal is to perform a robustness test of the detection statistics under different characterizations of the residual internal variability, taking as signals the global temporal response patterns derived in the four studies summarized in the previous section, we are constrained to use temporal patterns only. This constraint also implies that in our approach we ignore possible errors in the estimates of the temperature response or signals and uncertainties in the source of the forcings and how those sources are converted into radiative forcings for some of these signals. In summary, our attribution study only accounts for sampling uncertainty from observations (during the regression step).

[22] In order to explore different characterizations of the internal variability of the climate system, we will assume that $\mathbf{u}$ can be represented by a stochastic process that, in addition to the random component, includes some information about more complex temporal correlations between different states of the climate system. We refer to this temporal correlation between different states as the *memory* of the system; other authors refer to it as the climate persistence [*Beran*, 1994; *Percival et al.*, 2001]. As in Imbers et al. (submitted manuscript, 2013), we use two different stochastic models to represent internal variability. The first one is an autoregressive model of the first order (AR(1)), and the second one is a Fractional ARIMA model (fractionally differenced process (FD)) [*Percival et al.*, 2001]. The choice of the two statistical models is based on a simplicity criteria: the two models are characterized by a minimal number of parameters, and at the same time, they are representative of systems with very different internal temporal structures. While the AR(1) model is classified as a short memory process due to the exponential decay of its autocorrelation function, the FD model corresponds to a long memory process whose autocorrelation function decays algebraically. These different but
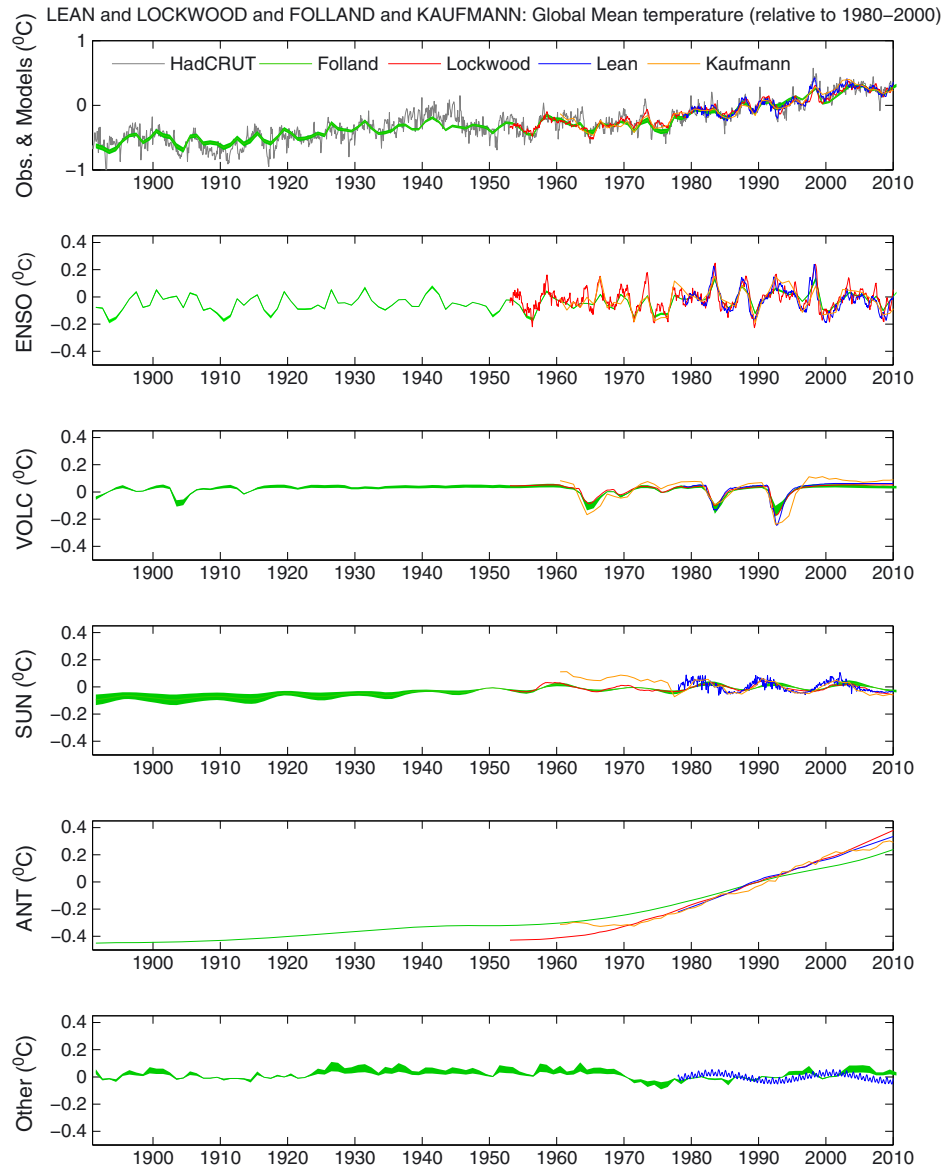
Figure 1. The top panel shows the observed global mean air surface temperature anomaly from HadCRUT3 (gray line) and the best multivariate fits using the methods of *Lean and Rind* [2009] (blue line), *Lockwood* [2008] (red line), *Folland et al.* [2011] (green line), and *Kaufmann et al.* [2011] (orange line). Note that in the case of Folland (green line), data are plotted with 94.4 % confidence interval based on the ensemble standard deviation [*Folland et al.*, 2011]. The remaining panels show the individual temperature contributions to the top panel fits from ENSO (second panel), volcanoes (third panel), solar irradiance (fourth panel), anthropogenic contribution (fifth panel), and other factors (sixth panel) that include the AMO for *Folland et al.* [2011] and minor annual, semi-annual, and 17.5 year cycle identified by *Kopp and Lean* [2011] in the residuals of *Lean and Rind's* [2009] model. In the case of Folland's data, the figure shows the mean ±1 standard deviation of the 120 time series ensemble.

plausible (see, for instance, *Hasselmann* [1979] and *Vyushin and Kushner* [2009]) characterizations of the internal variability in terms of simple parametric models allow for a sensitivity analysis of the detection statistics of the signals T_i in the observed temperature record, to well-defined parameters whose change is easily understood in terms of memory or unresolved variability (white noise) in the climate system. We therefore formulate the detection and attribution problem as in equation (1) but where **u** corresponds to

either AR(1) or an FD model whose main properties are described below.

[23] *Autoregressive model: AR(1)*

[24] The AR(1) model corresponds to a stationary stochastic process that can be written as follows:

$$u_t = a_1 u_{t-1} + a_0 \epsilon_t \qquad (4)$$

where u has zero mean, $E(u_t) = 0$, a_1 and a_0 are two defining parameters, and ϵ_i is white noise, i.e., $E(\epsilon_t \epsilon_{t'}) = \delta_{t,t'}$.

The lag-covariance matrix $\boldsymbol{\Omega} = <\mathbf{u}.\mathbf{u}^T>$ is given by the following:

$$\Omega^{AR_1}_{i,j} = \frac{a_0^2}{1-a_1^2} a_1^{|i-j|} \tag{5}$$

In the detection and attribution analysis, this covariance matrix is used when employing generalized linear regression to solve equation (1). The optimal scaling factors β_i and the parameters a_1 and a_0 of the noise u are simultaneously determined following the Hildreth-Lu method [*Kmenta*, 1971].

[25] *Fractional differenced model : FD*

[26] There is accumulated empirical evidence that the spectrum of global mean temperature is more complex than the spectrum of an AR(1) process [e.g., *Huybers and Curry*, 2006]. In order to account for this complexity in the depiction of the internal climate variability, we use an FD model [*Beran*, 1994] that has been considered in other climate studies as a plausible representation of internal variability [*Percival et al.*, 2001; *Vyushin and Kushner*, 2009, 2012].

[27] The FD model corresponds to a stationary stochastic process of zero mean defined by the following:

$$u_t = (1-B)^{-\delta} \epsilon_t, \tag{6}$$

where B is the backshift operator, i.e., $Bu_t = u_{t-1}$ [*Beran*, 1994]. Despite the extra complexity, this model is fully specified by just two parameters, δ and σ_e, and the white noise ϵ_t satisfying $E(\epsilon_t \epsilon_{t'}) = \sigma_e^2 \delta_{t,t'}$. The covariance matrix is in this case

$$\Omega^{FD}_{i,j} = \frac{\sigma_e^2 \sin(\pi\delta)\Gamma(1-2\delta)\Gamma(|i-j|+\delta)}{\pi\Gamma(|i-j|+1-\delta)}. \tag{7}$$

Additionally, it can be shown that for large τ, the autocorrelation functions decay algebraically as $\lim_{\tau\to\infty}\omega_{FD}(\tau) = |\tau|^{2\delta-1}$ [*Beran*, 1994]. Similarly to the AR(1) case, we use this covariance matrix, equation (7), to simultaneously determine the scaling factors β_i and the parameters δ and σ of the noise u in equation (1), applying the Hildreth-Lu method [*Kmenta*, 1971].

3. Results: Robustness of Anthropogenic Signal Detection

3.1. Robustness Under Changes in Characterization of Internal Variability

[28] In this section, the signals T_i in equation (1) are the responses to external forcings and modes of internal variability contributing to the global mean temperature time series, as calculated in the four studies discussed in section 2.1. The robustness of the detection statistics statement is assessed when modeling the residual internal variability as either an AR(1) or an FD process. Equation (1) is solved using generalized linear regression, assuming that the noise covariance matrix Ω in equations (2) and (3) is given by equation (5) or (7), respectively, and finding simultaneously the scaling factors for each signal and the stochastic parameters of the noise u.

[29] Hence, we use generalized multiregression and then test that the residuals of the fitted model in both cases are uncorrelated. We find that for all of the studies and given the observational record used in our analysis, the AR(1) and
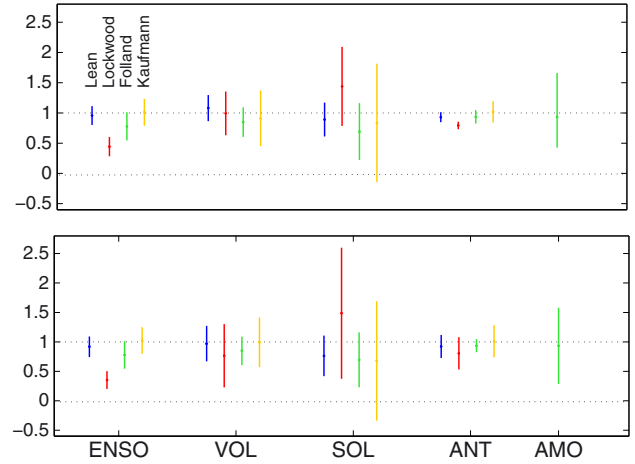
Figure 2. Scaling factors and their 95% confidence interval (vertical axis) for the different signals: ENSO, volcanic (VOL), solar (SOL), anthropogenic (ANT), and AMO (as indicated in the horizontal axis); using the data from the four studies: Lean (blue), Lockwood (red), Folland (green), and Kaufmann (orange) and assuming two stochastic models for the residual internal variability u, AR(1) (top panel) and FD (bottom panel) models.

the FD noise models are equally skillful at representing the residual internal variability. This ensures that, from a statistical point of view, both models are equally acceptable depictions of the "climate noise" in equation (1) and can be used to calculate the scaling factors and their confidence intervals.

[30] The results for the scaling factors β_i and their 95% confidence intervals are shown in Figure 2 for the AR(1) and the FD models in the top and bottom panels, respectively. We consider ENSO, solar, volcanic, and anthropogenic signals from *Lean and Rind* [2009], *Folland et al.* [2011], *Kaufmann et al.* [2011], and *Lockwood* [2008] as well as AMO for *Folland et al.* [2011].

[31] We find that although the estimated response temperature to the considered climate signals is different for each study, the scaling factors for any given signal, except for ENSO, are quantitatively similar in the sense that their confidence intervals overlap. In addition, we find that except for the solar signal using Kaufmann data, none of the scaling factors is consistent with zero, implying that all the other signals are detectable and features of the temperature record can be attributed to them. Figure 2 also shows that, except for Lockwood's ENSO signal, the scaling factors are consistent with unity, suggesting that the contributions of the forced responses to the observed global mean temperature are approximately correct in each of the four statistical approaches. This is to be expected since all of these estimated forced responses have been obtained by various forms of fit to observed temperatures. Our objective here is to put them all in a common statistical framework in order to assess the robustness of results.

[32] As expected, in most cases the confidence intervals are smaller when the internal variability is represented by a short memory process (AR(1) model) than by a long memory process (FD model) [*Beran*, 1994; *Percival et al.*, 2001].

Table 1. Estimated Parameters for the AR(1) Model $\left(a_1, a_0^2\right)$, and the FD Model (δ, σ^2)[a]

Model	AR(1) $\left(a_1, a_0^2\right)$	FD (δ, σ^2)
Lean	$(0.40, 5.0\times10^{-3})$	$(0.33, 5.1\times10^{-3})$
Lockwood	$(0.63, 6.3\times10^{-3})$	$(0.49, 6.2\times10^{-3})$
Folland	$(0.55, 4.6\times10^{-3})$	$(0.47, 4.6\times10^{-3})$
Kaufmann	$(0.01, 4.3\times10^{-3})$	$(0.00, 4.3\times10^{-3})$

[a]Used to characterize the residual internal variability u for the four studies *Lean and Rind* [2009], *Lockwood* [2008], *Folland et al.* [2011], and *Kaufmann et al.* [2011].

In the case of Kaufman's data, we find that the scaling factors are nearly identical in both cases, which is consistent with the fact that for this study the noise term u in equation (1) has nearly no memory, i.e., a_1 and δ are very close to zero (see Table 1).

[33] Table 1 summarizes the values of the stochastic models' parameters for the four studies. It is clear that, except for Kaufmann's signals, the best fit of the residual internal variability u in equation (1) does have nonvanishing temporal correlations, as captured by the nonzero values of the a_1 and δ parameters in the AR(1) and FD models, respectively. On the other hand, for Kaufmann's signals, the residual variability is found to have no or very little memory for both stochastic model characterizations of u. We note that Kaufmann's approach involves a statistical model that explains the observed global mean temperature in terms of responses to solar, volcanic, and anthropogenic forcings and ENSO variability in such a way that the residual noise is uncorrelated [*Kaufmann et al.*, 2006, 2011]. The results in Table 1 suggest that this is not the case for the other three studies. We notice that Kaufmann's statistical model has been shown [*Kaufmann et al.*, 2013] to be the discrete-time counterpart to a zero dimensional energy balance model, indicating that the underlying physics in both models are consistent with each other.

### 3.2. Robustness Under Changes in Choice of Forced Responses

[34] In many previous detection and attribution studies [e.g., *Allen et al.* 2000; *Allen and Stott* 2003; *Allen et al.* 2006; *Hegerl and Zwiers* 2011], only natural (solar and volcanic) and anthropogenic signals were included into the decomposition of the temperature (T_i in equation (1)). Hence, key features of modes of internal variability, such as ENSO, were implicitly included in the climate noise u in equation (1). Within the context of the four simple statistical models analyzed in this paper, we can perform a simple test of the sensitivity of the detection statistics to the statistical model employed by removing the ENSO signal or equivalently assuming that the ENSO variability is simply part of the internal variability in the model. In other words, we perform the multiregression analysis in equation (1) but remove the ENSO signal as one of the T_is in this equation. The scaling factors obtained in this case are shown in Figure 3. The top panel corresponds to modeling the internal variability as an AR(1) process, while the bottom panel corresponds to the FD model. The dashed lines show the same scaling factors as in Figure 2, while the solid lines show the new scaling factors when ENSO is not taken into account

as an explicit signal. We find that even though its confidence interval increases in most cases, the detection of the anthropogenic signal is robust under the two assumptions, i.e., the corresponding scaling factor is not consistent with zero in both cases. The most noticeable difference between the two cases (solid versus dashed lines) is on the estimated scaling factors for the solar and volcanic signals (and AMO in the case of Folland). For the volcanic forcing, the scaling factors decrease in both panels while their confidence interval increases, resulting in the loss of detection in some cases. Consistently with *Stone and Allen* [2005] and Imbers et al. (submitted manuscript, 2013), the solar signal is not detected (β consistent with zero) for two of the four studies for both models of internal variability. These results are consistent with the fact that by removing ENSO as a forced response, the internal variability of the system increases, making it harder to detect a variable signal such as the solar one in a noisier background.

3.3. Potentially Missing Signals

[35] In the previous sections, we have discussed statistical models that aim to reproduce the behavior of the global mean surface temperature justifying the inclusion of the different signals using physical arguments: anthropogenic forcings are expected to affect the global mean temperature on multidecadal time scales, while ENSO and AMO are expected to influence the shorter-term interannual variability. We have assumed these choices to be adequate and assessed the robustness of the detection statistics to the characterization of internal variability and the influence of the ENSO forced response.

[36] Recently, some authors concerned by the ability of GCMs to depict long-term climate trends and cycles detected in geological data sets [*Loehle and Scafetta*, 2011] proposed to study the detection and attribution problem using an empirical decomposition of the global mean temperature that includes signals associated to solar cycles. Based on the periodicity in the solar signal detected by *Scafetta* [2010], *Loehle and Scafetta* [2011] introduced
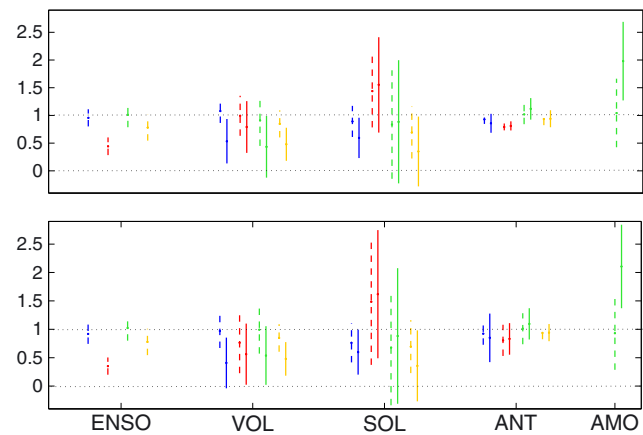
Figure 3. Same as in Figure 2 but illustrating the sensitivity of the detection statistics to ENSO, when its contribution to the global mean temperature is separated as a signal (dotted lines) or included in the residual internal variability u (thick lines).
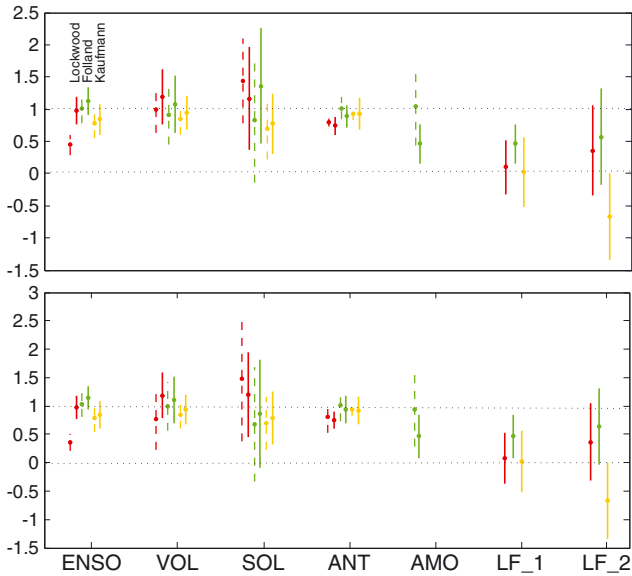
Figure 4. Comparison of scaling factors (vertical axis) using two different multiregression statistical models: (1) The same as in Figure 2 (dashed lines) and (2) adding explicitly 60 year (LF_1) and 20 year (LF_2) cycle signals separately (solid lines). The top panel corresponds to an AR(1) model for the internal variability, and the bottom panel to an FD model for the internal variability. The colors correspond to Lockwood, Folland, and Kaufmann signals as stated on the label.

cycles of 60 and 20 years associated to features of the solar motion and activity. The model was fit to HadCRUT3 global temperature data up to 1950 by nonlinear least squares estimation, assuming that the cycles have a fixed period but free phase and amplitude. Then the model was extrapolated from 1951 to 2010. In this paper, we use the model that best fits the observed record according to *Loehle and Scafetta* [2011] empirical decomposition.

[37] *Loehle and Scafetta* [2011] claim that GCMs do not include the 20 and 60 year low-frequency cycles [*Loehle and Scafetta*, 2011]. and thus the statistical robustness of detection and attribution methods that strongly rely on GCMs' simulations needs to be revisited. Since the physical mechanisms responsible for the cycles are not well understood [*Loehle and Scafetta*, 2011], it is difficult to evaluate whether or not these cycles are included in GCMs. This evaluation is beyond the scope of this paper. However, with our approach we can get some insight into what is the influence of the inclusion of these cycles on the detection statistics of the anthropogenic signal using the simplified models described in section 2.1.

[38] In order to do so, we build a matrix **T** in equation (1) for each of the studies above adding two new columns to account explicitly for the 60 and 20 year natural cycles following *Loehle and Scafetta*'s [2011] model and proceed to calculate the scaling factors and their variances following the generalized linear regression approach. Note that we are not able to use the data from Lean because this study starts in 1980, and due to the short length of the time series the resulting **T** is degenerate. The estimated scaling factors are shown in Figure 4. The top panel shows the scaling

factors assuming an AR(1) model for the internal variability, and the bottom panel shows the scaling factors when internal variability is modeled with a FD process. The dashed lines are the scaling factors shown in Figure 2 where no low-frequency cycles are included, and the thick lines are the new results. LF_1 and LF_2 indicate the scaling factors of the 60 and 20 year cycles, respectively. The results of this statistical study suggest that adding these two low-frequency oscillations does not change significantly the detection and attribution of the anthropogenic signal in any of the studies considered in this figure.

[39] The scaling factor that is more sensitive to the explicit inclusion of the 60 and 20 year natural cycles is the AMO (from Folland). The AMO has an oscillation of similar frequency to the low-frequency cycles from *Loehle and Scafetta* [2011]; therefore, part of the variability of the AMO is presumably represented by LF_1 and LF_2 in this statistical model. Note that for both the AR(1) and the FD cases, the AMO scaling factor remains detectable. On the contrary, we find that the LF_1 and LF_2 signals are not detectable when using Lockwood and Kaufmann's data and very close to including zero within the confidence interval when we use Folland's data. These results suggest that statistically these cycles are not significant and are not clearly distinguishable from the internal variability as described by our noise models.

[40] Note that the conclusions that one can draw from this analysis are heavily dependent on the goodness of fit achieved when new signals are included in the multiregression model. We have analyzed the residuals using standard statistical tests (Durbin-Watson test; *Kmenta* [1971]), and we have found that our model is still statistically adequate, but the results are less sound. In this particular case where there is no complete understanding of the physical mechanisms underlying these two cycles, one might conclude that these signals are not significant.

4. Conclusions

[41] In this paper, we have performed a sensitivity analysis of a range of recent empirical studies assessing the evidence for externally or internally forced temperature responses in the observed record of global mean temperatures. The external signals considered are responses to volcanic, solar, and anthropogenic forcings. Consistent with recent studies, we also include ENSO and AMO signals among the explanatory variables to assess whether we can still detect these key external influences, in order to be confident that these potentially confounding variables are not masquerading as externally driven signals.

[42] The danger of all such studies is over-fitting, given that multiple parameters are adjusted to reproduce the data. An indication of this problem would be high estimated errors on individual signal amplitudes. To check this, we place all the studies in a common statistical framework for a thorough error analysis.

[43] We show that key detection and attribution statements are robust to the inclusion of additional signals and also to the characterization of internal variability whether it is represented by a short memory AR(1) process or a long memory FD process. Hence, the detection of the anthropogenic signal is statistically robust independent of the model utilized to characterize the internal variability.

[44] We further extend the sensitivity analysis to include 20 and 60 year oscillations that have been proposed to explain recent apparent changes in rates of warming [*Loehle and Scafetta*, 2011]. We find that the detection of human influence on global temperature, and its magnitude, are insensitive to the inclusion of these natural cycles.

References

Allen, M., and P. Stott (2003), Estimating signal amplitudes in optimal fingerprinting, part I: Theory, *Clim. Dynam.*, *21*(5), 477–491, doi: 10.1007/s00382-003-0313-9.

Allen, M., and S. Tett (1999), Checking for model consistency in optimal fingerprinting, *Clim. Dynam.*, *15*(6), 419–434.

Allen, M., et al. (2000), Quantifying the uncertainty in forecasts of anthropogenic climate change, *Nature*, *407*(6804), 617–620.

Allen, M., et al. (2006), Quantifying anthropogenic influence on recent near-surface temperature change, *Surv. Geophysics*, *27*, 491–544, doi: 10.1007/s10712-006-9011-6.

Beran, J. (1994), *Statistics for Long-Memory Processes*, vol. 61, Chapman & Hall/CRC, Boca Raton, Florida.

Brohan, P., J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones (2006), Uncertainty estimates in regional and global observed temperature changes: A new dataset from 1850, *Journal of Atmospheric and Solar-Terrestrial Physics*, *111*, D12,106, doi:10.1029/2005JD006548.

Folland, C., A. Colman, J. Kennedy, J. Knight, D. Parker, P. Stott, D. Smith, and O. Boucher (2011), High predictive skill of global surface temperature a year ahead, in *AGU Fall Meeting Abstracts*, vol. 1, p. 01.

Hasselmann, K. (1979), On the signal-to-noise problem in atmospheric response studies, *Meteorology of Tropical Oceans*, 251–259.

Hegerl, G., and F. Zwiers (2011), Use of models in detection and attribution of climate change, *WIREs Clim. Change*, *2*, 570–591, doi: 10.1002/wcc.121.

Huybers, P., and W. Curry (2006), Links between annual, Milankovitch and continuum temperature variability, *Nature*, *441*(7091), 329–332.

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

Kaufmann, R., H. Kauppi, M. Mann, and J. Stock (2011), Reconciling anthropogenic climate change with observed temperature 1998–2008, *Proc. Natl. Acad. Sci.*, *108*(29), 11,790.

Kaufmann, R., H. Kauppi, M. Mann, and J. Stock (2013), Does temperature contain a stochastic trend: Linking statistical results to physical mechanisms, *Clim. Change*, 1–15, doi:10.1007/s10584-012-0683-2.

Kmenta, J. (1971), *Elements of Econometrics*, Macmillan, New York.

Kopp, G., and J. Lean (2011), A new, lower value of total solar irradiance: Evidence and climate significance, *Geophys. Res. Lett.*, *38*, L01,706, doi: 10.1029/2010GL045777.

Lean, J., and D. H. Rind (2008), How natural and anthropogenic influences alter global and regional surface temperatures: 1889 to 2006, *Geophys. Res. Lett.*, *35*, L18,701, doi:10.1029/2008GL034864.

Lean, J., and D. H. Rind (2009), How will Earth's surface temperature change in future decades? *Geophys. Res. Lett.*, *36*, L15,708, doi: 10.1029/ 2009GL038932.

Lee, T., and T. Ouarda (2011), Prediction of climate nonstationary oscillation processes with empirical mode decomposition, *J. Geophys. Res.*, *116*(D6), D06,107, doi:10.1029/2010JD015142.

Lee, T., and T. Ouarda (2012), An EMD and PCA hybrid approach for separating noise from signal, and signal in climate change detection, *Int. J. Climatol.*, *32*(4), 624–634, doi:10.1002/joc.2299.

Lockwood, M. (2008), Recent changes in solar outputs and the global mean surface temperature. III. Analysis of contributions to global mean air surface temperature rise, *P. Roy. Soc. A: Math. Phy.*, *464*(2094), 1387–1404, doi:10.1098/rspa.2007.0348.

Loehle, C., and N. Scafetta (2011), Climate change attribution using empirical decomposition of climatic data, *Open. Atmos. Sci. J.*, *5*, 74–86, doi: 10.2174/1874282301105010074.

Percival, D., J. Overland, and H. Mofjeld (2001), Interpretation of North Pacific variability as a short-and long-memory process*, *J. Clim.*, *14*(24), 4545–4559.

Scafetta, N. (2010), Empirical evidence for a celestial origin of the climate oscillations and its implications, *J. Atmos. Sol. Terr. Phys.*, *72*(13), 951–970, doi:10.1016/j.jastp.2010.04.015.

Stone, D., and M. Allen (2005), Attribution of global surface warming without dynamical models, *Geophys. Res. Lett.*, *32*, L18,711, doi: 10.1029/2005GL023682.

Vyushin, D., and P. Kushner (2009), Power-law and long-memory characteristics of the atmospheric general circulation, *J. Clim.*, *22*(11), 2890–2904, doi:10.1175/2008JCLI2528.1.

Vyushin, D., P. Kushner, and F. Zweirs (2012), Modeling and understanding persistence of climate variability, *J. Geophys. Res. Atmos.*, *117*, 1984–2012.

Wu, Z., N. Huang, S. Long, and C. Peng (2007), On the trend, detrending, and variability of nonlinear and nonstationary time series, *Proc. Natl. Acad. Sci.*, *104*(38), 14,889–14,894., doi:10.1073/pnas.0701020104.