# A new statistical approach to climate change detection and attribution

Aurélien Ribes, Francis W. Zwiers, Jean-Marc Azaïs, Philippe Naveau

# A new statistical approach to climate change detection and attribution

**Aurélien Ribes[1], Francis W. Zwiers[2], Jean-Marc Azaïs[3], Philippe Naveau[4]**

**Abstract** We propose here a new statistical approach to climate change detection and attribution that is based on additive decomposition and simple hypothesis testing. Most current statistical methods for detection and attribution rely on linear regression models where the observations are regressed onto expected response patterns to different external forcings. These methods do not use physical information provided by climate models regarding the expected response magnitudes to constrain the estimated responses to the forcings. Climate modelling uncertainty is difficult to take into account with regression based methods and is almost never treated explicitly. As an alternative to this approach, our statistical model is only based on the additivity assumption; the proposed method does not regress observations onto expected response patterns. We introduce estimation and testing procedures based on likelihood maximization, and show that climate modelling uncertainty can easily be accounted for. Some discussion is provided on how to practically estimate the climate modelling uncertainty based on an ensemble of opportunity. Our approach is based on the "*models are statistically indistinguishable from the truth*" paradigm, where the difference between any given model and the truth has the same distribution as the difference between any pair of models, but other choices might also be considered. The properties of this approach are illustrated and discussed based on synthetic data. Lastly, the method is applied to the linear trend in global mean temperature over the period 1951–2010. Consistent with the last IPCC assessment report, we find that most of the observed warming over this period (+0.65 K) is attributable to anthropogenic forcings ($+0.67 \pm 0.12$ K, 90 % confidence range), with a very limited contribution from natural forcings ($-0.01 \pm 0.02$ K).

**Keywords** Detection · Attribution · Climate change · Optimal fingerprint

✉ Aurélien Ribes
aurelien.ribes@meteo.fr

[1] CNRM, Météo France/CNRS, 42 avenue Gaspard Coriolis, 31057 Toulouse, France

[2] Pacific Climate Impacts Consortium, University of Victoria, Victoria, BC V8W 2Y2, Canada

[3] IMT, University of Toulouse, 118 route de Narbonne, 31062 Toulouse Cedex 9, France

[4] Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRSUVSQ, Université Paris-Saclay, 91191 Gif-sur-Yvette, France

## 1 Introduction

Detection and attribution of climate change has received much attention over the last two decades because it seeks to assess whether recent observed changes are consistent with internal climate variability only, or are consistent with the expected response to different combinations of external forcings and internal variability (Santer et al. 1995; Mitchell et al. 2001; Hegerl et al. 2007; Bindoff et al. 2013). The IPCC's definitions of these notions have partly varied over time, particularly in order to be more suitable to all IPCC working groups. The definitions used in the 5th assessment report (IPCC 2013, thereafter AR5) were taken from the IPCC guidance paper on detection and attribution by Hegerl et al. (2010), and were stated as follows. "Detection of change is defined as the process of demonstrating that climate [...] has changed in some defined statistical sense, without providing a reason for that change. [...] Attribution is defined as the process of evaluating the relative

contributions of multiple causal factors to a change or event with an assignment of statistical confidence". The definition for attribution that was previously used by WGI in the AR4 (IPCC 2007, thereafter AR4), was a bit more precise and mentioned three required conditions. A change Y was attributed to a forcing or a combination of forcings X if Y was detectable, consistent with the expected response to the cause X, and inconsistent with alternative, physically plausible causes. Note that "consistent with" may be understood in different ways that will be discussed further later. Taken together, these two definitions imply that detection and attribution (D&A) requires some knowledge of the statistical properties of the internal (unforced) climate variability and of the expected response to one or several external forcings. They also imply that D&A relies on statistical inference approaches.

Several statistical models have been used in D&A. Over the past two decades, the most commonly used method has relied on linear regression models where observations are regressed onto expected responses to external forcings, with various levels of complexity. Regression based models for D&A were first proposed by Hasselmann (1979), Hasselmann (1993), who introduced the "Optimal Fingerprints" terminology. The links with classical linear regression were further highlighted by Allen and Tett (1999a). More recently, this statistical model was made somewhat more complex so as to account for uncertainty in the estimates of the expected responses due to internal variability (there is uncertainty from this source because the expected responses are estimated from finite ensembles of forced simulations). This is referred to as the Total Least Squares approach (TLS, see Allen and Stott 2003; Ribes et al. 2013).

In all of these studies, a key assumption underlying the use of such a regression approach is that uncertainty in the simulated response to a specified external forcing is dominated by uncertainty in the amplitude of the spatio-temporal pattern of response rather than in the pattern itself. The justification for this assumption is generally that the magnitude of the response is often uncertain while the spatio-temporal pattern of response is more constrained by physical understanding. As an illustration, uncertainties in the amplitude of the response to increasing concentrations of well-mixed greenhouse gases (GHGs) has remained relatively large for more than three decades: about a factor of 2.5 for the transient climate response, and 3 for the equilibrium climate sensitivity (Knutti and Hegerl 2008; IPCC 2013), related to large uncertainty on feedbacks. In contrast, the response to GHGs is robustly expected to be larger over land than over oceans, and to be amplified in the Arctic region. With a very high level of confidence, it is also expected to strengthen with time, as a direct consequence of the historical evolution of emissions and concentrations. Robust, large scale features of response patterns have also been described for other variables—e.g. the temperature of the free atmosphere and some aspects of the precipitation response—and for other forcings, such as natural forcings (e.g. due to the well-known historical occurrence of major volcanic eruptions) and anthropogenic aerosols (as a direct consequence of the location and timing of their emissions). These arguments may justify the assumption that the amplitude of the response is a dominant source of uncertainty, which is subsequently treated as unknown in statistical regression models.

Despite these uncertainties, physical knowledge does provide some information about amplitude of the response to many forcings. As a very naive illustration, the sign of the response, e.g. in terms of the mean surface temperature, is known in many cases. Although uncertainty remains substantial, physical constraints often allow us to discard wide ranges of values for the magnitude of the response to a forcing. Regression based statistical models, by considering the amplitude of the response as being unknown, do not take these constraints into account, although the simulated response magnitudes are subsequently used to interpret the fitted model. This weakness was previously pointed out by Berliner et al. (2000), who argued for a Bayesian treatment. Huber and Knutti (2012) also proposed to incorporate prior knowledge on the magnitude to disentangle contributions from various external forcings. Furthermore, while a few features of the response pattern are well-known qualitatively, uncertainties may stand in the way of their quantification. The land-sea warming ratio, for instance, or the amount of Arctic amplification, vary from one model to another. Many other aspects of the spatial patterns of response also vary considerably among models (e.g. Shin and Sardeshmukh 2011).

The discussion about the relative importance of the uncertainties on the magnitude vs pattern can also be considered from a forcings and feedbacks perspective.

- Some forcings are quite uncertain, both in magnitude and pattern. For example, aerosols forcing is particularly uncertain (Boucher et al. 2013). Greenhouse gases forcing also has substantial uncertainties if effective radiative forcings are considered rather than radiative forcings (Myhre et al. 2013). The associated time-series have also been shown to be quite uncertain, particularly for aerosols (e.g. Rotstayn et al. 2015), suggesting that it may not be sufficient to represent forcing uncertainty through the use of a scaling coefficient.
- In addition to uncertainty on the forcing itself, global scale feedbacks like the water vapour feedback are likely to enhance or reduce the signal everywhere, consistent with a regression framework. In contrast, the cloud feedback, which is the most uncertain feedback in

the response to increasing GHGs (Dufresne and Bony 2008), is highly variable over space, and able to induce local and regional changes in the atmospheric circulation (Stevens and Bony 2013), that could contribute substantially to uncertainties on the response patterns.

- Further, possible unknown feedbacks within the climate system have been mentioned as factors that contribute uncertainty to the response magnitude. However, if we acknowledge that unknown feedbacks may impact the magnitude of the response as simulated by climate models, then it seems reasonable to expect that these feedbacks may also alter the expected patterns of response as they will probably act differently over different regions—e.g. feedbacks involving specifically the land surface, snow, sea ice, etc.

- Thus, even if the global rescaling of the model estimated patterns of forced responses has some advantages, we argue that uncertainties in the response patterns are also very substantial and that they should be treated symmetrically to the extent possible.

In terms of D&A, Ribes and Terray (2013) recently reported that inter-model discrepancies in the simulated response patterns were substantial, and potentially large enough to be detrimental to D&A results (see also Jones et al. 2013). This study showed in particular that, as a consequence of the discrepancies in response patterns simulated by different models, the use of regression based methods may lead to nonphysical D&A results such as negative scaling factors - even in cases where the sign of the response is not ambiguous. It therefore suggested that better accounting for physical knowledge could be of interest in D&A, and that climate modelling uncertainty should not be neglected. Note that "climate modelling uncertainty" here includes uncertainties in climate model parameters, and in the representation of physical processes in models, but does not include sampling uncertainty related to internal variability within climate model experiments.

A few approaches have previously been proposed to account for climate modelling uncertainty within a regression framework, using Errors-In-Variables approaches (Huntingford et al. 2006). However, statistical inference is then much more complicated (e.g. maximum likelihood estimates are not explicit), and there are remaining issues in the uncertainty analysis (Hannart et al. 2014). More importantly, if climate modelling uncertainty is explicitly taken into account in the spatio-temporal response pattern, we see no clear reason why the response magnitude should be treated differently. We argue that both the magnitude and the pattern could be treated similarly in order to reflect inter-model uncertainty. In this way, all the information provided by physical models on the response to each forcing could be appropriately taken into account, together with the associated modelling uncertainty. As further illustrated in this work, the climate modelling uncertainty in the estimated response to a given forcing may well be much larger in the magnitude than in the pattern—e.g. it may cover a range of values as large as the factor of 2.5 on the transient sensitivity to GHG forcing that is acknowledged in the IPCC AR5. In such cases, the method we propose will come close to the usual linear regression methods.

The main goal of this study is to describe an alternative to the use of linear regression based statistical models in D&A. This alternative basically proposes a symmetric treatment of the magnitude and the pattern of the response to each forcing. Our method involves simple hypothesis testing to check each of the three conditions mentioned in the IPCC AR4 attribution definition. In particular, we address the important questions such as the consistency between observed and simulated responses, not only in terms of the response magnitudes, but also in terms of the response patterns. We illustrate a few properties of this method and show its efficiency in particular in cases where linear regression is not efficient, e.g. where there are collinear or weak responses.

A second objective of this work is to provide a new statistical framework to deal with climate modelling uncertainty in D&A. This is done by considering the responses simulated by a wide range of climate models, and the corresponding uncertainty, as representing what we know about the physically plausible response to a forcing. Given modelling uncertainty, observations are used to further constrain the response to each forcing. In this way, our method also allows assessment of the contributions of different combinations of external forcings, consistent with the more recent definition of attribution (Hegerl et al. 2010). It is shown that accounting for climate modelling uncertainty in this way leads to a simpler and more accurate statistical treatment than under the EIV approach. Note that if modelling uncertainty is ignored, the method presented in this paper also leads to a simpler and more accurate treatment than under the widely used TLS approach.

An important requirement to deal with climate modelling uncertainty is to be able to estimate it from available ensembles of climate models. This requirement applies whatever the statistical method used. The inference method presented here assumes that such estimates are available. Since this is a challenging task, we provide a brief discussion of how such estimates might be derived. Our estimation is based on the "*models are statistically indistinguishable from the truth*" paradigm, where the difference between any given model and the truth has the same distribution as the difference between any pair of models. However, many different methods might be considered, and their efficiency

might depend on the variable under scrutiny, so we provide no definitive recommendation in this respect. In particular, the chosen paradigm may not be appropriate for cases where the models have large common errors and are unable to simulate the pattern or magnitude of a change.

In Sect. 2, we present how an attribution statement may be made based on a single univariate variable if information on the magnitude of the response is considered. Section 3 describes the general statistical framework as well as the proposed inference and hypothesis testing methods. Section 4 deals with a few implementation issues, in particular those related to the estimation of the variance-covariance matrix of climate modelling uncertainty. Section 5 compares our method with traditional linear regression methods, and provides a few illustrations of the method using synthetic data, in order to describe important properties. Lastly, Sect. 6 shows a first application to real-world data by focusing on the linear trend in global mean temperature over the 1951-2010 period.

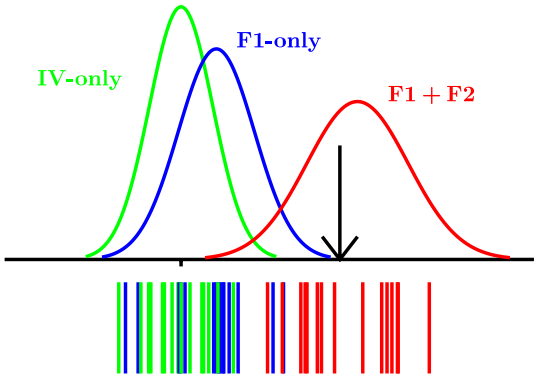


**Fig. 1** Schematic illustration of univariate D&A The observed scalar diagnostic (*black arrow*) is compared to internal variability (*green distribution*), the expected response to forcing $F_1$ only (*blue distribution*), or the expected response to forcings $F_1 + F_2$ taken together (*red distribution*). Samples obtained under each of these three distributions are illustrated below the $x$-axis (synthetic data, with a sample size of 15)

## 2 D&A based on a single observation of a scalar diagnostic

This section is intended to provide a short illustration of how detection and attribution may be applied to a scalar quantity such as the linear trend in the global mean temperature. Considering one unique variable, response patterns are not really defined (except for the sign of the response), and D&A will only focus on the magnitude of the (observed and simulated) changes. We stress that if more than one causal factor were to be considered, a regression model could not have been used as several regression coefficients cannot be inferred based on a single scalar observation. The discrimination between two forcings, if any, necessarily relies on the magnitude of the observed change. We also argue that such a very simple analysis could be very relevant to determining how much an observation exceeds the range of internal variability, or to assess whether this observation is consistent with an ensemble of simulated responses. It may also be particularly appropriate for ruling out a weak forcing from being a sufficient explanation of an observed change. An illustration of such a single scalar based D&A analysis applied to real-world data is provided in Sect. 6, based in the linear trend in global mean temperature over the 1951-2010 period.

The question of detection, first, is primarily related to assessing whether the observed indicator of climate change, say $y$, is consistent with internal variability only. To address this question, D&A studies usually compare observations with internal variability as simulated from unforced climate models runs (Hegerl and Zwiers 2011). In regression formulations, the regression residuals are interpreted as an observe realization of internal variability, and are additionally compared with model simulated internal variability to assess the ability of the models to simulate such variability (Hegerl and Zwiers 2011). Then considering a pool of unforced control run segments covering the same period, the question comes to a simple comparison of $y$ with the same indicator computed from each simulated segment, say $a_1, \ldots, a_p$. More precisely, the detection question could be rephrased as "Does $y$ come from the same population as the $a_i$'s?" Assuming a Gaussian distribution for the $a_i$'s, which is commonplace at least for mean temperature, this question could be easily addressed with a Student $t$ test. Note that we do not provide mathematical details in this section, as the statistical framework will be presented in Sect. 3 under more general assumptions. This simple way of performing detection has been used in several studies, e.g. Santer et al. (2013), Terray et al. (2011) or even much earlier as discussed by Hegerl and North (1997).

Regarding attribution, and based on the IPCC AR4 definition (IPCC 2007), two more questions have to be considered. The first regards the consistency of $y$ with the expected response to all forcings considered. To evaluate consistency, climate models are therefore also required to provide estimates of the expected response to all forcings (Hegerl and Zwiers 2011). If we assume that a sample $b_1, \ldots, b_q$ of simulated responses is available from a pool of climate models, the question, from a statistical point of view, comes to be identical to the previous one, and could be tackled with the same simple tool, i.e. a Student test. Note that here, consistency is meant in a broad sense, i.e. including climate modelling uncertainty. The last question—consistency with alternative, physically plausible causes—demands a third sample of climate model

simulations, e.g. only driven by a subset of external forcings. Denoting this sample $c_1, \ldots, c_r$, the statistical treatment could, again, be the same.

Figure 1 provides an illustration of a hypothetical D&A outcome based on a single scalar observable. We consider only two forcings, say $F_1$ and $F_2$. The observation (here assumed to be free of measurement uncertainty) is compared to three different distributions. Attribution of the change to $F_2$ may be claimed since consistency is not found with internal variability only, found with forcings $F_1 + F_2$ combined, and not found with forcing $F_1$ only. Thus a procedure such as that which is illustrated could be argued to comply with the IPCC AR4 definition of detection and attribution: a change has been detected, it is demonstrated to be consistent with a hypothesized combination of forcings, and inconsistent with other plausible explanations. This procedure is also similar to that used in pioneering detection studies (e.g. Hasselmann 1997; Hegerl et al. 1997). The "contributions of several causal factors" as strengthened in the IPCC AR5, however, are not really quantified. This is done in the next section under a more general statistical framework that also allows multivariate diagnostics to be used, such as a space-time pattern of change.

## 3 Statistical framework and inference approach

### 3.1 Statistical model

Our proposed statistical model is written as follows:

$$Y^* = \sum_{i=1}^{n_f} X_i^*, \tag{1}$$

$$Y = Y^* + \varepsilon_Y, \quad \varepsilon_Y \sim N(0, \Sigma_Y), \tag{2}$$

$$X_i = X_i^* + \varepsilon_{X_i}, \quad \varepsilon_{X_i} \sim N(0, \Sigma_{X_i}), \quad i = 1, \ldots, n_f, \tag{3}$$

where $Y^*$ is the true, unknown response of the climate system to all external forcings taken together, $X_i^*$ is the true, unknown response to each forcing $i$ among the $n_f$ forcings considered, $Y$ is the observation, $\varepsilon_Y$ describes noise in observations, $X_i$ is the simulated response to forcing $i$, and lastly, $\varepsilon_{X_i}$ is the deviation between the simulated and true response to forcing $i$. Note that all these variables are vectors of the same dimension $n$. Multiple factors could contribute to $\varepsilon_Y$, including observational error and internal variability. Multiple factors could also contribute to $\varepsilon_{X_i}$, including climate model uncertainty, forcing uncertainty and internal variability. The random variables $\varepsilon_{X_1}, \ldots, \varepsilon_{X_{n_f}}$ and $\varepsilon_Y$ are all assumed to be independent and to follow Gaussian distributions with zeros means and known variance-covariance matrices (respectively $\Sigma_{X_1}, \ldots, \Sigma_{X_{n_f}}$, and $\Sigma_Y$, hereafter variance matrices). These strong assumptions are further discussed below.

Equation (1) assumes additivity, i.e. that the response to a subset of forcings is the sum of the responses to each forcing taken individually. This strong assumption has been used in most previous D&A methods (see also Shiogama et al. 2013). Equation (2) describes how the external response is altered by noise in the observations $Y$. Noise refers here to both internal variability and measurements errors, possibly including instrumental errors, errors related to observation adjustments and sampling errors associated with the configuration of the observing network and its evolution over time (Brohan et al. 2006; Morice et al. 2012). References to observational uncertainty in this article refer to $\varepsilon_Y$. Previous studies have suggested that internal variability tends to be the dominant source of observational uncertainty, at least for global near-surface temperature (Jones and Stott 2011). Equation (3) provides a symmetrical representation for climate model output. $X_i$ is typically the mean response to forcing $i$ simulated by a set of climate models. The multimodel mean will be mainly considered, but we refer to Sect. 4 for a comprehensive discussion on deriving a response $X$ from a multimodel ensemble. The noise $\varepsilon_{X_i}$ that contaminates $X_i$ may be related to both internal variability within the climate model runs and other sources of variation such as forcing and modelling uncertainty, with the latter often being larger than the former. Discussion on how to estimate and combine these two terms is also provided in Sect. 4.

Our proposed statistical model (1)–(3) does not rely on a linear regression approach, but it is closely related to Errors-In-Variables (EIV) models, such as described in Fuller (1987), Hannart et al. (2014). Using our notation, an EIV model would leave (2)–(3) unchanged, while (1) would be replaced with

$$Y^* = \sum_{i=1}^{n_f} \beta_i X_i^*, \tag{4}$$

where the $\beta_i$'s are unknown scaling factors. Therefore, these two statistical models differ only by the introduction of unknown regression coefficients (or scaling factors) $\beta_i$ in (4). Similarly, the TLS statistical framework, which has been much more widely used than EIV, consists of equations (2)–(4), with only small differences in the definition of $\varepsilon_Y$ and $\varepsilon_X$. In linear regression, and more generally under the TLS and EIV models, using these coefficients $\beta_i$ means that the magnitude of the influence of $X$ on $Y$ is assumed unknown. Usually, such adjustments are useful as $X$ and $Y$ may not be directly comparable, e.g. may be given with different units, etc. This is obviously not the case in D&A. In particular, the most commonly used approach aims at assessing, in addition to scaling factors, the consistency between model output and

observations by testing whether $\beta$ is consistent with 1. This means that quantities $X$ and $Y$ are considered to be directly comparable. By removing these scaling factors $\beta$, we assume that the cohort of available climate models can appropriately simulate the response magnitude, with some modelling uncertainty, in addition to the assumption that they are able to appropriately simulate the response patterns, again with some modelling uncertainty. The latter assumption is implicit under the EIV approach. It is also implicit under the TLS approach, but the source of response pattern uncertainty is then limited to internal variability only. Our model thus proposes a more symmetric treatment of uncertainty in the magnitude and the pattern of the change. Note that the climate modelling uncertainty will be estimated from an ensemble of opportunity (e.g. Taylor et al. 2012).

A few important remarks should be made concerning this statistical model.

First, this model may be regarded as a linear Gaussian model. This alternative point of view is discussed in Appendix 8.1, and helps us derive some statistical properties. However, we will still focus on the previous point of view, i.e. (1)–(3), in the following, as the maximization of the likelihood is simpler in this way.

Second, the assumption that the variance matrices $\Sigma_Y, \Sigma_{X_1}, \ldots, \Sigma_{X_{n_f}}$ are known is strong. For example, in the single scalar case discussed in Sect. 2 it corresponds to using Gaussian tests instead of Student tests. This strong assumption, however, is consistent with previous attempts to account for climate modelling uncertainty in D&A, in particular Huntingford et al. (2006) and Hannart et al. (2014). Practically, these matrices need to be estimated from multi-model ensembles, as discussed in Sect. 4. Overall, this study proposes a "plug-in" method, where these matrices are estimated first, and then considered as fixed in the main statistical model. Providing a more comprehensive statistical treatment that also accounts for uncertainties in the estimation of these matrices would be very attractive, but is also very challenging and is beyond the scope of this paper. Note that efforts have been made in the standard D&A approach to deal with uncertainty on the covariance matrix related to internal variability (e.g. Allen and Tett 1999b; Ribes et al. 2013; Hannart 2015). It is more challenging here as the larger uncertainty will be, in many cases, related to climate modelling uncertainty, which can only be estimated from a very limited sample of climate models. To our knowledge, previous EIV implementations have used the same plug-in approach.

Third, as the random terms $\varepsilon$ are centered (i.e. $\mathrm{E}(\varepsilon) = 0$), (1)–(3) imply that

$$\mathrm{E}(Y - X) = 0, \tag{5}$$

where $X = \sum_{i=1}^{n_f} X_i$. That is, it is assumed that the $n_f$ forcings factors considered are sufficient for explaining the forced component of the observed change.

Fourth, all errors are assumed to follow a Gaussian distribution. Considering non-Gaussian distributions might be very attractive, but is beyond the scope of this paper, although one possibility might be to transform data to bring its distribution closer to being Gaussian.

### 3.2 Inference method

The proposed inference method is based on the method of maximum likelihood (Le Cam 1990). After writing the likelihood function for the model, we derive maximum likelihood estimates and exact confidence intervals are proposed, with no use of asymptotic theory. As all random variables are assumed to follow a Gaussian distribution, we consider the $-2$ log-likelihood function (i.e. the logarithm of the likelihood, multiplied by $-2$) and minimize it, instead of maximizing the likelihood—this makes the calculation easier.

In (1)–(3), the unknown parameters are $X_i^*, i = 1, \ldots, n_f$; assuming known variance matrices, the $-2$ log-likelihood function relative to observations of $(Y, X)$ can be written (up to an additive constant that plays no role) as

$$\ell(X_1^*, \ldots, X_{n_f}^*) = \left(Y - \sum_{i=1}^{n_f} X_i^*\right)' \Sigma_Y^{-1} \left(Y - \sum_{i=1}^{n_f} X_i^*\right) + \sum_{i=1}^{n_f} \left(X_i - X_i^*\right)' \Sigma_{X_i}^{-1} \left(X_i - X_i^*\right). \tag{6}$$

### 3.3 Maximum likelihood estimators (MLEs)

Maximum likelihood estimators $\widehat{X}_i^*$ of $X_i^*$ may be obtained by maximizing this likelihood function in $X_i^*$. Note that we will also denote $\widehat{Y}^*$ the MLE of $Y^*$, as it helps to derive analytic solutions. From (6), the first order maximization condition on $X_i^*$ gives

$$\Sigma_Y^{-1}(Y - \widehat{Y}^*) + \Sigma_{X_i}^{-1}(X_i - \widehat{X}_i^*) = 0, \tag{7i}$$

which yields

$$\widehat{X}_i^* = X_i + \Sigma_{X_i} \Sigma_Y^{-1}(Y - \widehat{Y}^*). \tag{8i}$$

We then recall that $X = \sum_{i=1}^{n_f} X_i$, define $\Sigma_X = \sum_{i=1}^{n_f} \Sigma_{X_i}$, and consider $\sum_{i=1}^{n_f} \widehat{X}_i^*$, which gives

$$\widehat{Y}^* = X + \Sigma_X \Sigma_Y^{-1}(Y - \widehat{Y}^*), \tag{9}$$

$$\left(I + \Sigma_X \Sigma_Y^{-1}\right) \widehat{Y}^* = X + \Sigma_X \Sigma_Y^{-1} Y. \tag{10}$$

From there, two equivalent expressions for the MLE of $Y^*$ can be derived

$$\widehat{Y}^* = (\Sigma_X^{-1} + \Sigma_Y^{-1})^{-1}(\Sigma_X^{-1}X + \Sigma_Y^{-1}Y), \tag{11}$$

$$= Y + \Sigma_Y(\Sigma_Y + \Sigma_X)^{-1}(X - Y). \tag{12}$$

Substituting (12) into (8i), the MLE of $X_i^*$ may also be derived as

$$\widehat{X}_i^* = X_i + \Sigma_{X_i}(\Sigma_Y + \Sigma_X)^{-1}(Y - X), \quad i = 1, \ldots, n_f. \tag{13}$$

This demonstrates that MLEs can be derived explicitly under our model as opposed to the EIV model, where MLEs can only be obtained numerically, with no guarantee that the maximum is actually reached (Hannart et al. 2014) and no possibility to derive their distribution explicitly. Note also that the maximization of this likelihood corresponds to least squares minimization. In this way, the estimators derived above can be of interest even if the Gaussian assumption is not satisfied in (2)–(3).

Lastly, it may be noted that the minimum $-2$ log-likelihood (which corresponds to the maximum likelihood) takes the value (still up to the same additive constant)

$$\ell(\widehat{X}_1^*, \ldots, \widehat{X}_{n_f}^*) = (Y - X)'(\Sigma_Y + \Sigma_X)^{-1}(Y - X). \tag{14}$$

### 3.4 Distribution of MLEs

Being linear combinations of independent Gaussian vectors $X_i, i = 1 \ldots n$, and $Y$, $\widehat{X}_i^*$ and $\widehat{Y}^*$ all follow Gaussian distributions. From (12) and (13), and noting that $\mathrm{E}(Y - X) = 0$, it is also easy to prove that they are unbiased estimates. Thus their distributions will be fully determined by their variances. The variance of $\widehat{Y}^*$ can be deduced from (11), leading to

$$Var(\widehat{Y}^*) = \left(\Sigma_X^{-1} + \Sigma_Y^{-1}\right)^{-1}. \tag{15}$$

and

$$\widehat{Y}^* \sim N\left(Y^*, \left(\Sigma_X^{-1} + \Sigma_Y^{-1}\right)^{-1}\right). \tag{16}$$

The case of $\widehat{X}_i^*$ can be treated very similarly, leading to

$$\widehat{X}_i^* \sim N\left(X_i^*, \left(\Sigma_{X_i}^{-1} + \left(\Sigma_Y + \sum_{j \neq i} \Sigma_{X_j}\right)^{-1}\right)^{-1}\right). \tag{17}$$

The above equations show that the distribution of the MLEs is explicit under our statistical model, which will allow the computation of exact confidence regions or hypothesis tests, assuming we know the variance matrices. Confidence regions for $X_i^*$ allow the quantification of uncertainties in the contribution of a particular forcing to the observed changes. In particular, if $Y$ is a time series of $n$ observations, attributable trends have been commonly used (e.g. Stott et al. 2006; Jones et al. 2013; Gillett et al. 2013; Bindoff et al. 2013) to estimate the relative contributions of several forcings to a change. Within our statistical framework, estimates and uncertainty analysis on trends attributable to a given forcing can be derived respectively from (13) and (17).

Under like-for-like assumptions, similar results on uncertainty analysis are not known under TLS or EIV models. Instead, the computation of confidence intervals on $\beta$ must be based on approximate results from asymptotic statistics, and may involve some computational challenges. Hannart et al. (2014) further suggested that the use of these asymptotic results in EIV models could lead to confidence intervals that are too permissive, which is not the case here. In addition, uncertainty analysis on attributable trends inferred from TLS or EIV models should in principle involve the computation of confidence regions for $\widehat{\beta}_i\widehat{X}_i^*$, which is challenging and usually not done.

### 3.5 Hypothesis testing

This subsection describes hypothesis tests that may be used for D&A under our statistical model. Consistent with Sect. 2, we will consider three different tests, corresponding to the three conditions required for attribution: consistency with internal variability only (i.e. detection), consistency with the expected response to all forcings, and consistency with the response to a subset of forcings. We provide a more comprehensive discussion of the tests in "Appendix 8.2". In particular, all three tests we propose are in a sense "goodness of fit" tests, as is discussed in "Appendix 8.2.1".

First, consistency with internal variability corresponds to the null hypothesis $H_0 : Y^* = 0$. Under this hypothesis, $Y \sim N(0, \Sigma_Y)$, and thus a natural test would be based on

$$Y' \Sigma_Y^{-1} Y \sim_{H_0} \chi^2(n). \tag{18}$$

However, as discussed in "Appendix 8.2", this test is not exactly the likelihood ratio test (LRT) of $H_0 : Y^* = 0$ under our model (1)–(3); in particular, the proposed responses $X_1, \ldots, X_{n_f}$ are not used. The test defined in (18) might be considered as testing $H_0 : Y^* = 0$ in (2), with (1) and (3) removed. This test only attempts to answer the question about whether there is evidence of a change in climate for an undefined reason (i.e., detection in the purely statistical sense, not detection of signals $1, \ldots, n_f$ per se). A few alternatives are discussed in Appendix 8.2.3. Note that the detection that does not involve any information about the expected climate change signal was rejected by Hasselmann (1993) as a "needle in a haystack" situation: the dimensionality might be too large and prevent efficient detection. In the test proposed here, the length of $Y$ has to be small. This is usually achieved by a pre-processing of the data. Although

preliminary to the main statistical analysis, this pre-processing often focuses the power of the test on the expected change, so the signal comes into consideration indirectly and is in effect not ignored. Such a "low dimension" condition is satisfied in the application presented in Sect. 6.

Second, consistency with the response to all forcings corresponds to testing whether the data ($X$, $Y$) are consistent with Model (1)–(3). In other words, we test the goodness of fit to our full statistical model. Denoting $X = \sum_{i=1}^{n_f} X_i$, this test may be implemented based on the minimum log-likelihood given in (14):

$$(Y - X)'(\Sigma_Y + \Sigma_X)^{-1}(Y - X) \sim_{H_0} \chi^2(n). \tag{19}$$

Here, the consistency between observed and simulated responses is addressed both in terms of the response patterns and the response magnitudes. This question was only indirectly addressed in previous linear regression approaches, as the results of two different tests —testing whether scaling factors are consistent with unity, and testing the overall goodness of fit to the regression model— had to be combined to answer it. Note that this consistency test will be more demanding if $\Sigma_X$ is smaller. Practically, this means using simulations with all forcings, if available, to limit the sampling uncertainty in $\Sigma_X$. Note also that consistency here is meant with respect to all sources of uncertainty simultaneously, including internal variability, climate modelling uncertainty and observational uncertainty.

Third, consistency with only a subset of external forcings corresponds to a null hypothesis where (1) is replaced by

$$Y^* = \sum_{i \in I} X_i^*, \tag{20}$$

where $I$ is a subset of $[\![1, n_f]\!]$. Very similarly to the previous test, this test could be based on

$$(Y - X_I)'(\Sigma_Y + \Sigma_{X_I})^{-1}(Y - X_I) \sim_{H_0} \chi^2(n), \tag{21}$$

where $X_I = \sum_{i \in I} X_i$ and $\Sigma_{X_I} = \sum_{i \in I} \Sigma_{X_i}$.

By accounting for information on the magnitude of a change, this test corrects a few deficiencies of the linear regression approach. Under linear regression, this test is usually performed in the presence of at least two forcings $F_i$, which means that two or more scaling factors $\beta_i$ are estimated simultaneously. Forcing $F_1$ is assessed to not be a sufficient explanation if, within, for example, a 2-forcing analysis, $\beta_2$ is significantly different from 0. Using this procedure, however, two forcings cannot be differentiated if their responses are collinear. Even worse, if the response to the forcing $F_1$ is very weak, its response pattern will be very uncertain, even if many $F_1$-only simulations are available. This will prevent rejecting the hypothesis that "$F_1$ alone can explain the change". This is a paradoxical situation because the smallest forcings may be the most difficult to exclude from a list of possible sufficient causes. In such a case, the main information provided by climate model simulations is actually that the magnitude of the response to $F_1$ is weak, and we argue that this information has to be taken into account.

## 4 Implementation: estimation of $X$ and $\Sigma_X$

In the method presented above, the multi-model response $X$ and the variance matrix $\Sigma_X$ describing the corresponding uncertainty are assumed to be known (as a single forcing is considered, the index $i$ is dropped in this section). However, the way a multi-model ensemble of opportunity like CMIP5 may be translated into $X$ and $\Sigma_X$ is not straightforward, and several approaches might be considered. Before providing illustrations of our new method, we briefly discuss this issue and present one possible way to estimate those quantities.

It is worth noting that a wide literature, and most notably the IPCC Assessment Reports, have addressed the question of quantifying uncertainties in terms of the Earth's climate sensitivity. However, far fewer studies have provided uncertainty assessments on patterns, i.e. not restricted to one single parameter. Quantifying the uncertainty in the response pattern is necessary to account for climate modelling uncertainty in D&A, whatever the statistical model used, whether EIV or that described in this paper. In particular, previous EIV studies provided only limited discussion on this topic (Huntingford et al. 2006; Gillett et al. 2013; Hannart et al. 2014). Note that other approaches like that by Huber and Knutti (2012) also usually neglect uncertainty on the forcing pattern or time-series.

### 4.1 Paradigms for climate modelling uncertainty

While several methods with various level of complexity could be considered to derive $X$ from a multi-model ensemble, this study focuses mainly on the multi-model mean. Empirical evidence from an increasing number of studies suggests that the multimodel mean is a better estimate than responses provided by any individual model (see e.g. Knutti et al. 2010 for a review). Also, the multi-model mean has been used extensively in the IPCC assessments (IPCC 2007, 2013), as well as in many individual studies, making it a well-described variable. Nevertheless, other estimates of $X$ might have been considered, e.g. the median, or more generally a trimmed mean (i.e. mean over a subset of climate models, in order to avoid or limit the impact of outlier models on the average). But such techniques may reduce physical consistency across space, as values at different locations may come from different models. Several authors also propose to eliminate some models prior to analysis because

of unrealistic behaviour or to seek subsets of models that exhibit adherence to an emergent constraint (e.g. Hall and Qu 2006).

The computation of $\Sigma_X$ is more debatable and subject to partly arbitrary choices. Note that we only focus on climate modelling uncertainty and ignore internal variability in this section; the influence of internal variability is discussed in Sect. 4.2. Several paradigms have been used in the literature to compare observations with simulations coming from various models, and more precisely to describe the distance from $X$ at which the observations are likely to be found (Annan and Hargreaves 2010; Knutti et al. 2010). Here we only consider the "models are *statistically indistinguishable* from the truth" paradigm. In a Bayesian perspective where the truth is treated as non-deterministic, this paradigm simply assumes that the models and the truth are taken from the same distribution. This paradigm can also be understood in a frequentist perspective, assuming dependence among models, as discussed below. Whatever the point of view, this paradigm assumes that the difference between any given model and the truth has the same distribution as the difference between any pair of models. By considering this paradigm, it is implicitly assumed that the true-world response is somewhere within the distribution of model responses, for the population of models from which the available ensemble of opportunity was drawn. Since this is an assumption about the population of models, it also implies that models not yet observed may lie outside the range of responses seen in the available ensemble of opportunity.

Under this paradigm, it is assumed that values simulated by individual models, $w_j$, are taken independently from the same distribution, i.e. $w_j \sim N(\mu, \Sigma_\mathrm{m})$, where $\mu$ denotes the mean response that would be obtained from an infinitely large ensemble, and $\Sigma_\mathrm{m}$ describes the climate modelling uncertainty. It is further assumed that $\mu$ is not equal to true value of the parameter of interest $w^*$. This may happen in particular if all the models involved share some common features and errors. For any value $w'$ simulated by a randomly selected model taken from the same population, we have $w' \sim N(\mu, \Sigma_\mathrm{m})$, and thus $(\mu - w') \sim N(0, \Sigma_\mathrm{m})$. The underlying idea of the *statistically indistinguishable* paradigm is to assume that the value $\mu$ differs from the truth $w^*$ as much as it differs from any independent realization $w'$. For this reason, we assume that $(\mu - w^*) \sim N(0, \Sigma_\mathrm{m})$. As the multi-model mean $\overline{w} = \frac{1}{n_m}\sum_{j=1}^{n_m} w_j$ satisfies $(\overline{w} - \mu) \sim N(0, \Sigma_\mathrm{m}/n_m)$ and is independent from the latter, it follows that $\overline{w} \sim N(w^*, (1 + 1/n_m)\Sigma_\mathrm{m})$, which corresponds to using $\Sigma_X = (1 + 1/n_m)\Sigma_\mathrm{m}$. Note that this paradigm is also equivalent to assuming that

$$w_j \sim N\big(w^*, 2\Sigma_m\big), \quad \text{and} \quad \mathrm{Cov}(w_j, w_{j'}) = \Sigma_m \text{ if } j \neq j'. \tag{22}$$

In this way, this paradigm assumes that models are centered on the truth $w^*$, with a particular type of positive dependence among models. This tends to make confidence region around the multi-model mean larger than if independence is assumed.

We use this paradigm for two main reasons.

First, this paradigm assumes large uncertainties around the multimodel mean. We argue that it is more appropriate to assume larger rather than narrower modelling uncertainty to provide a conservative statement. Following previous D&A studies, we are primarily interested in deriving an observational constraint, rather than a strong modelling constraint, on the forced responses $X_i^*$, since we regard the evidence contained by the observations as being paramount. For robustness, we might wish this observational constraint to hold even with an overestimate of modelling uncertainty. This paradigm is also quite pessimistic because it means that no matter how extensively the space of all plausible models is sampled, the multimodel mean will not converge to the truth.

Second, Annan and Hargreaves (2010) and van Oldenborgh et al. (2013) suggest that the *models are statistically indistinguishable from the truth* paradigm is reasonably well supported by observations, although they partly disagree on whether it tends to be overly conservative or not. Notably, this paradigm is better supported by observation than alternatives, in particular the *models centered on the truth* paradigm (e.g. Annan and Hargreaves 2010; Fyfe et al. 2013), which is briefly introduced and discussed in 8.4. Note however that those results are mainly valid for temperature, but might be discussed for other variables such as precipitation.

As a conclusion, the "models are statistically indistinguishable from the truth" paradigm provides a useful framework to compute estimates of climate modelling uncertainty. Basically, this approach assumes that the truth is somewhere within the model envelope, and a possible drawback is that values outside the model range (e.g. in terms of sensitivity) will not be considered. This is the reason why the estimation of climate modelling uncertainty should account for multiple lines of evidence in order to determine whether or not this paradigm might be considered as reliable - this discussion goes beyond this paper. If not, other estimates might be considered, e.g. with an inflated variance, if the ensemble is proven to be underdispersive, with some models discarded if they are proven unrealistic, or with more specific adjustments. The same conclusion applies if some component of the uncertainty, such as the forcing uncertainty is ignored in the ensemble design. The sensitivity of the results to considering such alternative estimates might also be explored.

Lastly, it may be noted that the EIV approach, via the introduction of unknown scaling factors, is able to cope

with an additional uncertainty on the response magnitude. However, this approach still ignores part of the physical knowledge of the response magnitude, and more importantly, we see no evidence why models would be underdispersive in the magnitude and not in the patterns. Thus some assessment of the paradigm used will also be needed regarding uncertainty in the response pattern.

### 4.2 Estimation of $\Sigma_X$ with modelling uncertainty and internal variability

A more comprehensive framework includes repeated experiments from each climate model. We add some complexity to the previous description by considering that the response $w_{jk}$ simulated in run $k$ from model $j$ can be decomposed as

$$w_{jk} = \mu + m_j + \epsilon_{jk}, \tag{23}$$

where $\mu$ is the mean value of the population of climate models, $\mu + m_j$ is the mean value in model $j$, and $\epsilon_{jk}$ denotes the particular realization of internal variability contained in simulation $k$ from model $j$. We set $\epsilon_{jk} \sim N(0, \Sigma_\mathrm{v})$, assuming the variance matrix $\Sigma_\mathrm{v}$ to be the same for all models (i.e. doesn't depend on $j$), consistent with some previous D&A studies, and $m_j \sim N(0, \Sigma_\mathrm{m})$. Within such a framework, uncertainty is related to both internal variability and climate modelling uncertainty, consistent with Huntingford et al. (2006) and Hannart et al. (2014). We further assume these two random terms to be independent, which leads to

$$w_{jk} \sim N(\mu, \Sigma_\mathrm{m} + \Sigma_\mathrm{v}). \tag{24}$$

Then, as above, we use the "models are statistically indistinguishable from the truth" paradigm and assume that $(\mu - w^*) \sim N(0, \Sigma_\mathrm{m})$. Finally, we consider the balanced case in which $n_r$ runs are available from each climate model.

Considering this whole set of assumptions, each individual ensemble mean $w_{j.} = 1/n_r \sum_{k=1}^{n_r} w_{jk}$ satisfies

$$w_{j.} \sim N\left(\mu, \Sigma_\mathrm{m} + \frac{\Sigma_\mathrm{v}}{n_r}\right), \tag{25}$$

and the multimodel mean $\overline{w} = 1/n_m \sum_{j=1}^{n_m} w_{j.}$ (assuming $n_m$ models are involved) satisfies

$$\overline{w} \sim N\left(\mu, \frac{\Sigma_\mathrm{m}}{n_m} + \frac{\Sigma_\mathrm{v}}{n_m n_r}\right). \tag{26}$$

Last,

$$(\overline{w} - w^*) \sim N\Big(0, \big(1 + \frac{1}{n_m}\big)\Sigma_\mathrm{m} + \frac{\Sigma_\mathrm{v}}{n_m n_r}\Big), \tag{27}$$

which suggests considering

$$\Sigma_X = \left(1 + \frac{1}{n_m}\right)\Sigma_\mathrm{m} + \frac{1}{n_m n_r}\Sigma_\mathrm{v}. \tag{28}$$

While (28) has been obtained in the balanced case where each model performs the same number of simulations $n_r$, modelling centers usually provide ensembles of various sizes. One possible way to obtain a balanced ensemble would be to consider only a fixed number of simulations from each model. This usually requires a compromise between $n_m$ and $n_r$, and tends to exclude some of the available simulations (i.e. not consider all available information). In order to avoid this, we show in "Appendix 8.3" how $\Sigma_X$ and $\Sigma_\mathrm{m}$ may be estimated in the more general unbalanced case (i.e. ensembles of various sizes).

## 5 Properties and several illustrations

This section illustrates a few properties of the proposed method. We first discuss how it compares with and relates to previous linear regression approaches. We then describe the results obtained in a few particular cases, based on synthetic data and with a very small dimension $n = 1, 2$, to illustrate how this method works.

### 5.1 Relationship with linear regression

In order to compare our method to regression-based methods, we consider the most basic regression framework, which assumes no error in the predictors $X$ (usually referred to as Ordinary Least Squares in D&A, following Allen and Tett 1999a), i.e.

$$Y = X\beta + \varepsilon_Y, \qquad \varepsilon \sim N(0, \Sigma_Y). \tag{29}$$

We also consider a very simple case where $n_f = 1$ and consequently, $X$ is a column vector.

If we consider the term $X\beta$ as being the response to a forcing $F_1$, the basic assumption underlying this framework is that the shape of the response to $F_1$ (i.e. $X$) is perfectly known, while the magnitude of the response to $F_1$ (i.e. $\beta$) is unknown. This assumption may be thought of as defining a very specific uncertainty structure on the $F_1$ response in (3). The "unknown magnitude" means that uncertainty proportional to the simulated response $X$ is large, while uncertainty vanishes in other directions.

To understand the implications of assuming exact knowledge of the direction of $X$, but uncertain knowledge of its amplitude, we may write the variance matrix of $X$ as

$$\Sigma_X = \lambda XX', \tag{30}$$

where $\lambda$ represents the variance of the magnitude of $X$. One natural question is to determine whether these two approaches provide consistent results if based on similar assumptions. The answer is yes because, based on (30),

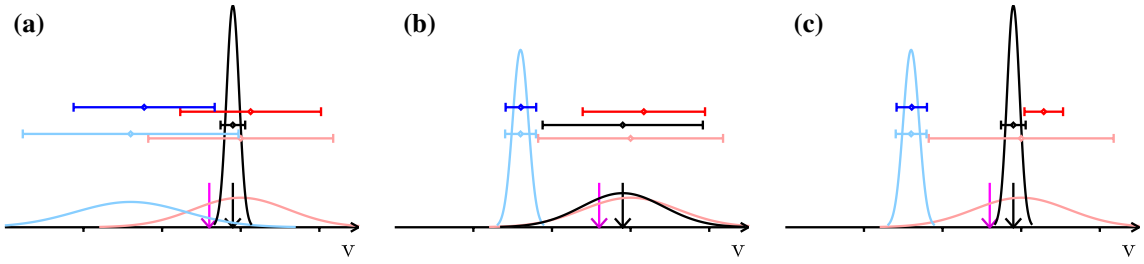


**Fig. 2** Schematic illustration of 1D reduction of uncertainties Contributions from two forcings $F_1$ and $F_2$ are assessed. *Black*: observation $Y$ (*arrow*), the related uncertainty, including both internal variability and measurements errors (pdf), and the confidence region for $Y^*$, as derived from (2) (interval). *Orange*: uncertainty on $X_1$ (pdf) and the corresponding confidence interval for $X_1^*$ as derived from (3) (interval). *Red*: confidence region for $X_1^*$ based on inference in the full model. *Sky blue*: uncertainty on $X_2$ (pdf) and the corresponding confidence interval for $X_2^*$, as derived from (3) (interval). *Blue*: confidence region, for $X_2^*$, based on inference in the full model. *Magenta*: sum of simulated responses $X = X_1 + X_2$ (*arrow*). Panels **a**, **b** and **c** are representative of three different configurations (see text)

and assuming that $\lambda \to \infty$, we can show that our method coincides with linear regression. In particular, noting that

$$\left(\Sigma_Y + \lambda XX'\right)^{-1} \xrightarrow[\lambda\to\infty]{} \Sigma_Y^{-1} - \Sigma_Y^{-1}X(X'\Sigma_Y^{-1}X)^{-1}X'\Sigma_Y^{-1}, \tag{31}$$

(12) becomes

$$\widehat{Y}^* = Y + \Sigma_Y\left(\Sigma_Y^{-1} - \Sigma_Y^{-1}X(X'\Sigma_Y^{-1}X)^{-1}X'\Sigma_Y^{-1}\right)(X - Y), \tag{32}$$

$$= X(X'\Sigma_Y^{-1}X)^{-1}X'\Sigma_Y^{-1}Y, \tag{33}$$

$$= X\widehat{\beta}, \tag{34}$$

where $\widehat{\beta}$ denotes the optimal least squares estimate in (29). The response $Y^*$ estimated by our method, $\widehat{Y}^*$, is then the same as that estimated using linear regression. This shows that linear regression may be regarded as a limiting case of our method (as $\lambda \to \infty$), and that both approaches are consistent. Similar consistency is also found in terms of the uncertainty analysis (e.g. confidence intervals). This result is important, because it means that if the assumptions underlying linear regression are valid—i.e. climate models do agree on the patterns and the magnitude is uncertain—then our approach will provide results similar to a regression based method.

Two comments should be mentioned on this topic however. First, comparison might have been performed with a more comprehensive framework for linear regression, such as the EIV approach proposed in Hannart et al. (2014). However, comparison to such a model is much more difficult because maximum likelihood estimates are not explicit. Instead, using OLS as a baseline allows us to make a simple comparison. Second, this result helps to illustrate the main difference between these two possible approaches. While linear regression is equivalent to assuming no error in the shape ($\Sigma_X$ is rank-1) and infinite uncertainty in the magnitude ($\lambda \to \infty$), our approach allows a more balanced point of view, basically assuming there is a limited (i.e. finite) amount of uncertainty in both the shape and the magnitude of the response $X$.

### 5.2 Toy examples

We here consider a few practical cases in order to illustrate particular features of our method, in particular with respect to its ability to estimate contributions from individual forcings.

#### *5.2.1 Single scalar analysis $n = 1$*

We first consider the case of a univariate analysis where the main concern is to assess the contributions of two different forcings $F_1$, $F_2$ to the observed change $Y$ on the climate variable v. More specifically, we assume estimation of $X_1$ to be our main concern, and discuss the accuracy of the estimation as a function of uncertainties in both $X_2$ and $Y$.

Figure 2 illustrates three typical cases that might be encountered in this configuration. In this figure, we distinguish between two possible ways to derive confidence intervals for the forced contributions $X_1^*$ and $X_2^*$. The first, very naive option is to consider only information provided by climate models (in terms of both the mean spread of multimodel ensembles) using (3). Another option is to consider our full statistical model (1)–(3) and thus information provided by both models and observations, which leads to our estimates $\widehat{X}_1^*$ and $\widehat{X}_2^*$. Figure 2 describes how these two options may provide different results.

In Fig. 2a, the two contributions $X_1^*$ and $X_2^*$ are very uncertain based on multimodel information. As one single observation is not sufficient to separately estimate the contributions from two different forcings, the added value from observations is limited and the resulting confidence intervals

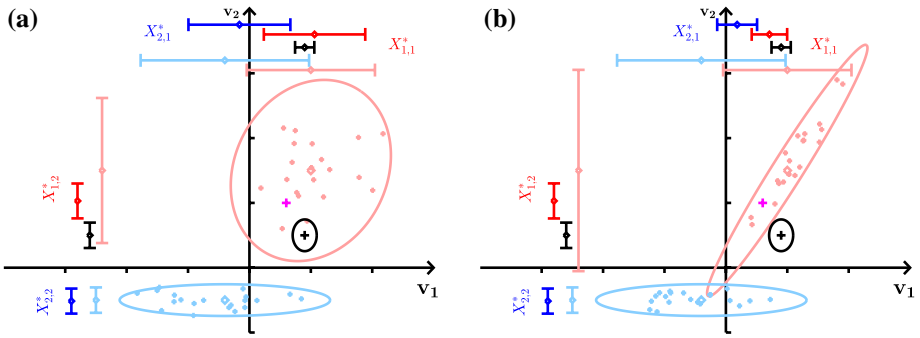


**Fig. 3** Schematic illustration of 2D reduction of uncertainties Contributions from two forcings $F_1$ and $F_2$ are assessed. *Black*: observation $Y$ (*cross*) and the related uncertainty (confidence intervals and ellipsoid for $Y^*$, as derived from (2)), including both internal variability and measurements errors. *Pink*: simulated response to forcing $F_1$, including the simulated response $X_1$ (*diamond*), individual synthetic models (*cross*), confidence region for $X_1^*$ as derived from (3) only (*ellipse*), and corresponding confidence intervals for each coordinate, $X_{1,1}^*$ and $X_{1,2}^*$. *Sky blue*: simulated response to forcing $F_2$ (similar). *Red*: confidence intervals for $X_{1,1}^*$ and $X_{1,2}^*$ based on inference in the full model. *Blue*: equivalent confidence intervals for $X_{2,1}^*$ and $X_{2,2}^*$. *Magenta cross*: sum of simulated responses $X = X_1 + X_2$. Panels **a** and **b** illustrate two typical cases (see text)

on $X_1^*$ and $X_2^*$ are only slightly reduced after accounting for the observational constraint. Note that observations (black arrow) fall above the overall forced response as simulated by climate models (magenta arrow). As a consequence, best-estimates $\widehat{X}_1^*$ and $\widehat{X}_2^*$ are slightly higher than the first guess $X_1$ and $X_2$ simulated by the climate models.

In Fig. 2b, the contribution $X_2^*$ is rather constrained by model simulations, but observations are dominated by noise, and so the resulting uncertainty on $X_1^*$ is, again, only slightly reduced. As in panel **a**, the best-estimate $\widehat{X}_1^*$ is found to have a higher value than the multimodel mean $X_1$ because the observation $Y$ falls above the simulated forced response $X = X_1 + X_2$. However, because observations are noise dominated, only part of this departure $(Y - X)$ is transferred to $\widehat{X}_1^* - X_1$; the remaining part is actually transferred to $(\widehat{Y}^* - Y)$.

Figure 2c illustrates a more favourable case in which both observations and the contribution from forcing $F_2$ are well-constrained. In this case, while climate models simulate a relatively uncertain response to forcing $F_1$ (large pink confidence interval), the use of our statistical model leads to a much better constrained estimate of the response $X_1^*$ (red confidence interval). This can be understood as a direct consequence of the additivity assumption. Basically, (1) can be re-written $X_1^* = Y^* - X_2^*$. Because of limited uncertainties (i.e. relatively small $\Sigma_Y$ and $\Sigma_{X_2}$), $Y$ and $X_2$ are close to $Y^*$ and $X_2^*$, respectively. So computing $X_1^*$ from the subtraction above allows uncertainties to be reduced substantially. Furthermore, the departure of $Y$ from $X$ is almost completely transferred to $\widehat{X}_1^* - X_1$, as the two other terms are much less uncertain. In this toy example, while $Y - X$ is actually not very large, the resulting confidence interval for $X_1^*$ doesn't contain $X_1$ (which could be considered as a naive first guess).

### 5.2.2 *Analysis when Y has dimension n = 2*

In order to illustrate how uncertainties may be substantially reduced in the presence of two somewhat uncertain responses, we now consider a 2-dimensional example, $n = 2$. We still consider two forcings $F_1$ and $F_2$ and assume that our primary interest is to estimate the contribution of these forcings to the variable $\mathrm{v}_1$ ($x$-axis on Fig. 3). Let $X_{1,1}^*$ and $X_{2,1}^*$ be the true responses to $F_1$ and $F_2$ respectively on $\mathrm{v}_1$, and let $X_{1,2}^*, X_{2,2}^*$ be the true responses to $F_1$ and $F_2$ on $\mathrm{v}_2$. In these examples, we assume that uncertainty in the observations $Y$ is relatively small, e.g. because internal variability has somehow been partly filtered out (see Sect. 6 for a concrete illustration). We then discuss how the accuracy of the final estimate of $X_{1,1}^*$ depends on the structure of the uncertainty on $X_1$ and $X_2$, i.e. $\Sigma_{X,1}$ and $\Sigma_{X,2}$.

In Fig. 3a, for both $X_1$ and $X_2$, the uncertainty in the $\mathrm{v}_1$ direction is essentially independent from that in the $\mathrm{v}_2$ direction, which means that $\Sigma_{X,1}$ and $\Sigma_{X,2}$ are close to being diagonal. Therefore, these two dimensions $\mathrm{v}_1$ and $\mathrm{v}_2$ may be considered separately. First, on $\mathrm{v}_1$, both $X_1$ and $X_2$ are quite uncertain. Consistent with Fig. 2a, the observational constraint is weak and $X_{1,1}^*$ is poorly estimated. Second, on $\mathrm{v}_2$, the situation is closer to Fig. 2c, and $X_{1,2}^*$ is more accurately estimated. This, however, has little impact on the estimation of $X_{1,1}^*$.

The case illustrated in Fig. 3b is much different; the simulated response $X_1$ has the same variance as in panel 3a on both $\mathrm{v}_1$ and $\mathrm{v}_2$, but there is strong dependence between the two components. The more $X_1$ deviates from $X_1^*$ on $\mathrm{v}_1$, the more it deviates on $\mathrm{v}_2$. Thus, based on (3) only, the confidence region for $X_1^*$ (pink ellipsoid) is stretched in one particular direction. We chose this direction to be the same as $X_1$, so that the uncertainty is closer to that which is assumed

when linear regression is performed (see Sect. 5.1)—but some constraint on the magnitude of the change is still available from climate models.

In Fig. 3b, under these assumptions, there is a strong observational constraint on estimates of both $X^*_{1,1}$ and $X^*_{2,1}$. This can be understood as follows. First, in the $v_2$ direction, the case is close to that of Fig. 3a, and the contribution $X^*_{1,2}$ of $F_1$ is well constrained. Second, the assumed shape of the uncertainty on $X_1$ ensures that this constraint on $v_2$ projects quite clearly on $v_1$, and thus a relatively small confidence interval is found for $X^*_{1,1}$. Third, this new constraint also allows accurate estimation of the contribution $X^*_{2,1}$ of forcing $F_2$ on $v_1$. Finally, thanks to the particular shape of the modelling uncertainty, the observational constraint is relatively strong on each component. This example is also particularly illustrative because, on $v_1$, the observation $Y$ falls above the expected total response (*magenta cross*), but the response to $F_1$ ($X^*_{1,1}$) is finally assessed to be smaller than simulated in $X_{1,1}$. This is due to two facts: (1) on $v_2$, the observation $Y$ falls below the expected total response, and (2) the observational constraint on $v_1$ comes from that on $v_2$.

The example of Fig. 3b also illustrates a strong discrepancy with the linear regression approach. Here, the simulated responses $X_1$ and $X_2$ are close to being collinear. Discrimination of the responses to $F_1$ and $F_2$ would not have been successful based on the linear regression approach, as a direct consequence of this collinearity. In particular the EIV inference method proposed by Hannart et al. (2014) provides unbounded confidence intervals if applied to the same data. Instead, the observational constraint found above is actually a consequence of the strong constraint provided by $X_2$ on the magnitude of $X^*_{2,2}$, which is appropriately taken into account here. It may also be noted that while the structure of the uncertainty on $X_1$ is consistent with a linear regression approach, that of $X_2$ is not. Our statistical model is also able to appropriately deal with such different cases.

## 6 Application to global mean temperature

As a simple illustration of an application of the method to real data, we consider the global mean warming over the period 1951-2010, as estimated with a linear trend. This period is selected in order to be consistent with the Fig. 10.5 of the IPCC Fifth Assessment Report (IPCC 2013). Based on this example, we illustrate the capabilities of our method, both in terms of hypotheses testing and in terms of estimation of individual forcing contributions. This is done in a 2-forcing analysis (considering both natural and anthropogenic external forcing) of a simple scalar diagnostic, as proposed in Sect. 2. Application of our method to more comprehensive datasets, possibly including time-series, and/or spatial information, is beyond the scope of this methodological paper. It may also be noted that, given the limited number of climate models available worldwide, estimating climate modelling uncertainty variance matrices necessitates working with a reasonably small dimension $n$.

### 6.1 Data

Observed temperature data are taken from the median realization of the HadCRUT4 merged land/sea temperature data set (Morice et al. 2012). We use outputs from unforced pre-industrial control simulations, historical simulations performed with all external forcings combined (ALL) and historical simulations with natural forcings only (NAT), from all available CMIP5 models (see Table 1). These data are available at: http://cmip-pcmdi.llnl.gov/cmip5/. For models providing both ALL and NAT ensembles, the response to anthropogenic forcings (ANT) is computed as the difference between the ALL and NAT responses. As our analysis is based on the 1951-2010 period, we did not consider models providing historicalNat simulations with end dates earlier than 2010. When required, ALL simulations were extended to 2010, either with historicalExt experiments, if available, or with RCP8.5 simulations.

The pre-processing of data involves the following steps. Model outputs are interpolated onto a common 5°× 5°regular grid. Long control simulations are divided into non-overlapping 60-year segments (their number varies among models, see Table 1). We compute anomalies with respect to the 1961–1990 period (or the corresponding period in control segments). The spatio-temporal observational mask is applied to each simulated 60-year period. Following the HadCRUT4 dataset, the global mean temperature is then computed as the average of Northern Hemisphere and Southern Hemisphere mean temperatures. Each of these is computed as the area-weighted average of available grid-points. Lastly, we compute the linear trend with a least square fit.

### 6.2 Results

Figure 4 illustrates results obtained at different stages of the D&A analysis. Note that 90 % confidence intervals are reported. The analysis presented in this section is performed under the "models are statistically indistinguishable from the truth" paradigm. The equivalent analysis assuming the "models are centered on the truth" paradigm is presented and briefly discussed in "Appendix 8.4".

**Table 1** Ensembles of CMIP5 simulations used

| Climate model | Nb 60-year PICTL seg. | Nb ALL Runs | Nb NAT Runs |
|---|---|---|---|
| ACCESS1-0 | 9 | 1 | – |
| ACCESS1-3 | – | 1 | – |
| bcc-csm1-1 | 22 | 3 | 1 |
| bcc-csm1-1-m | – | 1 | – |
| BNU-ESM | 24 | 1 | – |
| CanESM2 | 46 | 5 | 5 |
| CCSM4 | 22 | 6 | – |
| CESM1-BGC | – | 1 | – |
| CMCC-CM | – | 1 | – |
| CMCC-CMS | – | 1 | – |
| CMCC-CESM | – | 1 | – |
| CNRM-CM5 | 47 | 10 | 6 |
| CSIRO-Mk3-6-0 | 22 | 10 | 5 |
| EC-EARTH | – | 5 | – |
| FGOALS-g2 | – | 2 | 2 |
| FGOALS-s2 | 22 | 3 | – |
| FIO-ESM | – | 1 | – |
| GFDL-CM3 | 22 | – | – |
| GFDL-ESM2G | 4 | 1 | – |
| GFDL-ESM2M | – | 1 | – |
| GISS-E2-H | 41 | 5 | 5 |
| GISS-E2-R | 70 | 5 | 5 |
| HadGEM2-ES | 9 | 4 | 4 |
| HadGEM2-CC | – | 1 | – |
| inmcm4 | 22 | 1 | – |
| IPSL-CM5A-LR | 47 | 4 | 3 |
| IPSL-CM5A-MR | – | 1 | – |
| IPSL-CM5B-LR | 12 | 1 | – |
| MIROC5 | 7 | 4 | – |
| MIROC-ESM | – | 1 | – |
| MIROC-ESM-CHEM | – | 1 | – |
| MPI-ESM-LR | 6 | 3 | – |
| MPI-ESM-MR | 22 | 1 | – |
| MRI-CGCM3 | 22 | 3 | – |
| NorESM1-M | 22 | 3 | 1 |

Figure 4a illustrates the detection step, where observations are only compared to unforced simulations. Estimated from the linear trend over the whole period, the observed warming is about +0.65K. By contrast, the linear warming of global mean temperature found in segments taken from unforced simulations has mean zero and a standard deviation of about 0.08K. The observed value is thus well outside the range of values expected as a consequence of internal variability alone (±0.13K at the 90 % confidence level). Detection, based on (18), is very significant, as the $p$ value is numerically indistinguishable from 0.

Figure 4b illustrates the comparison of observations to the ALL-forcings simulations. This comparison allows us to test the consistency between the observed and simulated global warming, and to estimate the overall forced response. The range of values obtained from individual climate models is quite large, ranging from +0.44K to +1.16 K. Consistent with this, the forced warming estimated from (3) only is about +0.80K, with a 90 % confidence range of [+0.46K, +1.15K]. The observed value (+0.65K) is smaller than the center of this distribution. Given the expected range of internal variability, the forced response based on (2) only is within [+0.52K, +0.78K]. Considering the whole model (1)–(3), the confidence interval for the overall forced response is very similar, [+0.55K, +0.79K], with a best estimate at +0.67K. Observations are also found to be very consistent with simulated responses, as the $p$ value of the test (19) is 0.51.

Figure 4c illustrates how the contributions from natural and anthropogenic external forcings may be estimated. Note that a smaller ensemble of models has been used here as many CMIP5 models did not run D&A simulations (i.e. historical simulations using specified subsets of forcings). Models simulate a response to natural forcing of −0.01K ([−0.03K, +0.02K]). This range of values is particularly narrow, as the NAT response is weak in all models. Surprisingly, this range of values is also much narrower than that reported in Fig. 4a from internal variability only. This is partly because ensemble means are considered, which are less impacted by internal variability than individual runs. Related to this, the estimated modelling uncertainty on this term is exactly 0 (which may happen with the truncated estimate used, see "Appendix 8.3"). This means that the discrepancy between models is consistent with internal variability only. The simulated response to anthropogenic forcing is much more uncertain, with a warming of +0.80K on average, and a 90 % range of [+0.39K, +1.21K]. Applying our method leads to reduced uncertainty on the latter term, [+0.55K, +0.80K] with a best-estimate of +0.67K, while the natural contribution is left virtually unchanged. Note that the main limitation to a stronger observational constraint comes from noise in the observations, which is dominated by internal variability, and not from uncertainty in the contribution of natural forcings, which is very small here. Based on this 2-forcing analysis, observations are also found to be consistent with models, with a $p$ value of 0.60. A final piece of information from this analysis is the attribution test of *other plausible causes*. Based on this analysis of linear warming trends only, we find that the anthropogenic influence is unequivocal, as the hypothesis that *natural forcings alone can explain observations* is strongly rejected (the $p$ value is, again, numerically indistinguishable from 0). The symmetric hypothesis that *anthropogenic forcings alone can explain observations* is not rejected ($p$

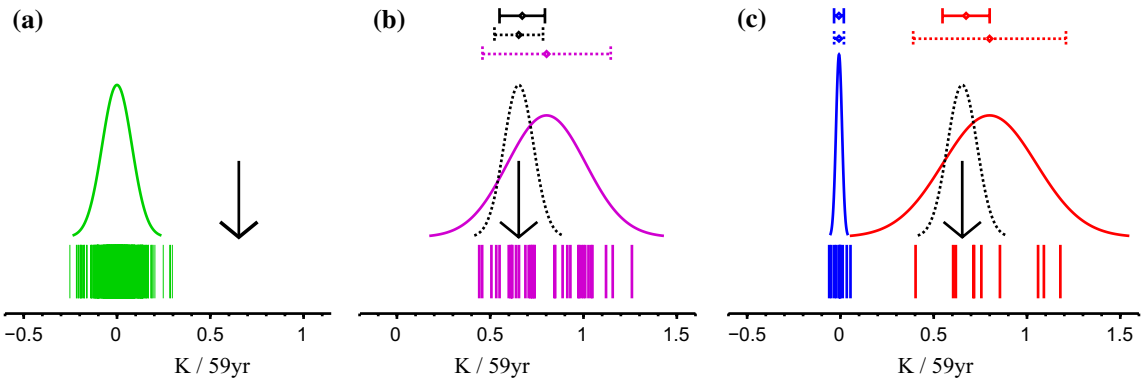


**Fig. 4** Univariate D&A on the 1951–2010 linear trend on global mean temperature. Comparison of observations (*black arrow*) to unforced control simulations (*left*), historical simulations with all external forcings (*middle*), or historical simulations with natural forcings only in blue and anthropogenic forcings only in *red* (*right*). *Middle*: confidence intervals of $Y^*$ based on: climate models only (i.e. Eq. (3), *dotted magenta*), observations only (i.e. Eq. (2), *dotted black*), and the full inference proposed (*black*). *Right*: confidence intervals of $X^*_{\mathrm{ANT}}$ (*red*) or $X^*_{\mathrm{NAT}}$ (*blue*) are based on: climate models only (i.e. Eq. (3), *dotted*) or the full inference proposed (*solid*). Uncertainty related to internal variability in the observations is reported as a dotted black pdf (*middle* and *right*)

value 0.58). The natural influence is thus much more difficult to demonstrate, consistent with the very small response simulated to natural forcings over this period of time.

Our estimates of natural or anthropogenically-induced warming are well consistent with Figure 10.5 from the last IPCC report, in the sense that the IPCC estimates are included within our intervals. Our ranges, however, are slightly wider. Part of this might be due to fact that we compute 90 % confidence ranges, while the IPCC ranges were assessed to be only *likely* (i.e. 66 %). Part of this difference is also expected as no temporal or spatial information is accounted for here. Our estimates come only from the constraint provided by the linear warming over this period. Applying our procedure to a more comprehensive observed vector $Y$ in order to more efficiently distinguish between internal variability and the expected responses to forcing, and to further reduce these uncertainties, would be a natural continuation of this work.

## 7 Conclusion

D&A has, for the most part, used regression-like approaches for the past two decades, where observations are regressed onto expected response patterns. These methods tend to ignore the information provided by climate models on the magnitude of the forced responses. Accounting for climate modelling uncertainty in such regression-based approaches is quite challenging, and thus it has also been common practice to neglect this type of uncertainty.

We have introduced a revised statistical framework for D&A that overcomes these weaknesses. Our approach relies on the additivity of the forcing responses to provide an observational constraint on the contribution of each forcing. This method is able to deal with climate modelling uncertainty. The information provided by an ensemble of climate models is then used both in terms of the response patterns and the response magnitudes, in a very symmetrical way.

This paper describes statistical inference methods required for D&A within this new statistical framework. The estimation of each forced response is based on a maximum likelihood method. Closed-form estimators and exact confidence regions are provided, as opposed to regression-based methods like EIV. Hypothesis tests that are of interest to formally attribute an observed change to some combination of forcings are also presented and discussed. In particular we provide likelihood ratio tests and their null-distribution.

We provide some guidelines on quantifying climate modelling uncertainty from an ensemble of opportunity. Following previous studies, we consider the *models are statistically indistinguishable from the truth* paradigm, where the truth is assumed to lie somewhere within the model envelope, to obtain a more conservative estimate of this uncertainty. The reliability of such a paradigm, however, might be investigated further. We also discuss in an Appendix how this uncertainty may be estimated if the number of simulations available varies among models. This approach is not expected to strongly affect inferences because the strongest constraints on the estimated responses to forcing is from observations.

Our additive decomposition method is shown to have good properties based on simple synthetic examples. In particular, we illustrate how the observational constraint on forced contributions is influenced by the structure of the climate modelling uncertainty - i.e. the corresponding variance matrices. We also demonstrate that our approach is

equivalent to linear regression in the particular case where models agree on the response pattern but widely disagree on the response magnitude.

Application of this method to the analysis of the contributions of anthropogenic and natural external forcing to the linear 1951-2010 trend in global mean temperature provides results that are very consistent with the recent IPCC AR5. We find that the observed warming over this period (+0.65K) is mostly related to anthropogenic forcings ($+0.67 \pm 0.12$K), with a very limited contribution from natural forcings ($-0.01 \pm 0.02$K). Application of the same method with space-time information might further reduce these ranges.

Assessing the extent to which this new method may improve the observational constraint on other variables or other external forcings would be a natural continuation of this work.

**Acknowledgments** The authors are grateful to the two anonymous referees for their constructive comments, which were of great value in improving the paper. Part of this work has been supported by the Fondation STAE, via the project Chavana, and by the Extremoscope and ANR-DADA projects.

## Appendix

### Model (1)–(3) as a linear regression Gaussian model

Based on the notation used in (1)–(3), we define

$$Z = \begin{pmatrix} Y \\ X_1 \\ \vdots \\ X_{n_f} \end{pmatrix}, \qquad Z^* = \begin{pmatrix} X_1^* \\ \vdots \\ X_{n_f}^* \end{pmatrix}, \quad \text{and} \quad \varepsilon = \begin{pmatrix} \varepsilon_Y \\ \varepsilon_{X_1} \\ \vdots \\ \varepsilon_{X_{n_f}} \end{pmatrix}. \tag{35}$$

which are vectors of size $(n_f+1)n$, $n_f n$, and $(n_f+1)n$, respectively. Then, (2) and (3) may be written simultaneously as

$$Z = AZ^* + \varepsilon, \tag{36}$$

where

$$A = \begin{pmatrix} I_n & I_n & \dots & I_n \\ I_n & 0 & \dots & 0 \\ 0 & I_n & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & I_n \end{pmatrix}, \tag{37}$$

is the $(n_f+1)n \times n_f n$ design matrix of the linear Gaussian model (36).

This representation is useful to derive some of the statistical properties of our model. It is used, for instance, in "Appendix 8.2.3". $A'A$ has also a simple closed form. However, it is difficult to derive the precision matrix $\left(A'A\right)^{-1}$ in closed form, which would be required to deduce, e.g., the MLEs more directly than in Sect. 3.3.

### Hypothesis testing details

*Goodness of fit tests*

The tests proposed in Sect. 3.5 are goodness of fit tests constructed as follows.

If $Z \in \mathbb{R}^n$ is a random vector with distribution $N(\mu, \Sigma)$ under a given model, with $\mu$ and $\Sigma$ known, we will call "goodness of fit" test of this model the deviance test, i.e. the likelihood ratio test (LRT) with respect to a saturated alternative hypothesis $H_1 : Z \sim N(\theta, \Sigma)$, where $\theta \in \mathbb{R}^n$ is unknown. The log-likelihood is 0 under this alternative, so the LRT only involves $-2$ log-likelihood under $H_0$, the considered model. Therefore the LRT statistic is

$$T = (Z-\mu)'\Sigma^{-1}(Z-\mu), \tag{38}$$

which follows a $\chi^2(n)$ distribution under $H_0$.

*Minimized likelihood under $H_0 : Y^* = 0$ in model (1)–(3)*

Under $H_0 : Y^* = 0$, each $X_i^*$ has to be estimated under the additional constraint that $\sum_{i=1}^{n_f} X_i^* = 0$. Under $H_0$, (6) becomes

$$\ell_{H_0}(X_i^*) = Y'\Sigma_Y^{-1}Y + \sum_{i=1}^{n_f} \left(X_i - X_i^*\right)'\Sigma_{X_i}^{-1}\left(X_i - X_i^*\right). \tag{39}$$

The gradient of $\ell_{H_0}$ with respect to $X_i^*$ is

$$\Sigma_{X_i}^{-1}(\widehat{X}_i^* - X_i). \tag{40}$$

The gradient of the constraint is $\mathbb{1}_n$, i.e. the "all-one" vector of size $n$. The theory of Lagrange multipliers imposes that at the constrained minimum, these two gradients are proportional, so that

$$\Sigma_{X_i}^{-1}(\widehat{X}_i^* - X_i) = \lambda \mathbb{1}_n, \tag{41}$$

where $\lambda$ is some constant. From (41),

$$\widehat{X}_i^* = X_i + \lambda \Sigma_{X_i} \mathbb{1}_n, \tag{42}$$

and thus the constraint $\sum_{i=1}^{n_f} \widehat{X}_i^* = 0$ gives

$$\lambda \mathbb{1}_n = -\Sigma_X^{-1} X, \tag{43}$$

where $X = \sum_{i=1}^{n_f} X_i$ and $\Sigma_X = \sum_{i=1}^{n_f} \Sigma_{X_i}$. The maximum likelihood estimate of $X_i^*$ under $H_0$ is therefore

$$\widehat{X}_i^* = X_i - \Sigma_{X_i} \Sigma_X^{-1} X. \tag{44}$$

Finally, the minimized value of $-2$ log-likelihood under $H_0$ is

$$\ell_{H_0}(\widehat{X}_i^*) = Y' \Sigma_Y^{-1} Y + X' \Sigma_X^{-1} X. \tag{45}$$

This result is used in "Appendix 8.2.3", Eq. (46), to derive the LRT of $H_0 : Y^* = 0$ within the statistical model defined by (1)–(3).

*Choice of the null and the alternative hypotheses in the detection test*

The "detection" test deals with the null hypothesis of no change, i.e. $H_0 : Y^* = 0$. The detection test we propose in (18) is a goodness of fit test based on (2) only. In particular, it doesn't consider (1) and (3). Here, we present two alternatives that might be considered to test the same null hypothesis. Both are LRTs between two nested well-defined hypotheses on the data $(Y, X)$.

- In the statistical model (1)–(3), consider the null-hypothesis $H_0 : Y^* = 0$ versus the alternative hypothesis $H_1$ :(1)–(3), ie the same model with an unspecified $Y^*$. Following (45) and (14), this test would be based on the statistic

$$\ell_{H_0} - \ell_{H_1} = Y' \Sigma_Y^{-1} Y + X' \Sigma_X^{-1} X - (Y - X)' (\Sigma_Y + \Sigma_X)^{-1} (Y - X) \sim_{H_0} \chi^2(n). \tag{46}$$

  Note that the distribution under $H_0$ is known to be $\chi^2(n)$ since $H_0$ is a linear sub-hypothesis of $H_1$, and they differ by a dimension of $n$ (see 8.1).
- In the statistical model (1)–(3), consider the null-hypothesis $H_0 : Y^* = 0$ versus the saturated alternative hypothesis. This test is a goodness of fit test as defined above. Such a test would be based on the statistic (see (45))

$$\ell_{H_0} = Y' \Sigma_Y^{-1} Y + X' \Sigma_X^{-1} X \sim_{H_0} \chi^2(2n), \tag{47}$$

  where the null-distribution is deduced directly from (2) and (3).

Both of these tests would treat information from $Y$ and $X$ very symmetrically because, given (1), $Y^* = 0$ implies not only that $Y$ is small, but also that $X$ is small (as $X^* = 0$). Therefore, rejection of these tests may happen because either $Y$ or $X$ are "large". In the case where $X$ is large while $Y$ is not, detection would not be a consequence of an abnormal observation, but rather the consequence of too large a response simulated in climate models. This, of course, is not consistent with the definition of detection. In our opinion, it is then more appropriate to discuss the first term in the right hand side of (45) separately.

We further argue that there is a fundamental distinction between detection and attribution in this respect. As detection only assesses whether observations are consistent with internal variability, historical simulations by climate models ($X$) are not required for detection. The only requirement is to quantify internal variability—which is usually done based on other simulations by climate models. Attribution, however, definitively relies on historical simulations to disentangle contributions from different forcings based on some physical knowledge. This fundamental difference is the main reason why we propose to base detection on (2) only, while (1)–(3) are considered as a whole to perform attribution, and in particular to estimate the contributions of individual forcings.

As a last remark, the test defined in (18) may also be considered as a test of the null hypothesis $H_0$ : "$Y^* = 0$ and (1) does not hold". Indeed, removing (1) implies that the second term in 45 is zero, and then our test (18) is actually a LRT against a saturated alternative hypothesis.

**Estimation of $\Sigma_{\mathrm{m}}$ and $\Sigma_X$ with unbalanced data**

This section deals with the realistic case where models have run ensembles of historical simulations of various sizes. For instance, in the CMIP5 archive, the number of historical simulations with all external forcings varies from 1 to 10 depending on the model considered. Here, we mainly discuss the estimation of $\Sigma_{\mathrm{m}}$, which is not explicitly addressed in the main text. We also mention what $\Sigma_X$ should be considered in this unbalanced case.

Consistent with Sect. 4, we assume that simulation $k$ of model $j$ can be decomposed as

$$w_{jk} = \mu + m_j + \epsilon_{jk}, \quad j = 1, \ldots, n_m, \quad k = 1, \ldots, n_j, \tag{48}$$

where $m_j \sim N(0, \Sigma_{\mathrm{m}})$, and $\epsilon_{jk} \sim N(0, \Sigma_{\mathrm{v}})$, leading to

$$w_{jk} \sim N(\mu, \Sigma_{\mathrm{m}} + \Sigma_{\mathrm{v}}). \tag{49}$$

We then introduce

$$w_{j.} = \frac{1}{n_j} \sum_{i=1}^{n_j} w_{jk} \sim N\left(0, \Sigma_{\mathrm{m}} + \frac{\Sigma_{\mathrm{v}}}{n_j}\right). \tag{50}$$

This kind of framework is actually a multivariate linear mixed model, and useful references could be found in Rao and Kleffe (1988). Within such models, the main challenge comes from the estimation of variance components (here $\Sigma_{\mathrm{m}}$ and $\Sigma_{\mathrm{v}}$), and no optimal estimator is known in the general unbalanced case (i.e. if the ensemble sizes $n_j$ are not all equal). We propose to use a method of moments approach very similar to that proposed by Henderson (1953), but with a couple of specific features:

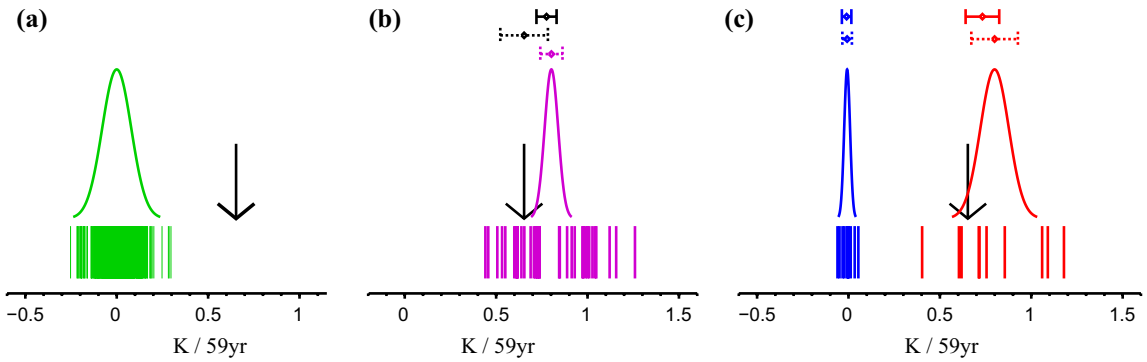


**Fig. 5** Same as Fig. 5 based on the "models are centered on the truth" paradigm

- Consistent with common practice in climate science, we estimate the fixed effect $\mu$ as the mean of the ensemble means from each model. In this way each model is given equal weight, disregarding the number of simulations performed. Consequently, we consider $\widehat{\mu} = \overline{w} = \frac{1}{n_m}\sum_{j=1}^{n_m} w_{j.}$, whereas the common approach in statistics uses $\widehat{\mu} = w_{..} = \frac{1}{n}\sum_{j,k} w_{jk}$.
- As a large number of unforced simulations are available to estimate $\Sigma_{\mathrm{v}}$ (Table 1), we assume that this term is already known, and only focus on the estimation of $\Sigma_{\mathrm{m}}$. Note that within-ensemble differences might also be used to estimate $\Sigma_{\mathrm{v}}$.

- One has:

$$\overline{w} = \frac{1}{n_m}\sum_{j=1}^{n_m} w_{j.} \quad \sim \quad N\left(\mu, \frac{1}{n_m}\Sigma_{\mathrm{m}} + \frac{1}{n_m^2}\sum_{j=1}^{n_m}\frac{1}{n_j}\Sigma_{\mathrm{v}}\right). \tag{51}$$

$$\mathrm{Var}(w_{j.} - \overline{w}) = \left(1 - \frac{1}{n_m}\right)\Sigma_{\mathrm{m}} + \left(\frac{1}{n_j} - \frac{2}{n_m n_j} + \frac{1}{n_m^2}\sum_{j=1}^{n_m}\frac{1}{n_j}\right)\Sigma_{\mathrm{v}}, \tag{52}$$

$$\mathrm{E}\left(\sum_{j=1}^{n_m}(w_{j.} - \overline{w})^2\right) = (n_m - 1)\,\Sigma_{\mathrm{m}} + \frac{n_m - 1}{n_m}\sum_{j=1}^{n_m}\frac{1}{n_j}\Sigma_{\mathrm{v}}. \tag{53}$$

Using the method of moments, we estimate this quantity with

$$SSM = \sum_{j=1}^{n_m}(w_{j.} - \overline{w})^2. \tag{54}$$

Finally, given an estimate of $\Sigma_{\mathrm{v}}$, we estimate $\Sigma_{\mathrm{m}}$ with

$$\widehat{\Sigma}_{\mathrm{m}} = \frac{1}{n_m - 1}\left(SSM - \frac{n_m - 1}{n_m}\sum_{j=1}^{n_m}\frac{1}{n_j}\Sigma_{\mathrm{v}}\right)_+, \tag{55}$$

where $A_+$ means that negative eigenvalues of $A$ are set to 0. $\widehat{\Sigma}_{\mathrm{m}}$ is a truncation of a quadratic unbiased estimator, very similar to Henderson (1953) or MIVQUE estimators (Rao and Kleffe 1988). A nice property of this approach is that it can be used even if the $w_{j.}$ only are known (e.g. the $w_{jk}$ are not observed). This happens, for instance, if one ensemble of simulations has not been performed, and the response of each model is estimated by subtraction, e.g. $w^{ANT} = w^{ALL} - w^{NAT}$. In such a case, we could assume, for model $j$:

$$w_{j.}^{ANT} = w_{j.}^{ALL} - w_{j.}^{NAT} \sim N\left(0, \Sigma_{\mathrm{m}}^{ANT} + \frac{\Sigma_{\mathrm{v}}}{n_j^{ANT}}\right),$$

where $n_j^{ANT}$ is defined by

$$\frac{1}{n_j^{ANT}} = \frac{1}{n_j^{ALL}} + \frac{1}{n_j^{NAT}}. \tag{56}$$

Finally, under this unbalanced assumption, putting (515253) together with $(\mu - w^*) \sim N(0, \Sigma_{\mathrm{m}})$ suggests considering

$$\Sigma_X = \left(1 + \frac{1}{n_m}\right)\widehat{\Sigma}_{\mathrm{m}} + \frac{1}{n_m^2}\sum_{j=1}^{n_m}\frac{1}{n_j}\Sigma_{\mathrm{v}}. \tag{57}$$

### Application to global mean temperature using the "models are centered on the truth" paradigm

In this section, we present and briefly discuss the results obtained in the analysis of global mean temperature, when the "models are centered on the truth" paradigm is used instead of the "models are statistically indistinguishable from the truth" paradigm, which was used in Sect. 6.2.

Under this paradigm, outputs from each climate model, $w_j$, may be regarded as being sampled independently from

the same distribution, which is centered on the truth, i.e. $w_j \sim N(w, \Sigma_m)$, where $w$ is the true value of the simulated parameter, and $\Sigma_m$ describes the climate modelling uncertainty on this parameter. The distribution of the multi-model mean is then given by $\overline{w} \sim N(w, \Sigma_m/n_m)$. Finally, $\Sigma_X = \Sigma_m/n_m$ has to be considered under this paradigm (if internal variability is neglected).

The primary consequence of considering this alternative paradigm is to narrow the climate modelling uncertainty. Consequently, the assessment of consistency between models and observations is usually more demanding, while more weight is given to models in the estimation of individual forcing contributions.

Under this revised assumption, panel **a)** is unchanged.

In panel **b)**, observations are found to be barely consistent with models ($p$ value 0.09). The expected ALL warming, based on climate models only, is [+0.74K, +0.86K] (90 % confidence interval). This is a much narrower interval than reported in Sect. 6.2 ([+0.44K, +1.16K]). The weak consistency with observations mentioned above is then related to internal variability, which has to be added to these numbers. After inference, the estimated past warming lies between [+0.72K, +0.83K], which is still substantially greater than the observed value of $+0.65K$. This can be understood as follows: because there is little uncertainty in the estimate provided by climate models, the method considers that internal variability is partly responsible for the low observed value. The overall forced change is then estimated to be higher than that found in raw observations.

In panel **c)**, results are similarly impacted. The expected (climate models only) NAT response is unchanged, and the ANT response is expected to lies within [+0.67K, +0.93K], which is again quite a narrow interval. After the inference is performed, the ANT response is estimated to be within [+0.64K, +0.82K]. If compared to the results given in Sect. 6.2, the impact of changing the paradigm is limited. The main impact of changing the paradigm is to discard the lowest values, from +0.55K to +0.64K.

Overall, these results suggest that the "models are statistically indistinguishable from the truth" paradigm, which was used in the main text, is more appropriate to ensure consistency between models and observations, and avoids over emphasizing the climate models outputs.

## References


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

Allen M, Tett S (1999a) Checking for model consistency in optimal fingerprinting. Clim Dyn 15(6):419–434

Allen M, Tett S (1999b) Checking for model consistency in optimal fingerprinting. Clim Dyn 15(6):419–434

Annan J, Hargreaves J (2010) Reliability of the cmip3 ensemble. Geophys Res Lett 37(L02703). doi:10.1029/2009GL041994

Berliner L, Levine R, Shea D (2000) Bayesian climate change assessment. J Clim 13(21):3805–3820

Bindoff N, Stott P, AchutaRao K, Allen M, Gillett N, D Gutzler D, K Hansingo K, Hegerl G, Hu Y, Jain S, Mokhov I, Overland J, Perlwitz J, Sebbari R, Zhang X (2013) Detection and attribution of climate change: from global to regional. In: Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds) Climate change 2013: the physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge

Boucher O, Randall D, Artaxo P, Bretherton C, Feingold G, Forster P, Kerminen V-M, Kondo Y, Liao H, Lohmann U, Rasch P, Satheesh SK, Sherwood S, Stevens B, Zhang XY (2013) Clouds and aerosols. In: Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds) Climate change 2013: the physical science basis. Contribution of Working Group I to the fifth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge and New York

Brohan P, Kennedy J, Harris I, Tett S, Jones P (2006) Uncertainty estimates in regional and global observed temperature changes: a new data set from 1850. J Geophys Res 111:D12106. doi:10.1029/2005JD006548

Dufresne JL, Bony S (2008) An assessment of the primary sources of spread of global warming estimates from coupled atmosphere-ocean models. J Clim 21(19):5135–5144

Fuller WA (1987) Measurement error models. Wiley, Amsterdam

Fyfe J, Gillett N, Zwiers F (2013) Overestimated global warming over the past 20 years. Nat Clim Change 3(9):767–769

Gillett NP, Arora V, Matthews D, Allen M (2013) Constraining the ratio of global warming to cumulative CO2 emissions using CMIP5 simulations. J Clim 26(18):6844–6858

Hall A, Qu X (2006) Using the current seasonal cycle to constrain snow albedo feedback in future climate change. Geophys Res Lett 33:L03502. doi:10.1029/2005GL025127

Hannart A (2015) Integrated optimal fingerprinting: method description and illustration. J Clim 29(6):1977–1998. doi:10.1175/JCLI-D-14-00124.1

Hannart A, Ribes A, Naveau P (2014) Optimal fingerprinting under multiple sources of uncertainty. Geophys Res Lett 41:1261–1268. doi:10.1002/2013GL058653

Hasselmann K (1979) On the signal-to-noise problem in atmospheric response studies. In: Meteorology of Tropical Oceans. Royal Meteorological Society, pp 251–259

Hasselmann K (1993) Optimal fingerprints for the detection of time-dependent climate change. J Clim 6(10):1957–1971

Hasselmann K (1997) Multi-pattern fingerprint method for detection and attribution of climate change. Clim Dyn 13(9):601–611

Hegerl G, Zwiers F (2011) Use of models in detection and attribution of climate change. Wiley Interdiscip Rev Clim Change 2(4):570–591. doi:10.1002/wcc.121

Hegerl G, Hasselmann K, Cubash U, Mitchell J, Roeckner E, Voss R, Waszkewitz J (1997) Multi-fingerprint detection and attribution analysis of greenhouse gas, greenhouse gas-plus-aerosol and solar forced climate change. Clim Dyn 13(9):613–634

Hegerl G, Zwiers F, Braconnot P, Gillet N, Luo Y, Marengo Orsini J, Nicholls N, Penner J, Stott P (2007) Understanding and attributing climate change. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) 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 University Press, Cambridge

Hegerl GC, North GR (1997) Comparison of statistically optimal approaches to detecting anthropogenic climate change. J Clim 10(5):1125–1133

Hegerl GC, Hoegh-Guldberg O, Casassa G, Hoerling M, Kovats R, Parmesan C, Pierce D, Stott P (2010) Good practice guidance paper on detection and attribution related to anthropogenic climate change. In: Stocker TF, Field CB, Qin D, Barros V, Plattner G-K, Tignor M, Midgley PM, Ebi KL (eds) Meeting report of the intergovernmental panel on climate change expert meeting on detection and attribution of anthropogenic climate change IPCC working group I technical support unit. University of Bern, Bern

Henderson CR (1953) Estimation of variance and covariance components. Biometrics 9(2):226–252

Huber M, Knutti R (2012) Anthropogenic and natural warming inferred from changes in earth/'s energy balance. Nat Geosci 5(1):31–36

Huntingford C, Stott P, Allen M, Lambert F (2006) Incorporating model uncertainty into attribution of observed temperature change. Geophys Res Lett 33:L05710. doi:10.1029/2005GL024831

IPCC (2007) Climate change 2007: the physical science basis. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge

IPCC (2013) In: Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds) Climate change 2013: the physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge

Jones G, Stott P (2011) Sensitivity of the attribution of near surface temperature warming to the choice of observational dataset. Geophys Res Lett 38(L21702). doi:10.1029/2011GL049324

Jones GS, Stott P, Christidis N (2013) Attribution of observed historical near surface temperature variations to anthropogenic and natural causes using CMIP5 simulations. J Geophys Res Atmos 118(10):4001–4024

Knutti R, Hegerl G (2008) The equilibrium sensitivity of the earth's temperature to radiation changes. Nat Geosci 1(11):735–743. doi:10.1038/ngeo337

Knutti R, Abramowitz G, Collins M, Eyring V, Gleckler PJ, Hewitson B, Mearns L (2010) Good practice guidance paper on assessing and combining multi model climate projections. In: Stocker TF, Qin D, Plattner G-K, Tignor M, Midgley PM (eds) Meeting report of the intergovernmental panel on climate change expert meeting on assessing and combining multi model climate projections. IPCC working group I technical support unit. University of Bern, Bern, Switzerland

Le Cam L (1990) Maximum likelihood—an introduction. Int Stat Inst Rev 58(2):153–171. doi:10.2307/1403464

Mitchell J, Karoly D, Hegerl G, Zwiers F, Allen M, Marengo J (2001) Detection of climate change and attribution of causes. In: Houghton et al (ed) Climate change 2001: the scientific basis. Contribution of working group I to the third assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge

Morice C, Kennedy J, Rayner N, Jones P (2012) Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: the hadcrut4 data set. J Geophys Res 117(D8). doi:10.1029/2011JD017187

Myhre G, Shindell D, Bréon F-M, Collins W, Fuglestvedt J, Huang J, Koch D, Lamarque J-F, Lee D, Mendoza B et al (2013) Anthropogenic and natural radiative forcing. In: Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds) Climate change 2013: the physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, pp 571–657

Rao CR, Kleffe J (1988) Estimation of variance components and applications. North-Holland, New York

Ribes A, Terray L (2013) Application of regularised optimal fingerprinting to attribution. Part II: Application to global near-surface temperature. Clim Dyn 41(11–12):2837–2853. doi:10.1007/s00382-013-1736-6 on line

Ribes A, Terray L, Planton S (2013) Application of regularised optimal fingerprinting to attribution. Part I: Method, properties and idealised analysis. Clim Dyn 41(11–12):2817–2836. doi:10.1007/s00382-013-1735-7

Rotstayn LD, Collier MA, Shindell DT, Boucher O (2015) Why does aerosol forcing control historical global-mean surface temperature change in CMIP5 models? J Clim 28(17):6608–6625

Santer B, Wigley T, Barnett T, Anyamba E (1995) Detection of climate change and attribution of causes. Cambridge University Press, Cambridge

Santer B, Painter J, Mears C, Doutriaux C, Caldwell P, Arblaster J, Cameron-Smith P, Gillett N, Gleckler P, Lanzante J, Perlwitz J, Solomon S, Stott P, Taylor K, Terray L, Thorne P, Wehner M, Wentz F, Wigley T, Wilcox L, Zou CZ (2013) Identifying human influences on atmospheric temperature. Proc Natl Acad Sci 110(1):26–33. doi:10.1073/pnas.1210514109

Shin SIS, Sardeshmukh D (2011) Critical influence of the pattern of tropical ocean warming on remote climate trends. Clim Dyn 36(7–8):1577–1591

Shiogama H, Stone D, Nagashima T, Nozawa T, Emori S (2013) On the linear additivity of climate forcing-response relationships at global and continental scales. Int J Climatol 33(11):2542–2550. doi:10.1002/joc.3607

Stevens B, Bony S (2013) What are climate models missing. Science 340(6136):1053–1054

Stott P, Mitchell J, Allen M, Delworth D, Gregory J, Meehl G, Santer B (2006) Observational constraints on past attributable warming and predictions of future global warming. J Clim 19(13):3055–3069

Taylor K, Stouffer R, Meehl G (2012) An overview of CMIP5 and the experiment design. Bull Am Meteorol Soc 93(4):485–498. doi:10.1175/BAMS-D-11-00094.1

Terray L, Corre L, Cravatte S, Delcroix T, Reverdin G, Ribes A (2011) Near-surface salinity as nature's rain gauge to detect human influence on the tropical water cycle. J Clim. doi:10.1175/JCLI-D-10-05025.1

van Oldenborgh G, Doblas Reyes F, Hawkins E (2013) Reliability of regional climate model trends. Environ Res Lett 8(1):014,055. doi:10.1088/1748-9326/8/1/014055