Sensitivity of climate change detection and attribution to the characterization of internal climate variability

Jara Imbers *

OCCAM, Mathematical institute, University of Oxford

Ana Lopez

Center for the Analysis of time series, London School of Economics

Chris Huntingford

Centre for Ecology and Hydrology

Myles Allen

Atmospheric, Oceanic and Planetary Physics, University of Oxford

* *Corresponding author address:* Jara Imbers, OCCAM , Mathematical Institute, 24 - 29 St Giles', Oxford, OX1 3LB.

E-mail: imbers@maths.ox.ac.uk

ABSTRACT

The Intergovernmental Panel on Climate Change (IPCC) "very likely" statement that anthropogenic emissions are affecting climate is based on a statistical detection and attribution methodology that strongly depends on the characterization of internal climate variability. In this paper, we test the robustness of this statement in the case of global mean surface air temperature, under different representations of such variability. The contributions of the different natural and anthropogenic forcings to the global mean surface air temperature response are computed using a box diffusion model. Representations of internal climate variability are explored using simple stochastic models that nevertheless span a representative range of plausible temporal autocorrelation structures, including the short-memory first-order autoregressive (AR(1)) process and the long-memory fractionally differencing (FD) process. We find that, independently of the representation chosen, the greenhouse gas signal remains statistically significant under the detection model employed in this paper. Our results support the robustness of the IPCC detection and attribution statement for global mean temperature change under different characterizations of internal variability, but also suggest that a wider variety of robustness tests, other than simple comparisons of residual variance, should be performed when dealing with other climate variables and/or different spatial scales.

1. Introduction

At the centre of the climate change debate is the question of whether global warming can be detected, and if that is the case, whether or not it can be attributed to anthropogenic causes. Optimal fingerprinting is a powerful method of detection and attribution of climate change (Hasselmann 1979, 1993; Hegerl et al. 1996) used widely in this area of research. In essence, optimal fingerprinting is a multi-regression analysis that searches for the observed temperature record response to external drivers or forcings such as changing levels of greenhouse gases, and aerosol loading (human-induced), volcanic activity and variations in solar radiation (naturally induced). A key input in the procedure of fitting this multiple regression model is an estimate of the internal variability of the climate system, against which the statistical significance of anthropogenic and natural signals must be compared. Hence, an accurate depiction of this variability is crucial for the robustness of the results.

In this work we refer to internal variability as the characterization of the variations in the climate system that would occur in the absence of natural or anthropogenic forcings, solely due to the coupling of atmosphere, ocean, biosphere and cryosphere dynamics. In most cases Global Climate Models (GCMs) are used to estimate climate internal variability due to the fact that the instrumental records are both, too short to provide a reliable estimate, and contaminated by the effects of external forcings. Typically, long GCM control simulations are employed for this purpose. This is such a key step in the process of detecting and attributing climate change that in fact, for some authors (e.g. (Huybers and Curry 2006)), the debate surrounding global warming centers on the uncertainties in the structure and magnitude of the internal variability of the climate system.

Previous studies (Allen and Stott 2003; Huntingford et al. 2006) used increasingly sophisticated variations of the multiregression technique in order to quantify the statistical significance of the anthropogenic signal in temperature trends as simulated by a range of climate models. In these studies, long GCMs control simulations are used to estimate internal variability on the temporal and spatial scales that are retained in the analysis. Although these

authors are careful to attempt the inclusion of model uncertainty in the regression model, and test the robustness of their results under changes in the amplitude of the estimated internal variability, it is not clear whether or not other aspects of the internal variability poorly represented by the climate models (Newman et al. 2009; DelSole and Shukla 2010; Klein et al. 1999) do bias statistical estimations of the significance of the anthropogenic signal in the observations.

In this paper we investigate this question by assuming that the internal climate variability can be represented by a stochastic process that includes, apart from a white noise component, 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 (also named climate persistence by some authors (Beran 1994; Percival et al. 2001). Understanding and characterizing the memory of the climate system is problematic due to the short length of the observational records when compared to the wide range of interconnected timescales. In fact, numerous explanations have been advanced regarding internal variability, e.g. Wunsch (2003); Mitchell et al. (1976); Hays et al. (1976) but the full characterization of its properties and its interplay with external forcings remains elusive (Ghil 2012).

We use two different stochastic models to represent internal variability: an auto-regressive model of the first order (AR(1)) and a fractionally differencing model (FD) (Percival et al. 2001). These correspond to the two simplest stochastic models (minimal number of parameters) that can represent significantly different assumptions about the internal temporal structure of the system they describe. While the AR(1) model has the short memory characteristic of an exponentially decaying autocorrelation function, the FD model has the long memory associated to an algebraically decaying autocorrelation function. These two models have been considered before as two different but plausible(e.g. Hasselmann (1979), Vyushin and Kushner (2009)) characterizations of the climate internal variability in terms of equally simple parametric models. In addition, choosing these simple models allows us to carry out

a sensitivity analysis of detection and attribution to well defined parameters whose change is easily understood in terms of memory or unresolved variability (white noise) in the climate system.

The paper is organized as follows. In Section 2 we describe the data analyzed. We briefly discuss the detection and attribution approach as applied to the one dimensional climate model used in our study, and the two stochastic models, exploring the arguments to justify using each of them to represent internal climate variability. In Section 3, we discuss how the significance of the anthropogenic signal depends on the model chosen to represent internal variability. We include an analysis of how consistent are our estimates of internal variability with the ones estimated from the CMIP3 control simulations, in order to evaluate whether or not the use of these control runs for detection and attribution can potentially bias the results. Section 4 is devoted to the conclusions.

We remark that our goal is to explore the sensitivity of the detection and attribution statistics to the representation of internal variability. Therefore, the main assumptions of detection and attribution of climate change, namely that the forced responses can be linearly superimposed on internal variability and that there are no interactions between forced and unforced variability , are assumed to be valid.

2. Data and Method

We analyze the problem of the sensitivity of detection and attribution results to internal variability in the simplest case, i.e. for the global mean surface air temperature as simulated by a one dimensional climate model.

To estimate the temperature responses to individual forcings we use the box diffusion model (BDM) described in Andrews and Allen (2008) and Allen et al. (2009), which can be written as

$$c\frac{dT}{dt} = F - \lambda T - \frac{c}{d_{ml}}\sqrt{\frac{\kappa}{\pi}}\int_0^t \frac{dT(t')}{dt'}\frac{dt'}{\sqrt{(t-t')}} \tag{1}$$

where T is the global mean temperature and F is the external forcing. In Allen et al. (2009), the heat capacity $c = 7.22 \frac{W\,y}{m^2\,^\circ K}$ corresponds to the heat capacity of an ocean mixed layer of depth $d_{ml} = 75m$ assuming that the ocean covers 70% of the Earth surface. Best estimates for the climate feedback parameter λ and effective ocean diffusivity κ are determined using the linear temperature trend attributable to the increase in greenhouse gases over the 20th century based on fingerprint attribution results (Stott et al. 2006), and the effective heat capacity of the atmosphere-land-ocean system implied by the combination of observed surface warming (Brohan et al. 2006) and the total ocean heat uptake over the period 1955-98 (Levitus et al. 2005). This results in $\kappa = 0.10 \frac{(mld)^2}{y} = 562.5 \frac{m^2}{y}$, and $\lambda = 1.29 \frac{W}{m^2\,^\circ K}$.

This BDM, with the specified parameters, can then be used to find the temperature responses T_i to different forcings: volcanic (VOL), solar (SOL), greenhouse gases (GHG), sulphates (SUL), and all anthropogenic forcings together (ANT). In this way, the temperature responses to individual forcings are computed without relying on GCMs.

Note that observed surface temperatures are used in the estimation of parameters in this model, albeit indirectly through the fingerprint results and estimates of effective heat capacity. The main impact of varying parameters in the model, however, is to change the magnitude of the responses to different forcings. The shape, or time-evolution, of the response is primarily driven by the forcings themselves. In our subsequent analysis, we use only the temporal shape of the responses, not their magnitude, hence minimizing the risk of "double-counting" of data.

The forcings time series required to estimate the corresponding temperature responses using the model in Eq. (1) are obtained from the CMIP5 recommended data sets [1] (Meinshausen et al. 2011). To carry out the detection and attribution analysis observed time series of annual global mean temperature are required. We use the observed data from HadCRUT3 [2] for the period $1850 - 2005$ (Brohan et al. 2006), and the HadCRUT4 [3] (Morice et al. 2012)

[1] http://www.pik-potsdam.de/ mmalte/rcps/

[2] http://www.cru.uea.ac.uk/cru/data/temperature/

[3] http://www.metoffice.gov.uk/hadobs/hadcrut4/data/current/download.html

data set to test the sensitivity of the results to the addition of the last seven years of observations up to 2012 (see section 3). Uncertainties in observed temperatures and estimates of forcings are ignored in this paper.

We additionally use the World Climate Research Programme (WCRP) CMIP3 multi-model archive of control simulations to study the internal variability simulated by the state of the art climate models (Solomon 2007). For completeness, we have used all the control simulations, regardless of drifts. We will comment on the effect of drifts in the control segments on the final results in Section 3.

(i) Detection and Attribution

The detection of climate change is the process of demonstrating that climate has changed in some well defined statistical sense, without providing a reason for that change. Attribution of causes of climate change is the process of establishing the most likely causes for the detected change with some defined level of confidence (Solomon 2007). In this work we aim to detect and attribute climate change by estimating the contribution to the observational record T_{obs} of each of the response temperatures T_i calculated using Eq.(1). In other words, we want to obtain the amplitudes β_i in the following expression:

$$T_{obs} = T\beta + u, \tag{2}$$

where T is a matrix with $n+1$ columns including the n forced responses T_i, and a constant term to remove the mean. u is an stochastic term that represents the internal climate variability with covariance matrix is given by $\Omega = E(uu^{\dagger})$. Under the assumption that u is multivariate normal (Allen and Tett 1999), the optimal scaling factors, $\beta = (\beta_1, \beta_2, ..\beta_{n+1})$ are given by (Kmenta 1971):

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

and their variance :

$$V(\hat{\beta}) = \left(T^{\dagger}\Omega^{-1}T\right)^{-1}, \tag{4}$$

where † is used to denote the transpose of a matrix.

In this work, following standard detection and attribution studies, we consider the following external forcings: greenhouse gases, sulphates, volcanic and solar. It has long been recognized however, that the detection and attribution results are sensitive to the omission of potentially important forcings and/or internal modes of variability. Likewise, if signals that have some degree of collinearity are included, this can affect the robustness of the results. This will be tested in Section a by performing the detection and attribution study considering solar, volcanic, and all anthropogenic (ANT) forcing together instead of separating greenhouse gases and sulphates into two different signals. The robustness of the detection and attribution statistics to separating other modes of internal variability such as ENSO or AMO (Atlantic Multi-decadal Oscillation) from the noise u in Eq.(2) has been analyzed in references Zhou and Tung (2013) and Imbers et al. (2013). In particular, in Imbers et al. (2013) the forced temperatures responses to anthropogenic, solar, volcanic and ENSO and/or AMO, are obtained from a series of studies that use different statistical models to single out each forced temperature response. Using the same approach as in this paper, it is found that the ANT detection statistic is robust in all cases.

Typically, detection of anthropogenic climate and its attribution to external forcings requires defining space and time dependent response patterns (Solomon 2007; Stone et al. 2007). These patterns are obtained from GCMs transient simulations. On the other hand the spatio-temporal structure of internal variability in Ω is estimated from averaging GCMs' control simulations over space, time and model ensembles. These calculations are high dimensional and require sensible truncation of the space and time domain using techniques such as principal components analysis.

In this paper we use a simpler version of the detection and attribution approach since we analyze only the global mean surface temperature, introducing parametric models to charac-

terize the global mean internal variability u explicitly as a stationary stochastic process. In other words, we formulate the detection and attribution problem as in Eq.(2) but with u a function of stochastic parameters that are estimated simultaneously with the scaling factors $\hat{\beta}$ using a minimum squared error algorithm.

The first challenge is to choose an adequate stochastic representation for the internal variability. The difficulties finding the appropriate stochastic model are due to the uncertainties in characterizing internal variability from the observational record, which as discussed before, is contaminated by the external forcings and too short relative to the long time scales potentially relevant to the current climate variability . In particular, in the observed record it is not clear how to separate the decadal from centennial or even longer time scales (Percival et al. 2001). Given these uncertainties in the characterization of the internal climate variability we choose to describe it using two models that span a wide range of plausible temporal autocorrelations (Vyushin and Kushner 2009). This choice is important to address the fact that GCMs simulations do not necessarily capture all the modes of internal variability in the system, certainly not variability at longer time scales than centennial. We then choose stochastic processes that allow to explore how the results of detection and attribution of climate change would change if the internal variability has either long or short memory, and assume that this is a necessary (not sufficient) test to evaluate the robustness of the results under a wide range of plausible characterizations of the memory of the climate system.

(ii) Short memory process: AR(1)

The best known and simplest stochastic representation for discrete geophysical time series is the AR(1) model (Ghil et al. 2002; Bretherton and Battisti 2000). In the continuous time domain the AR(1) process corresponds to diffusion, which in turn, is the simplest possible mechanism of a physical process with inertia and subjected to random noise. In the context of climate, this model was first introduced by Hasselmann (1979) to describe the internal variability of the climate system under the assumption of time scale separation

between oceanic and atmospheric dynamics. In this framework, the faster dynamics of the atmosphere can be modeled as white noise acting on the slower and damped dynamics of the ocean. Thus, the AR(1) is the simplest model that can explain the "weather " and the "climate" fluctuations as two components of the internal variability. Mathematically, the AR(1) is a stationary stochastic process that can be written as:

$$u_t = a_1 u_{t-1} + a_0 \epsilon_t \tag{5}$$

where $E(u_t) = 0$, a_1 and a_0 are parameters, and ϵ_t represents white noise, i.e. $E(\epsilon_t \epsilon_{t'}) = \delta_{tt'}$. The autocovariance function of this process is determined by a_0 and a_1 as follows:

$$\omega_{AR1}(\tau) = \frac{a_0^2}{1 - a_1^2} a_1^{|\tau|} \tag{6}$$

where τ is the time lag. Notice that a_1 controls the decaying rate of the autocorrelation function and in that sense we can associate it to the *memory* of the system. On the other hand a_0 is related to the amplitude of the white noise in the system. From Eq.(6) the covariance matrix Ω results:

$$\Omega_{i,j}^{AR} = \frac{a_0^2}{1 - a_1^2} a_1^{|i-j|} \tag{7}$$

Eq.(5) models the memory of the process such that at a given time t the state of the system is a linear function of the previous state $(t - 1)$ and some random noise with amplitude a_0^2 jittering, and hence moving the system away from equilibrium. The autocovariance of the process, Eq.(6), decays exponentially with time, so the system has always a much better memory of the near past than of the distant past. a_1 can take any value in the interval $[0, 1)$, $a_1 = 0$ represents the limit in which the system is purely white noise, and $a_1 \to 1$ is the extreme in which a system is dominated by inertia. In our case, we are characterizing annual global mean temperature internal variability with this model, so we are trying to quantify the impact of the natural fluctuations in the year to year variation.

In the detection and attribution analysis, the parametric form of the covariance matrix, Eq.(7), is used to simultaneously determine the optimal scaling factors β_i in Eq.(3) and

the parameters a_1 and a_0 of the climate noise in Eq.(5) following the Hildreth-Lu method (Kmenta 1971).

(iii) Long memory process: FD

There is 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)). Different power-law behaviors have been identified in globally and hemispherically averaged surface air temperature (Bloomfield 1992; Gil-Alana 2005), station surface air temperature (Pelletier 1997) and temperature paleoclimate proxies (Huybers and Curry 2006). These findings suggest that in order to thoroughly test the sensitivity of the detection and attribution statements to the representation of internal variability, modeling it with other than a short memory process such as the AR(1) model might be in order. We then alternatively assume that the global mean temperature internal variability autocorrelation decays algebraically, allowing for all time scales to be correlated. This long time correlation will clearly have an effect in the statistical significance of the anthropogenic signal (see Eq.(4)).

Long memory models were motivated initially by hydrology studies (Hurst 1951, 1957) and have been employed to model paleoclimatic time series (e.g., Huybers and Curry 2006; Pelletier 1997). An spectrum corresponding to algebraic decaying correlations can be constructed for a prescribed range of frequencies as the sum of AR(1) processes or as solutions of more complex stochastic differential equations (Erland et al. 2011; Kaulakys et al. 2006; Granger 1980). Therefore, a plausible justification to use a long memory process to represent the internal variability of the global mean temperature is that it could be thought as the result of the superposition of several diffusion processes (AR(1)).

Applying the law of parsimony, we choose a long memory process with the same level of complexity as the AR(1) model. The fractional differencing (FD) model (Beran 1994; Percival et al. 2001; Vyushin and Kushner 2009; Vyushin and P.J. Kushner 2012) is defined as a stationary stochastic process with zero mean u such that:

$$u_t = (1 - B)^{-\delta} \epsilon_t. \tag{8}$$

where B is the backshift operator, i.e. $Bu_t = u_{t-1}$ (Beran 1994). The model is fully specified by the parameters δ and the standard deviation σ_e of the white noise ϵ_t. The autocovariance function is given by the equation:

$$\omega_{FD}(\tau) = \frac{\sigma_e^2 \sin(\pi\delta)\Gamma(1-2\delta)\Gamma(\tau+\delta)}{\pi\Gamma(\tau+1-\delta)} \tag{9}$$

As a result the covariance matrix becomes,

$$\Omega_{i,j}^{FD} = \frac{\sigma_e^2 \sin(\pi\delta)\Gamma(1-2\delta)\Gamma(|i-j|+\delta)}{\pi\Gamma(|i-j|+1-\delta)}. \tag{10}$$

For large τ the autocorrelation function satisfies $\lim_{\tau\to\infty} \omega_{FD}(\tau) = |\tau|^{2\delta-1}$ (Beran 1994). From this expression one can see that the autocorrelation decays algebraically, thus the name "long memory". Since δ controls the decaying rate of the autocorrelation function it can be associated to the *memory* of the system, while σ_e is characterizes the amplitude of the white noise.

Similarly to the $AR(1)$ case, we use this covariance matrix, Eq.(10), and Eq.(2) and Eq.(3) to simultaneously determine the scaling factors β_i and the parameters δ and σ_e following the Hildreth-Lu method (Kmenta 1971).

3. Results

a. Robustness of detection statistics

In order to test the robustness of the detection statistics, we find simultaneously the scaling factors β_i and the stochastic parameters of the internal variability u, using generalized linear regression to solve Eq.(2). Notice that when u is modeled as an AR(1) or an FD process, the noise covariance matrix Ω in Eq.(3) and Eq.(4) is given by Eq.(7) or Eq.(10) respectively. The best estimates of the scaling and noise parameters are chosen as those that

minimize the residual white noise in u (Kmenta 1971). Using the Akaike Information Criteria we find that both models for u are equally skilful at representing the internal variability given the observational record used in our analysis.

Fig.(1) shows the values of the optimal scaling factors with their 95% confidence intervals using the AR(1) (grey line) and the FD (black line) models, when T_{obs} is the HadCRUT3 global mean temperature record for the period 1850-2005. In the detection and attribution approach, a signal is detected when the corresponding scaling factor is different from 0 with 95% confidence, while the attribution of a signal requires confidence intervals that include one (Allen and Stott 2003; Allen and Tett 1999). When the scaling factors are larger or smaller than one, the simulated forced responses are assumed to be over or under estimated by the climate model used to simulate them.

Therefore, in order to test the robustness of the detection statistics , we need to evaluate the statistical significance of the scaling factors and the uncertainty in the determination of the stochastic models' parameters. A scaling factor β is defined as statistically significant if the null hypothesis ($\beta = 0$) can be rejected with 95% confidence. A standard approach to find the confidence interval is to define the $z\_score = \beta i / V(\beta_i))^{\frac{1}{2}}$; if this quantity i sampled from, for instance, a t-distribution with more than 60 degrees of freedom, the scaling factor β will result statistically significant with 95% confidence when $z\_score \geq 2$ (Kmenta 1971). In our analysis, due to the correlations present in the noise models, an estimation of the number of degrees of freedom is problematic. We use instead a Monte Carlo approach that allows a testing of the null hypothesis as follows. For each of the noise models (AR(1) or FD) we generate , using the optimal values of the models'parameters, 1000 surrogate samples with the same length as the observed record (156 years). We then replace T_{obs} in Eq.(2) by each of these surrogate series, and perform the generalized linear regression with the four forced responses on the right hand side of the equation; the aim of this exercise being to estimate the optimal values of the scaling factors β and the u parameters that best fit each of the surrogate series. We can then perform an empirical evaluation of the null hypothesis: for any

given value of the z_score or, equivalently, the size of the confidence interval, we aim to find what is the proportion of cases where the scaling factor β is different from 0. In particular the value of the z_score that gives β different from 0 in at most 5% of the cases determines the 95% confidence interval. We find that for the GHG signal the z_score is 2.22 in the case of the AR(1) model and 2.45 in the case of the FD model. In addition, and since we expect that due to the stochastic nature of the noise models there will be some uncertainty in the determination of their parameters, the values of the noise model parameters estimated with this Monte Carlo approach provide an estimate of the uncertainty of the best fit noise model parameters when regressing the forced responses on T_{obs} in Eq.(2).

Fig.(1) shows that for our detection model, the greenhouse gas signal is detected and attributed, the volcanic signal is only detected, and the solar signal is not detected nor attributed for both models of internal variability. In the case of the sulphates forcings, the result depends on the representation of the internal variability.

The robustness of the GHG signal detection can be analyzed using Fig.(2) when the internal variability is characterized by the AR(1) model or by the FD model in the upper or lower panels respectively. The horizontal and vertical axes show the white noise amplitude and memory parameters respectively, and the contour lines indicate the significance level of the scaling factor β_{GHG}. The diamond symbol shows the best fit of internal variability (for each model) when the observed record T_{obs} is the HADCRUT3 data for the period 1850-2005. The uncertainty in the estimation of the best fit, computed using the Monte Carlo approach, is shown as the grey cloud of points. It is clear that even when taking into account this uncertainty in the parameters, the significance of the detection of the greenhouse gas signal is not affected.

As expected, the significance of the greenhouse gas signal is lower when we represent the internal variability as an FD than as an AR(1) process. We find that both stochastic models' best fit have similar white noise amplitude, showing that statistically they are similarly good at explaining variability, given that this is the residual of the linear fit. The bigger difference

between the two models arises in the memory parameter.

In the case of the AR(1), a_1 is bounded between $a_1 = 0.25$ and $a_1 = 0.70$, and the best estimate is $a_1 = 0.53$. In a short-memory process we can translate these values into a decay time, which is a well defined time scale given by $\tau = -1/\ln(a1)$ in units of years (Kmenta 1971). Using the range of values of a_1 above, the uncertainty in the decay time remains below ten years. This means that, according to the AR(1) model, we can explain the fluctuations of internal variability by being affected mainly by the previous few years and some random white noise.

In the case of the FD model, the uncertainty in the estimation of δ is much larger and spans almost all the allowed values from nearly white noise, $\delta = 0.12$, to $\delta = 0.5$, with a best estimate of $\delta = 0.43$. All these values are broadly consistent with the result in Huybers and Curry 2006. In particular, we find that there is a 10% probability of the estimated parameter corresponding to a non-stationary process (i.e $\delta \to 0.5$). Using Eq.(10), we find that, for δ close to the limiting value 0.5, the amplitude of the resulting variations would be inconsistent even with relatively high-variance reconstructions of paleo-climatic data over the past 1-2 millennia (Esper et al. 2012). The presence of poorly-known forced responses on these timescales makes it difficult to use paleo-climate data as an explicit quantitative constraint in our analysis, but it does suffice to indicate that the relative stability of the climate of the Holocene would be unlikely if internal variability of the climate system were to conform to an FD process with very high values of δ. Based on this arguments, we can ignore the values of δ close to the non-stationary limit ($\delta \to 0.5$). Furthermore, in Appendix A we explore how the estimation of the stochastic model parameters depends on the length of the time series considered and show that values of the parameters corresponding to non-stationary processes are likely to be an artifact of the short length of the time series.

For an algebraically decaying autocorrelation function there is no associated time scale, therefore a long memory process does not have a decay time (Beran 1994). Nevertheless, to have an intuition about the time scales associated to particular values of δ, one can calculate

the time it takes for the autocorrelation function to reduce to $1/e$ of its initial value (in analogy with the e-folding time for the AR(1) model). For the best fit value of $\delta = 0.43$ for instance, this calculation gives a much longer time than the length of the observational record (156 years). This suggests that, according to this model, in the 156 years long record all points are highly correlated. Overall, we find that, despite the very different time scales that are relevant for the AR(1) and FD characterizations of internal variability, the GHG signal detection statistics is robust for both models.

One interesting question that can be explored using our results is how wrong one would have to get the model parameters of the internal variability in order to change the detection statement of the greenhouse gas signal. In the case of the AR(1) model we find that the greenhouse gas signal would become not statistically significant in a world in which higher values of a_1 and/or a_0 were needed to describe internal variability. In the upper panel of Fig.2 we see that, to loose statistical significance, one would have to increase the time correlation characterized by a_1 to more than 0.8 , or triple the white noise parameter a_0.

Hence, the detection statistics for the AR(1) model is very sensitive to the memory parameter and relatively less sensitive to the amount of white noise in the process. Thus, in terms of the global mean temperature internal variability as simulated by GCMs, our findings suggest that the relevant aspect that should be taken into account in a robustness test should be the models' ability to capture correctly the temporal correlations more than the total variance, which is in turn conditioned by their ability to capture the most relevant dynamical processes, their couplings and feedback mechanisms.

For the FD process we find a different result. In the lower panel of Fig.(2) we can see that for the estimated σ_e there is no δ for which the process has a greenhouse gas scaling factor which is not statistically significant. Thus, this suggests that the greenhouse gases detection results are robust under changes in the memory parameter. In fact, for very high values of δ, one would still need to double the amplitude of the white noise to change the detection statistics. Results in Appendix A suggest that with a longer observational record

we would have estimated a smaller δ , but also, that the estimated white noise amplitude would not increase significantly. This suggests that if the observed record is relatively short to accurately characterize the memory in the FD model, the detection of the greenhouse gas signal would still be robust when the length of the record increases. In terms of the global mean temperature simulated by GCMs, our results using an FD model suggest that the emphasis should be placed on accurately depicting the amplitude of the white noise in order to be confident about the detection and attribution statistics.

In conclusion our results suggests that, in the presumably more realistic case in which the internal variability of the global mean temperature is best characterized by a process whose temporal structure lies somewhere in between that of an AR(1) and of an FD process, both its temporal correlation structure and the white noise amplitude are important for assessing the robustness of the signals.

To close this section we include a brief discussion about the robustness of the detection results to the inclusion of the last seven years of observations (up until 2012), and the potential effect of the collinearity of the greenhouse gas and sulphates temperature responses. We used the HadCRUT4 data set to include the last seven years of data in T_{obs}. Differences between the HadCRUT3 global mean temperature time series and the median of the HadCRUT4 global mean temperature time series result in slightly different values of the scaling parameters (see Table 1). Therefore, in order to ensure that the results are comparable, we use only HadCRUT4 data to analyze potential dependencies on the inclusion of the more recent observations and to the number of signals considered.

Figure (3) shows the results of this sensitivity analysis. We observe that if instead of SUL and GHG signals, we only consider a single total ANT signal, the scaling factors for the latter are smaller than the GHG ones (see Table (1)). However the ANT signal remains detectable for both characterizations of internal variability. Similarly, when adding the last seven years of the observed record, the GHG and the ANT signals remain detectable for both noise models, but attribution is lost in the case of the GHG.

b. CMIP-3 control runs

In this section we use the same techniques as above to evaluate the control simulations used in the detection and attribution of climate change included in the 4^{th} Assessment Report of the IPCC. Our goal is to get some insight about the controls' potential limitations to estimate internal variability and how this might impact in the robustness of the detection and attribution statistics.

We take annual global mean temperature segments from the CMIP3 control simulations that have the same length as the observational record, 156 years, and fit them to an AR(1) and a FD model. Thus, we characterize each control by a set of parameters, a_1 and a_0^2 when using an AR(1) model, and δ and σ_e^2 for the FD model. The results are indicated in Fig.(2) by numbers representing different GCMs' control segments (see table (2)). In both cases the spread of points is larger than the spread of the Monte Carlo experiment that characterizes the uncertainty in the estimation of the parameters of the internal variability for the observed record. Note that the number of control segments for each GCM depends on the available number of years of the control simulation in the CMIP3 database. We have taken segments of controls which are fully non-overlapping and assume that they are independent realizations. Only in the case of the HadCM3 and the CCMA-CGCM3 models the number of segments (identified by numbers 22 and 1 respectively in both panels of Fig.(2)) is large enough to get some intuition about the uncertainty in the estimate of the parameters for those particular GCMs' controls. The spread of the points corresponding to each of these two models suggests that, have we had many more control segments, the uncertainty in the estimation of the parameters would have been given by a cloud of points with a similar spread as the uncertainty estimate for the parameters corresponding to the observational record (grey cloud); and that both uncertainty estimates would have had a significant overlap. However, there are other models for which this is not the case.

The control segments we are investigating are not identical to the ones used in the detection and attribution studies, as their intradecadal variability is typically smoothed (5 year

means) and segments with drift are discarded (Stone et al. 2007). The argument for smoothing the temporal variability of the control segments for the 4^{th} Assessment of the IPCC is that some modes of internal variability such as ENSO (with a 2-7 years characteristic scale of variability) are often not properly depicted by all the GCMs, and this would subsequently introduce errors in the estimate of the covariance matrix. In addition, control segments with drifts are discarded attributing the drifts to numerical errors. In our case, the control segments with a drift (Stone et al. 2007) are few and correspond to those with the highest memory parameter values. In the case of the FD model, these are the segments with $\delta = 0.5$ in the lower panel of Fig. (2) and in the case of the AR(1) model the a_1's are such that, they all lie around the contour line for which the GHG scaling factor is not statistically significant (thick blue contour line).

Interestingly, we find in Appendix A that there is a very high correlation between the estimates of a_1 and δ and the amplitude of white noise for for any given control segment. From the point of view of our analysis, one reason for this is that both stochastic models can separate the same amount of correlated data from the white noise, and each model explains the dynamics with a different memory parameter according to the relevant covariance matrix. As a result, although the underlying physical assumptions are very different in these two models, we find that the numerical value of the autocorrelation function of both models are very similar for the first 156 years as expected.

To finish this section, we analyze the power spectra of the climate models' control runs and the observations. In Fig.(4) we show the power spectra of the CMIP3 models' control segments and the power spectra of the residuals from the best fit to T_{obs}, i.e., $T_{obs} - \hat{\beta}T$. The latest residuals are the unexplained fluctuations of the climate in our model, after removing the temperature response to the forcings. The figure shows that the power spectra of the residuals are very similar independently of whether the internal variability is characterized as and AR(1) (thick grey line) or an FD process (black line). Since the power spectrum is the Fourier transform of the autocorrelation function, finding similar power spectra is

equivalent to finding similar covariance matrices; hence this figure is consistent with our previous findings about the similarity in magnitude of the autocorrelation functions of the fitted internal variability to the 156 years observed record. It is clear that a much longer time series is required to appreciate more significant differences in the variability simulated by the two stochastic models.

We can also analyze the link between the ability of a GCM to model different modes of internal variability and the implications for the significance of detection and attribution. It is clear from Fig.(4) that some control segments display peaks corresponding to the ENSO signal with unrealistic high amplitudes , as shown by the high power at the 2-5 years frequency range. However, Fig.(2) shows that most of these control segments fall in the area of the plots that correspond to a significant greenhouse gas signal. Consistently wit the findings in Allen and Tett (1999), this analysis suggests that an accurate depiction of all modes of internal variability might not be required to ensure the robustness of the detection statistics under our detection model.

Finally, our analysis point towards the need to develop a wider range of techniques to assess the robustness of detection and attribution of climate change. The "consistency test" described in Allen and Tett (1999) is equivalent to look at the power spectra of GCMs runs and compare their (typically) decadal internal variability with the decadal internal variability retained in the residuals of the fit to the observed record. The aim of this test is mainly to discard the possibility of over-attributing climate change to the anthropogenic signal only because climate models under-represent decadal variability. However, studying just the amplitude (or power) of internal variability in Fig.(4) does not give us information about all the possible impacts that a model imperfection might have on the detection and attribution statistics. Thus, there is a need to develop techniques that provide a way to evaluate the impact of specific modes of variability and their interactions, and not just their amplitude, on the detection and attribution of climate change. Many interesting studies have been developed recently (eg. DelSole et al. (2011)) but more work is needed. One

advantage of our method is that it does not require to depict modes of internal variability accurately, but instead, we can test different assumptions and hypothesis about the internal variability structure by assuming that it can be represented by different physically plausible stochastic models. The generalization of this approach taking into account spatial patterns of variability is work in progress.

4. Conclusions

The IPCC very likely statement that anthropogenic emissions are affecting the climate system is based on the statistical detection and attribution methodology, which in turn is strongly dependent on the characterization of internal climate variability as simulated by GCMs.

The understanding of the internal climate variability has been identified as one of the hardest geophysical problems of the 21st century (e.g., Ghil 2001). One of the barriers we face to advance our understanding is the lack of long enough reliable observational records. We are then left with the problem of having to characterize internal natural variability with a relatively short observational record that is in fact contaminated by natural and anthropogenic forcings. The alternative to that is to use control simulations of GCMs, with the limitations in this case imposed by the fact that many aspects of the internal variability are poorly represented by the climate models (Newman et al. 2009; DelSole and Shukla 2010; Klein et al. 1999). The way in which these inaccuracies might bias the statistical significance of the detection and attribution results is hard to identify.

In this paper, we test the robustness of the detection and attribution statements in the case of global mean surface air temperature, under different representations of such variability. We use two different physically plausible stochastic models to represent the internal climate variability, and investigate the impact of these choices on the significance of the scaling factors in the detection and attribution approach. The two simple stochastic

models are chosen to span a wide range of plausible temporal autocorrelation structures, and include the short-memory first-order autoregressive (AR(1)) process and the long-memory fractionally differencing (FD) process). We find that, independently of the representation chosen, the greenhouse gas signal remains statistically significant under the detection model employed in this paper. Thus, our results support the robustness of the IPCC detection and attribution statement for global mean temperature change.

Our results also emphasize the need to apply a wider variety of test to assess the robustness of detection and attribution statistics. Previous studies carried out a "residual consistency test" which is used to assess the GCMs simulated variability on the scales that are retained in the analysis (Allen and Tett 1999) , and tests involving doubling the amplitude of the simulated variability (Tett et al. 1999). However, in the past variations in the correlation (and hence the memory) of the data have not been taking into account in the sensitivity tests. In the context of our study, the "residual consistency test" mentioned above is equivalent to exploring the sensitivity of the detection of the greenhouse gas signal to variations in the amplitude of the white noise (i.e., shifts on the horizontal direction only in both the upper and lower panels of Fig.(2)). We see that for the AR(1) process, this "consistency test" is not very helpful and a more appropriate robustness test should include constraining the values of the correlation parameter. For a FD process, however, the "consistency test" adequately explores the robustness of the results as varying the amplitude of the white noise can change their significance.

We conclude by emphasizing that in this study, headline attribution conclusions for GHG and total anthropogenic forcings were found to be insensitive to the choice between two representations of internal variability that were deliberately chosen to span a broad range of behaviors. Nevertheless, we did find that the significance of detection results were affected by the choice of a short-memory versus a long-memory process, indicating a need for checks on not only the variance but also the autocorrelation properties of internal variability when detection and attribution methods are applied to other variables and regional indices.

Acknowledgments.

This work was based on work supported in part by Award No. KUK-C1-013-04, made by King Abdulah University of Science and Technology (KAUST). M.R.A. acknowledges financial support from NOAA/DoE International Detection and Attribution Group (IDAG). A.L. was funded by the ESRC Centre for Climate Change Economics and Policy, funded by the Economic and Social Research Council and Munich Re. The authors also thank Dr. Ingram for and one of the referees for valuable discussions.

APPENDIX A

We use a HadCM3 control simulation of 1000 years to assess how the uncertainty of the stochastic parameters depends on the length of the segment, and we refer to this as a finite size effect. We estimate the stochastic parameters from the same control simulation but increasing its length by 20 years in each step , starting from 99 years. We do this for both stochastic models considered in this paper, the AR1 and FD model. In Fig.(5) and Fig.(6) the horizontal axis shows the length of the segment and the vertical axis shows the estimated parameter. Fig.(5) shows the result for the AR(1) model, in this case the estimated a_1 oscillates around a fixed value from the beginning. Fig.(6) shows the results for the FD model, in this case δ decreases its value until the segment is reaching a length of 300 years. Given that the observed record is of 156 years and that the best estimate of the white noise amplitude is larger than what we found for the long HadCM3 control run, we can expect an overestimation of the δ parameter in the observed 156 years. Thus, we expect that the 10% probability of δ being such that $\delta > 0.5$ for the Monte Carlo estimation of uncertainty, would decrease if we had a longer record.

We also investigated the correlation between the memory parameters and the white noise parameters when fitting and AR(1) and FD stochastic model to the CMIP3 control segments. The upper panel of Fig.(7) shows a very high correlation between a_1 and δ, each color represents a different GCM. The lower panel in Fig.(7) shows a very high correlation between a_0^2 and σ_e.

REFERENCES

Allen, M., D. J. Frame, C. Huntigford, C. D. Jones, J. Lowe, M. Meinshausen, and N. Meinhausend, 2009: Warming caused by cumulative carbon emissions towards the trillionth tonne. *Nature*, **458**, 1163–1166.

Allen, M. and P. Stott, 2003: Estimating signal amplitudes in optimal fingerprinting, part I: Theory. *Climate Dynamics*, **21 (5)**, 477–491.

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

Andrews, D. G. and M. R. Allen, 2008: Diagnosis of climate models in terms of transient climate response and feedback response time. *Atmospheric Science Letters*, **9 (1)**, 7–12.

Beran, J., 1994: *Statistics for long-memory processes*, Vol. 61. Chapman & Hall, New York.

Bloomfield, P., 1992: Trends in global temperature. *Climatic Change*, **21 (1)**, 1–16.

Bretherton, C. and D. Battisti, 2000: Interpretation of the results from atmospheric general circulation models forced by the time history of the observed sea surface temperature distribution. *Geophysical research letters*, **27 (6)**, 767–770.

Brohan, P. et al., 2006: Uncertainty estimates in regional and global observed temperature changes: A new dataset from 1850. *J. Geophys. research*, **111**, D12 106.

DelSole, T. and J. Shukla, 2010: Model fidelity versus skill in seasonal forecasting. *Journal of Climate*, **23 (18)**, 4794–4806.

DelSole, T., M. Tippett, and J. Shukla, 2011: A significant component of unforced multidecadal variability in the recent acceleration of global warming. *Journal of Climate*, **24 (3)**, 909–926.

584 Erland, S., P. Greenwood, and L. Ward, 2011: $1/f^{\alpha}$ noise is equivalent to an eigenstructure
585 power relation. *Europhysics Letters, Volume*, **95 (6)**, 60 006.

586 Esper, J., et al., 2012: Orbital forcing of tree-ring data. *Nature Climate Change*.

587 Ghil, M., 2001: Hilbert problems for the geosciences in the 21st century. *Nonlinear Processes*
588 *in Geophysics*, **8 (4/5)**, 211–211.

589 Ghil, M., 2012: Climate variability: Nonlinear and random effects. *Encyclopedia of Atmo-*
590 *spheric Sciences*.

591 Ghil, M., et al., 2002: Advanced spectral methods for climatic time series. *Rev. Geophys*,
592 **40 (1)**, 1003.

593 Gil-Alana, L., 2005: Statistical modeling of the temperatures in the northern hemisphere
594 using fractional integration techniques. *Journal of climate*, **18 (24)**, 5357–5369.

595 Granger, C., 1980: Long memory relationships and the aggregation of dynamic models.
596 *Journal of Econometrics*, **14 (2)**, 227–238.

597 Hasselmann, K., 1979: On the signal-to-noise problem in atmospheric response studies.
598 *Meteorology of tropical oceans*, 251–259.

599 Hasselmann, K., 1993: Optimal fingerprints for the detection of time-dependent climate
600 change. *Journal of Climate*, **6 (10)**, 1957–1971.

601 Hays, J., J. Imbrie, N. Shackleton, et al., 1976: Variations in the earth's orbit: pacemaker
602 of the ice ages. American Association for the Advancement of Science.

603 Hegerl, G. C., H. A. N. von Storch, K. Hasselmann, B. Santer, U. Cusbasch, and P. Jones,
604 1996: Detecting greenhouse-gas-induced climate change with an optimal fingerprint
605 method. *Journal of Climate*, **9 (10)**, 2281–2306.

606 Huntingford, C., P. Stott, M. Allen, and F. Lambert, 2006: Incorporating model uncertainty
607 into attribution of observed temperature change. *Geophysical Research Letters*, **33 (5)**,
608 L05 710.

609 Hurst, H., 1951: Long-term storage capacity of reservoirs. *Trans. Amer. Soc. Civil Eng.*,
610 **116**, 770–808.

611 Hurst, H., 1957: A suggested statistical model of some time series which occur in nature.
612 *Nature*, **180**, 494.

613 Huybers, P. and W. Curry, 2006: Links between annual, Milankovitch and continuum tem-
614 perature variability. *Nature*, **441 (7091)**, 329–332.

615 Imbers, J., A. Lopez, C. Huntingford, and M. Allen, 2013: Testing the robustness of the an-
616 thropogenic climate change detection statements using different empirical models. *Journal*
617 *of Geophysical Research: Atmospheres*, 1–8.

618 Kaulakys, B., J. Ruseckas, V. Gontis, and M. Alaburda, 2006: Nonlinear stochastic models
619 of 1/f noise and power-law distributions. *Physica A: Statistical Mechanics and its Appli-*
620 *cations*, **365 (1)**, 217–221.

621 Klein, S., B. Soden, and N. Lau, 1999: Remote sea surface temperature variations during
622 enso: Evidence for a tropical atmospheric bridge. *Journal of Climate*, **12 (4)**, 917–932.

623 Kmenta, J., 1971: *Elements of econometrics*. Macmillan New York.

624 Levitus, S., J. Antonov, and T. Boyer, 2005: Warming of the world ocean, 1955-2003.
625 *Geophys. Res. Lett*, **32**, L02 604.

626 Meinshausen, M., et al., 2011: The rcp greenhouse gas concentrations and their extensions
627 from 1765 to 2300. *Climatic Change*, **109 (1-2)**, 213–241.

628 Mitchell, J. et al., 1976: An overview of climatic variability and its causal mechanisms.
629 *Quaternary Research*, **6 (4)**, 481–493.

Morice, C. P., J. J. Kennedy, N. A. Rayner, and P. D. Jones, 2012: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The hadcrut4 data set. *Journal of Geophysical Research: Atmospheres (1984–2012)*, **117 (D8)**.

Newman, M., P. Sardeshmukh, and C. Penland, 2009: How important is air-sea coupling in enso and mjo evolution? *Journal of Climate*, **22 (11)**, 2958–2977.

Pelletier, J., 1997: Analysis and modeling of the natural variability of climate. *Journal of climate*, **10 (6)**, 1331–1342.

Percival, D., J. Overland, and H. Mofjeld, 2001: Interpretation of north pacific variability as a short-and long-memory process. *Journal of climate*, **14 (24)**, 4545–4559.

Solomon, S., 2007: *Climate change 2007: the physical science basis: contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge Univ Pr.

Stone, D., M. Allen, and P. Stott, 2007: A multimodel update on the detection and attribution of global surface warming. *Journal of climate*, **20 (3)**, 517–530.

Stott, P. et al., 2006: Observational constraints on past attributable warming and predictions of future global warming. *Journal of climate*, **19**, 3055–3069.

Tett, S. F., P. A. Stott, M. R. Allen, W. J. Ingram, and J. F. Mitchell, 1999: Causes of twentieth-century temperature change near the earth's surface. *Nature*, **399 (6736)**, 569–572.

Vyushin, D. and P. Kushner, 2009: Power-law and long-memory characteristics of the atmospheric general circulation. *Journal of Climate*, **22 (11)**, 2890–2904.

Vyushin, D. and P. P.J. Kushner, 2012: Modelling and understanding persistence of natural climate variability, submitted.

Wunsch, C., 2003: The spectral description of climate change including the 100 ky energy. *Climate Dynamics*, **20 (4)**, 353–363.

Zhou, J. and K.-K. Tung, 2013: Deducing multidecadal anthropogenic global warming trends using multiple regression analysis. *Journal of the Atmospheric Sciences*, **70 (1)**, 3–8.

List of Tables

	AR(1) 1850-2005	AR(1) 1850-2012	FD 1850-2005	FD 1850-2012
VOL	0.46	0.48	0.51	0.53
SOL	2.26	2.03	1.14	0.99
GHG	0.94	0.71	0.91	0.66
SUL	2.47	1.44	2.04	0.93
VOL	0.54	0.52	0.55	0.53
SOL	0.98	1.24	0.58	0.83
ANT	0.76	0.71	0.81	0.73

TABLE 1. Scaling factors β obtained from the linear regression when using HADCRUT4 observations for two time periods (1850 to 2005 and to 2012), and the forced temperature responses to VOL,SOL,GHG and SUL forcings , or to VOL, SOL and ANT forcings.

CCMA-CGCM3
CCCMA-CGCM3-1-T63
CNRM-CM3
CSIRO-MK3-0
GFDL-CM2-0
GFDL-CM2-1
GISS-AOM
GISS-AOM
GISS-Model-E-H
GISS-Model-E-R
IAP-FGOALS1-0-G
IAP-FGOALS1-0-G
IAP-FGOALS1-0-G
INMCM3-0
IPSL-CM4
MIROC3-2-HiRes
MIUB-ECHO-G
MPI-ECHAM5
MRI-CGCM2-3
NCAR-CCSM3
NCAR-PCM1
UKMO-HadCM3

TABLE 2. CMIP-3 General circulation models used partly on the 4th IPCC Assessment report. The order on the table is the same as the numbering in previous figures.

List of Figures

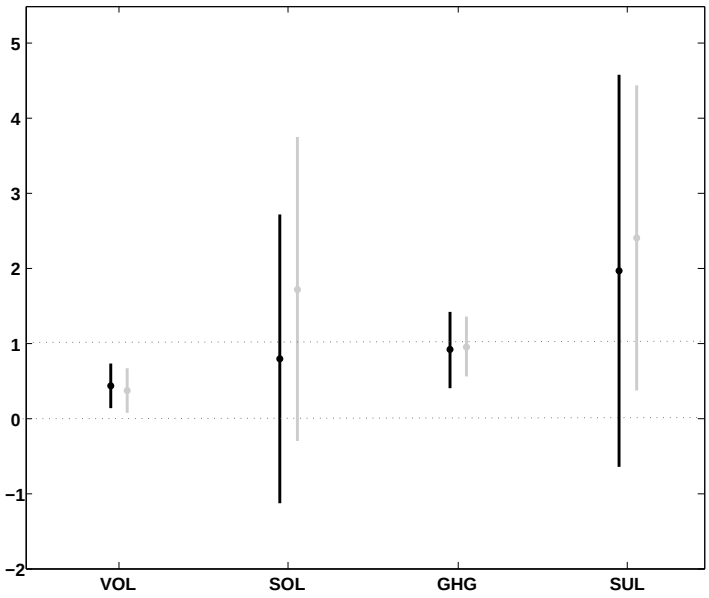

FIG. 1. The 95% confidence intervals of the scaling factors β_i derived from the multiregression of observed temperature changes onto the BDM estimates of the forced responses. The internal variability is represented by an AR(1) model (grey line) or an FD model (black line) for the period $1850 - 2005$

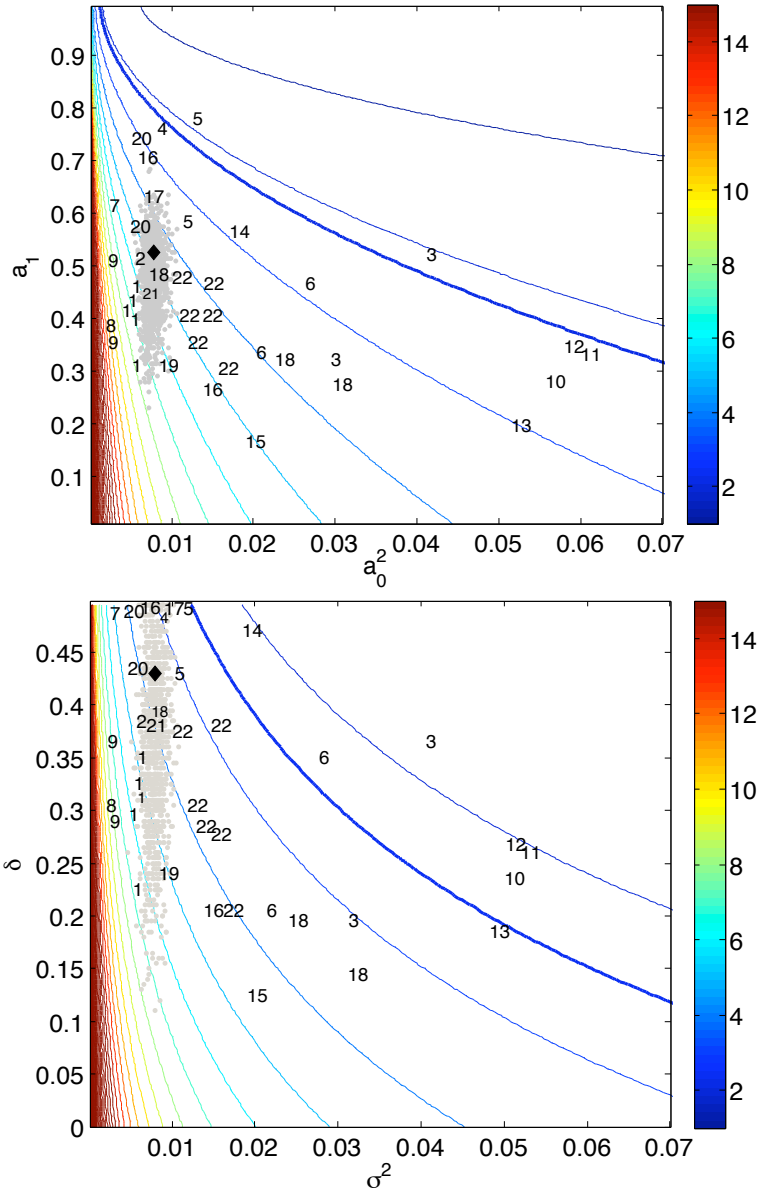

FIG. 2. Upper panel: Significance of the greenhouse gas signal as a function of the two stochastic parameters of the AR(1) model: a_1 in the y-axis and a_0^2 in the x-axis. The best fit of the observed record is displayed (diamond), showing that the significance is much greater than 2.22 (thick blue contour line) even considering the uncertainty of the Monte Carlo experiment (cloud of grey points). Best fits of the CMIP3 control segments of the same length as the observed record are shown with numbers, where each number represents a GCM (1-22). A total of 33 non-overlapping segments were selected. Lower panel: Significance of the greenhouse gas signal as a function of the two stochastic parameters of the FD model: δ in the y-axis and σ^2 in the x-axis. The best fit of the observed record is displayed (diamond), showing its significance is greater than 2.45 (thick blue contour line), even considering the uncertainty of the Monte Carlo experiment (cloud of grey points).

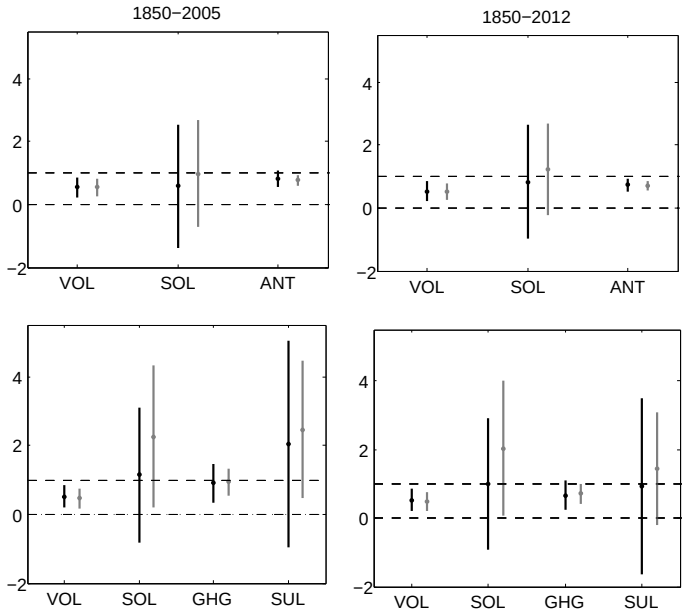

FIG. 3. The 95% confidence intervals of the scaling factors β_i derived from the multiregression of observed temperature changes onto the BDM estimates of the forced responses to the three signals VOL, SOL and ANT (top panels) and VOL, SOL, GHG and SUL (bottom panels). The internal variability is represented by an AR(1) model (grey line) or an FD model (black line) for the period $1850 - 2005$ (left hand side) and the period $1850 - 2012$ (right hand side), using HadCRUT4

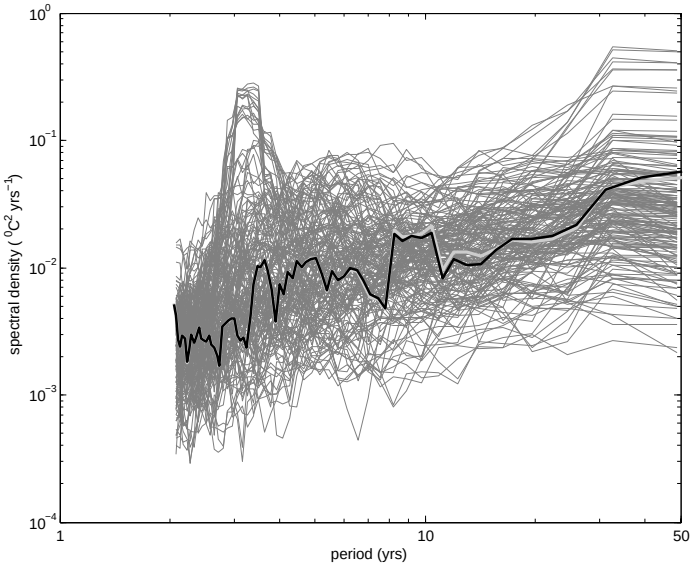

FIG. 4. Spectra from the individual GCM control simulations (gray), and the spectra of the residuals of the linear fit to T_{obs}: $T_{obs} - \hat{\beta}T$, when the internal variability is modeled as an AR(1) (thick grey line) and an FD (black line) process. We use a logarithmic scale in the horizontal axis (period) and the vertical axis (spectral density).

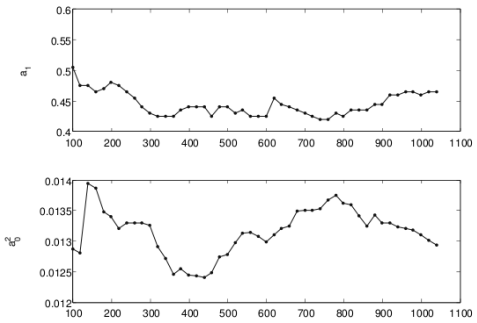

FIG. 5. AR(1) results of estimating a_1 (upper panel) and a_0^2 (lower panel) as a function of the length of the control segment sampled from the 1000 years long HadCM3 control run.

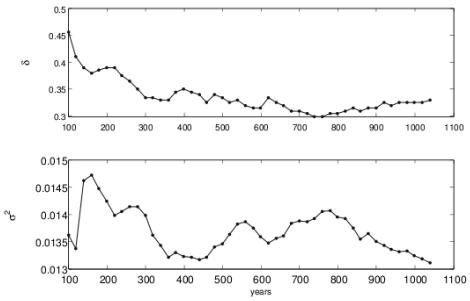

FIG. 6. FD results of estimating δ (upper panel) and σ_e (lower panel) as a function of the length of the control segment sampled from the 1000 years long HadCM3 control run.

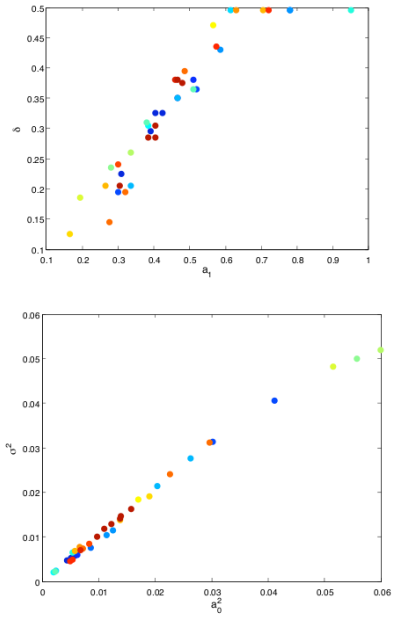

FIG. 7. Upper panel: correlation between the memory parameter of both stochastic models, δ values (vertical axis) versus a_1 values (horizontal axis) obtained from the CMIP3 control segments considered in our analysis. Lower panel: same for the white noise parameter of both stochastic models, σ_e (vertical axis) versus a_0^2 (horizontal axis). Each color corresponds to a different GCM.