What Surface Observations Are Important for Separating the Influences of Anthropogenic Aerosols from Other Forcings?

XIAOQIN YAN AND TIMOTHY DELSOLE

George Mason University, Fairfax, Virginia, and Center for Ocean–Land–Atmosphere Studies, Calverton, Maryland

MICHAEL K. TIPPETT

Department of Applied Physics and Applied Mathematics, Columbia University, New York, New York, and Center of Excellence for Climate Change Research, Department of Meteorology, King Abdulaziz University, Jeddah, Saudi Arabia

ABSTRACT

This paper shows that joint temperature–precipitation information over a global domain provides a more accurate estimate of aerosol forced responses in climate models than does any other combination of temperature, precipitation, or sea level pressure. This fact is demonstrated using a new quantity called potential detectability, which measures the extent to which a forced response can be detected in a model. In particular, this measure can be evaluated independently of observations and therefore permits efficient exploration of a large number of variable combinations before performing optimal fingerprinting on observations. This paper also shows that the response to anthropogenic aerosol forcing can be separated from that of other forcings using only spatial structure alone, leaving the time variation of the response to be inferred from data, thereby demonstrating that temporal information is not necessary for detection. The spatial structure of the forced response is derived by maximizing the signal-to-noise ratio. For single variables, the north–south hemispheric gradient and equator-to-pole latitudinal gradient are important spatial structures for detecting anthropogenic aerosols in some models but not all. Sea level pressure is not an independent detection variable because it is derived partly from surface temperature. In no case does sea level pressure significantly enhance potential detectability beyond that already possible using surface temperature. Including seasonal or land–sea contrast information does not significantly enhance detectability of anthropogenic aerosol responses relative to annual means over global domains.

1. Introduction

Numerous studies suggest that radiative forcing from anthropogenic aerosols (AA) plays an important role in climate variability [see Boucher et al. (2013) for a comprehensive review]. In addition to global climate change, anthropogenic aerosols also have been implicated in multidecadal variability of the North Atlantic (Booth et al. 2012), precipitation changes over the Sahel and South and East Asia (Shindell et al. 2012), warming patterns in the Indian Ocean (Dong and Zhou 2014),

southward shifts in the ITCZ (Rotstayn and Lohmann 2002), reduction of rainfall in the tropics (Ridley et al. 2015), the Pacific decadal oscillation and expansion and contraction of the tropical belt (Allen et al. 2014), and tropical Atlantic rainfall (Chang et al. 2011).

Unfortunately, the radiative forcing due to AA is highly uncertain because of uncertainties in emission levels, size distribution, optical properties, and the interaction of aerosols with clouds (Boucher et al. 2013). Current estimates of global radiative forcing suggest that the forcing due to AA is more uncertain than that due to well-mixed greenhouse gases (GHG) and hence dominates most of the uncertainty in total forcing (Myhre et al. 2013, their Fig. 8.16). Also, aerosol cooling partially cancels greenhouse warming. Thus, climate simulations could match past climate change through

Corresponding author address: Xiaoqin Yan, 132 Research Hall, Mail Stop 2B3, George Mason University, 4400 University Drive, Fairfax, VA 22030.
E-mail: xyan@masonlive.gmu.edu

fortuitous cancellation of errors, allowing models with very different responses to individual forcings to be equally consistent with observations.

In principle, aerosol-forced climate changes can be estimated directly from observations using energy balance constraints, provided that such changes can be distinguished from those by other forcings. A major difference in forcing is that greenhouse gases tend to be well mixed and have relatively small spatial gradients, whereas anthropogenic aerosols are short lived and tend to be more concentrated around their (land) sources. These differences in forcing lead to differences in hemispheric gradients and land–sea gradients in the temperature and precipitation responses. Forcings also differ temporally: forcing from well-mixed greenhouse gases has increased monotonically over the past century, while forcing from aerosols has leveled in the last few decades. In addition, both forcings have distinct seasonal cycles. Despite these forcing differences, the associated responses tend to be similar to each other (Xie et al. 2013), leading to collinearity problems in separating these responses in data.

The most widely used method for estimating different forced responses from observations is optimal fingerprinting. Stott et al. (2006) used optimal fingerprinting to rescale model responses to match observations and thereby inferred changes due to aerosol and greenhouse gas forcing. Features of the response that appeared to provide the most discrimination between greenhouse warming and aerosol cooling were the differential warming rates between the hemispheres, between land and ocean, and between mid- and low latitudes. For example, Stott et al. (2006) showed that removing one or more of these features from the response vector increases uncertainty.

The purpose of this paper is to quantify the effectiveness of different variables for detecting the response to aerosol forcing and thereby identify the best surface variables for detecting aerosol cooling. A key quantity in our methodology is *potential* detectability, which measures the detectability of a forced response in a model. For a single forcing, detectability is measured by the total-to-noise ratio, defined as the ratio of the total variance (forced plus unforced) to the variance of internal variability. For two forcings, detectability is measured by the product of the total-to-noise ratio of the given forcing times a measure of multicollinearity between the two forcings (as we will show). Both quantities can be estimated from single-forcing simulations; hence, potential detectability can be estimated as soon as the single-forced simulations are available.

To ensure that the more effective variables for representing a forced response are used, we select the variables that maximize the total-to-noise ratio (or equivalently, the signal-to-noise ratio plus one). These variables maximally discriminate between a climate change signal and internal variability and can be identified using the method of Jia and DelSole (2012) (discussed shortly). However, these variables may not be the best variables to distinguish different forcings. For instance, two forcings may produce highly detectable but virtually identical responses in a given set of variables, in which case those variables could not be used to distinguish the forced responses. Our goal is to find the best variables for separating forced responses from each other and from internal variability.

The relative importance of spatial versus temporal structure for separating different forced responses is unclear. Part of the reason for this is that most optimal fingerprinting studies use predictors that contain both spatial and temporal structure, leaving the individual contributions unclear. The extent to which different forced responses can be separated based on time series of global mean quantities has been investigated extensively (Stern and Kaufmann 2000; Lean and Rind 2009; Zhou and Tung 2013). In this paper, we investigate the extent to which forced responses can be separated based on spatial structure alone. Thus, our approach to optimal fingerprinting differs from previous approaches in that we do not use the full space–time evolution of the climate change signal as a predictor. Instead, we use only the spatial structure of a variable, as well as its covariability with other variables, as input to fingerprinting. The output is an estimate of the time histories of the forced responses. This approach avoids making the (strong) assumption that the climate model correctly simulates the spatial–temporal evolution of the forced response.

This paper will show that the more accurate estimates of the response to aerosol forcing in model simulations are obtained using two variables rather than a single variable such as surface temperature alone. This conclusion is consistent with the fact that the joint behavior of two variables is more informative than their separate behavior. Moreover, previous studies also have found that combining different variables can improve inferences, including using joint changes of temperature at the surface and aloft (Jones et al. 2003), joint changes of temperature, diurnal temperature range and precipitation (Schnur and Hasselmann 2005), joint changes of temperature and salinity (Stott and Jones 2009; Pierce et al. 2012), joint changes of maximum amount and location of zonal mean precipitation (Marvel and Bonfils 2013), and joint changes on river flows, winter air temperature, and snowpack over the western United States (Barnett et al. 2008). While these studies illustrate the

success that can be achieved using multiple variables, it is not always clear the extent to which multiple variables were required.

To avoid complications due to model error and observation error, we perform our analysis on model simulations instead of real observations. Our methodology for separating the response time series is described in section 2 and then applied to model simulations, as described in sections 3 and 4. We conclude with a summary and discussion of our results.

2. Methodology

As in previous studies (Bindoff et al. 2013 and references therein), we assume that a climate variable can be modeled as a linear combination of responses to external forcing, plus noise. Previous studies have shown that the linearity assumption is a good assumption for both temperature and precipitation for annual mean indices over the global domain (Marvel et al. 2015) and for 5-yr-mean indices over continental scale (Shiogama et al. 2013). We also performed more comprehensive linearity tests based on discriminant analysis methods, but these calculations are not discussed here because they merely confirm previous studies.

Let the observed change at the nth time step and sth spatial location be denoted $o_{n,s}$. Thus, $\mathbf{O}$ is an $N \times S$ matrix for a total of N time steps and S spatial locations. Observed changes $\mathbf{o}$ are modeled as a linear sum of simulated changes from forcings $\mathbf{f}_1, \ldots, \mathbf{f}_M$ plus random internal variability $\boldsymbol{\eta}$:

$$\mathbf{o} = \sum_{m=1}^{M} \mathbf{f}_m a_m + \boldsymbol{\eta}, \tag{1}$$

where $a_1, \ldots, a_M$ are scaling factors to account for amplitude errors. More sophisticated approaches account for uncertainty in the forced response vectors, but the above model is adequate for our purposes. Internal variability $\boldsymbol{\eta}$ is assumed to be white noise such that each time step is drawn independently from a Gaussian distribution with zero mean and S-dimensional covariance matrix $\boldsymbol{\Sigma}_\eta$. One might question whether the (temporal) white noise assumption for internal variability is appropriate. We have found that the white noise assumption gives reasonable estimates of uncertainty, in the sense that the results derived from separate ensemble members frequently lie (at the expected rate) within the confidence interval derived from any one of the members. This consistency suggests that the above model (including the white noise assumption) is reasonable.

The standard approach to estimating the forced response vector $\mathbf{f}_m$ is to compute an average over an ensemble of simulations driven by the mth forcing. We have obtained more accurate results by estimating the forced response by maximizing the signal-to-noise (S/N) ratio (i.e., maximizing the ratio of the variance of the forced response to the variance of internal variability). These vectors can be obtained from the method of Jia and DelSole (2012), which decomposes the forced response into a sum of components ordered by S/N ratio, where each component is a fixed pattern multiplied by an associated time series (similar to the way principal component analysis decomposes data by variance). In all cases, we find that only one component has a statistically significant S/N ratio, implying that other forced patterns, if they exist, are not detectable after the most detectable pattern has been removed. Consequently, the forced response pattern is represented as follows:

$$\mathbf{f}_m = \mathbf{r}_m \mathbf{p}_m^{\mathrm{T}}, \tag{2}$$

where $\mathbf{r}_m$ is an N-dimensional vector specifying the response time series, $\mathbf{p}_m$ is an S-dimensional vector specifying the spatial pattern, and superscript T denotes the matrix transpose. Jia and DelSole (2012) show that the pattern $\mathbf{p}_m$ maximizing the S/N ratio satisfies the eigenvalue problem

$$\boldsymbol{\Sigma}_m \boldsymbol{\Sigma}_\eta^{-1} \mathbf{p}_m = \lambda \mathbf{p}_m, \tag{3}$$

where $\boldsymbol{\Sigma}_m$ is the covariance matrix for the single-forcing simulation for the mth forcing and anomalies are measured relative to the control mean.

Because the forced response $\mathbf{f}_m$ is separable in space and time [see (2)], we can modify optimal fingerprinting to estimate the time series $\mathbf{r}_m$ (which subsumes the scaling factor a_m). This approach assumes that the climate model correctly simulates only the spatial structure of the response and does not impose any constraints on the time variability. Unfortunately, the uncertainties become unacceptably large when time series for two or more forcings are estimated for each year separately (e.g., confidence intervals for the scaling factors include 0, implying that the forced response cannot be detected). Consequently, we assume that the response evolves slowly in time, which allows the estimation method to pool data across years to reduce uncertainty. Specifically, the temporal variation is represented by a Jth-order polynomial in time:

$$r_m(t) = \sum_{j=1}^{J} c_{m,j} L_j(t), \tag{4}$$

where $L_j(t)$ is the jth Legendre polynomial in time t. This formulation requires estimating J coefficients per forcing.
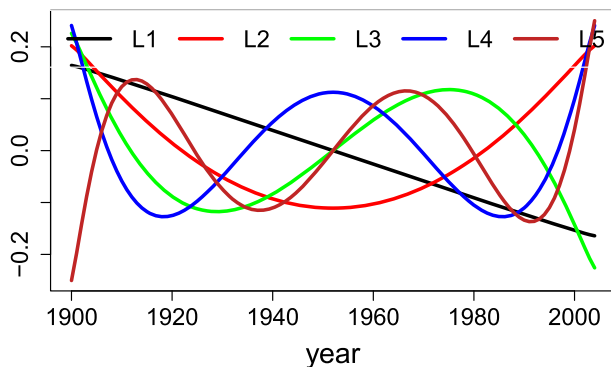
FIG. 1. Temporal evolution of Legendre polynomials 1–5.

Legendre polynomials 1–5 are shown in Fig. 1. The polynomial representation is most questionable during volcanic eruptions, but since the dominant response to volcanic eruptions spans only a few years (Iles et al. 2013), smoothing over periods of major eruptions leads to only minor errors on multidecadal time scales.

The above regression model assumes that the spatial structure of the forced response $\mathbf{p}_m$ is known but leaves the associated time series $\mathbf{r}_m$ (or more precisely, the coefficients of the Legendre polynomials) to be inferred from data. This reformulated model also can be expressed in the form of optimal fingerprinting; hence, solutions to this formula can be estimated similarly as the standard solutions to the optimal fingerprinting. Detail derivation of solutions of the polynomial coefficients and their uncertainty are given in the appendix.

The mth forced response is said to be detected if one can reject the null hypothesis H_m that all polynomial coefficients for the mth forced response vanish:

$$H_m: c_{m,1} = c_{m,2} = \cdots = c_{m,J} = 0. \qquad (5)$$

If only one forced response exists (i.e., $M = 1$), then the statistic for testing H_m is

$$\text{TNR}_m = \left(\sum_{j=1}^{J} \hat{c}_{m,j}^2 \right) (\mathbf{p}_m^{\text{T}} \Sigma_\eta^{-1} \mathbf{p}_m). \qquad (6)$$

The appendix shows that under the null hypothesis H_m, the statistic TNR_m has a chi-squared distribution with J degrees of freedom. This fact provides the basis for testing the null hypothesis at a prescribed significance level. The statistic TNR_m is proportional to the total-to-noise ratio, where "total" refers to forced and unforced variability and "noise" refers to unforced variability.

We are particularly interested in identifying the response to anthropogenic aerosols. Therefore, two response vectors will be considered: the response to anthropogenic

aerosols, denoted $\mathbf{p}_{\text{AA}}$, and the response to all forcings except AA, denoted $\mathbf{p}_{\text{noAA}}$. As shown in the appendix, the statistic for testing the null hypothesis H_m in the two-pattern detection problem is

$$\phi_m^2 = \text{TNR}_m (1 - \rho^2), \qquad (7)$$

where TNR_m is the total-to-noise ratio in (6) except using coefficients $\hat{c}_{m,j}$ derived from the two-pattern regression model, and the parameter ρ is defined as follows:

$$\rho = \frac{\mathbf{p}_{\text{AA}}^{\text{T}} \Sigma_\eta^{-1} \mathbf{p}_{\text{noAA}}}{\sqrt{\mathbf{p}_{\text{AA}}^{\text{T}} \Sigma_\eta^{-1} \mathbf{p}_{\text{AA}}} \sqrt{\mathbf{p}_{\text{noAA}}^{\text{T}} \Sigma_\eta^{-1} \mathbf{p}_{\text{noAA}}}}. \qquad (8)$$

The parameter ρ measures the cosine of the angle between the response vectors in "whitened space" (DelSole and Tippett 2007) and can be interpreted as a generalized spatial correlation. The squared pattern correlation ρ^2 is always between 0 and 1, where one implies that the response vectors are perfectly collinear while zero implies that the response vectors are orthogonal. Equation (7) shows that detectability decreases as the response vectors become more collinear.

The total-to-noise ratio TNR_m and pattern correlation ρ can be computed immediately after the response vectors have been derived from the single-forcing runs. Accordingly, we define potential detectability to be (7) after substituting TNR_m from single-forcing runs and realized detectability to be (7) after substituting TNR_m derived from coefficient estimates $\hat{c}_{m,j}$ from historical simulations. It follows that potential detectability can be computed directly from the single-forcing runs and does not require simulations with multiple forcings. Note that the method of Jia and DelSole (2012) effectively maximizes TNR_m for the single-forcing case.

The noise covariance matrix Σ_η can be estimated from control runs, but the result is singular (i.e., Σ_η has no matrix inverse) because the state space dimension exceeds the sample size by orders of magnitude. The most common approach to regularizing the estimation problem is to project the data onto the leading empirical orthogonal functions (EOFs) of internal variability (i.e., to represent the data in a reduced EOF space). Unfortunately, EOFs are data dependent and lead to difficulties in interpretation and statistical overfitting (Lawley 1956; Allen and Tett 1999; Jones et al. 2013; DelSole and Tippett 2015). In this paper, we use a different basis set to avoid these problems—namely, the eigenvectors of the Laplacian operator. These eigenvectors depend only on the domain and hence are data independent, and the leading eigenvectors often have

Spherical Harmonic 2

Spherical Harmonic 3

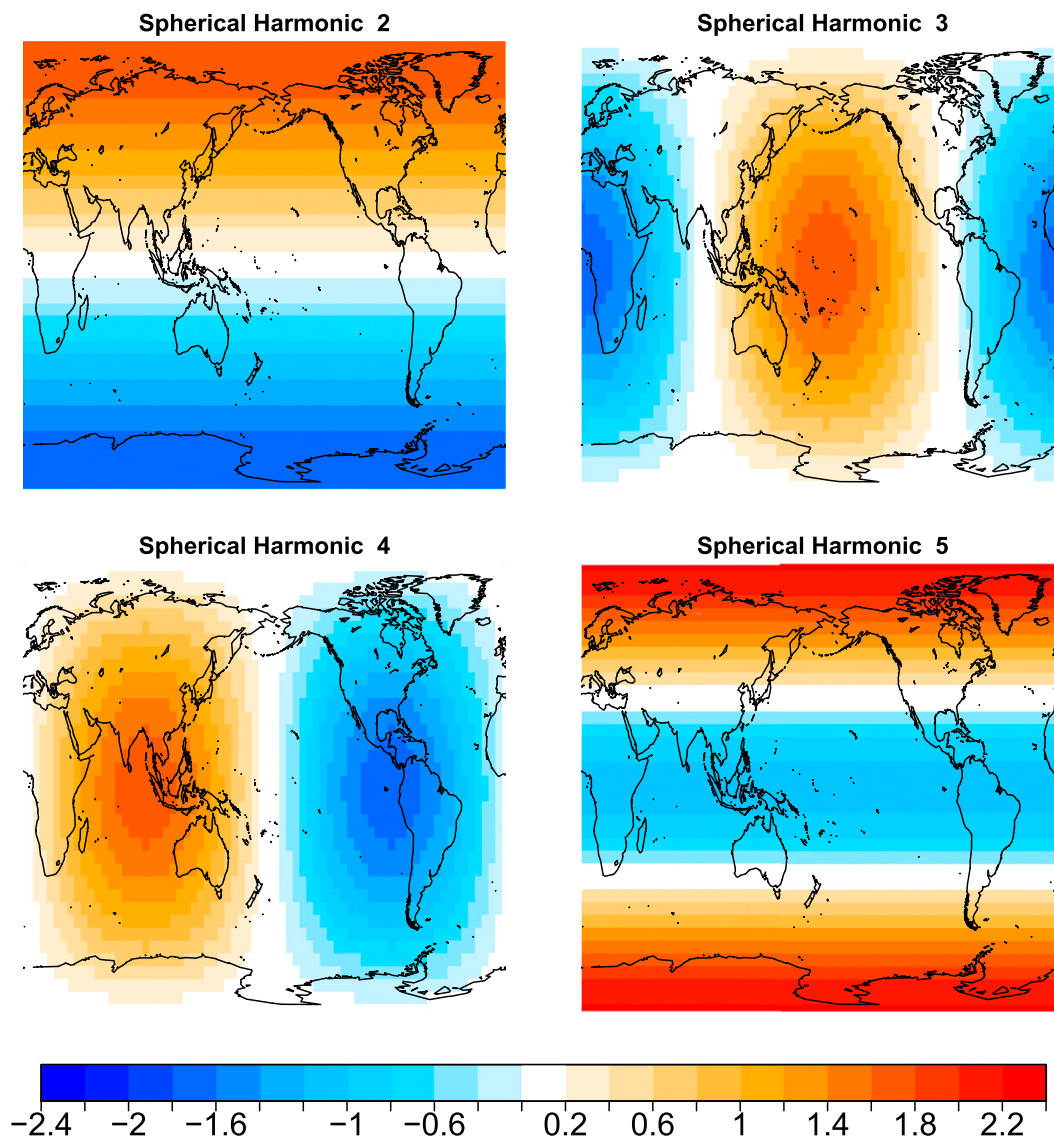## Spherical Harmonic 4

Spherical Harmonic 5

FIG. 2. Spatial patterns of spherical harmonics 2–5.

simple physical interpretation. The eigenfunctions can be ordered by decreasing spatial scale. For a global domain, Laplacian eigenfunctions are equivalent to spherical harmonics and illustrated in Fig. 2. The first eigenfunction, not shown, is merely a constant and corresponds to the global mean. The second eigenfunction measures the gradient between the Northern and Southern Hemispheres, the next two measure the east–west gradients, and the fifth measures the pole-to-equator gradient. For restricted domains, such as ocean only or land only, the Laplacian eigenfunctions are computed by the method discussed in DelSole and Tippett (2015), which is based on a Green's function method. For these restricted domains, the first Laplacian eigenfunction still corresponds to a spatially uniform

pattern over the domain, while the subsequent eigenvectors correspond to dipoles, tripoles, quadrapoles, and so forth of decreasing length scale.

In this paper, we consider detectability of not only a single variable like temperature but also combinations of variables like temperature and precipitation. Combinations of variables present no special difficulty in the above methodology. In particular, the above methodology does not involve computing EOFs, and the signal-to-noise maximization procedure is invariant to linear transformation, so the fact that the state vector contains different variables with different units is not problematic. As an example, suppose we want to detect AA-forced responses based on the first five Laplacian eigenvectors of temperature and the first five Laplacian eigenvectors of

precipitation. To do this, we construct a state vector in which the first five elements give the instantaneous amplitude of the first five Laplacian eigenvectors for temperature, and the second five elements give the instantaneous amplitude of the first five Laplacian eigenvectors for precipitation. The resulting vector is 10-dimensional and evolves in time for N time steps, giving a $10 \times N$ data matrix. This data matrix can be analyzed just like a single-variable data matrix, so the method of Jia and DelSole (2012) can be applied to AA simulations to determine the 10-dimensional "response vector" $\mathbf{p}_{AA}$ that maximizes detectability of AA forced responses. The elements of $\mathbf{p}_{AA}$ give the amplitudes of the first five Laplacian eigenvectors of temperature and of precipitation for detecting AA forced response. Multiplying these amplitudes by the appropriate Laplacian eigenvectors produces two spatial patterns, one for temperature and one for precipitation, that characterize the joint temperature–precipitation pattern for detecting AA forced response. The same procedure can be applied to simulations without AA forcing (noAA) to determine the response vector $\mathbf{p}_{noAA}$. Then, optimal fingerprinting can proceed just as in the single-variable case. The parameter ρ still measures the degree of multicollinearity but represents a generalized measure of pattern correlation involving temperature and precipitation.

3. Data

We apply the above methodology to simulations from phase 5 of the Coupled Model Intercomparison Project (CMIP5; Taylor et al. 2012). Estimates of internal variability are derived from preindustrial control runs, which have no year-to-year variation in external forcing. The response to particular forcings are estimated from simulations that contain only anthropogenic aerosol forcing, called AA simulations, and simulations that contain all forcings except anthropogenic aerosols, called noAA simulations. In place of observations, we use simulations of the recent past (1900–2004) driven by anthropogenic and natural forcing, called historical simulations. Only two models, CSIRO Mk3.6.0 (CSIRO) and IPSL-CM5A-LR (IPSL), have both noAA and AA simulations. Thus, only these models are considered in this study. Both the CSIRO and IPSL models use the same aerosol emission inventory (Lamarque et al. 2010) over the historical period from 1850 to 2000 in decadal increments at a horizontal resolution of 0.5° in latitude and longitude. CSIRO scaled all BC emissions by 1.25 and all OA emissions by 1.5 (Rotstayn et al. 2012). Both models include aerosol indirect effects and represent a wide range of aerosol sensitivity (Gettelman et al. 2015). We analyze annual means of three commonly used

variables for climate detection: 2-m air temperature (TAS), precipitation (PR), and sea level pressure (hereinafter PSL).

4. Results

a. Detection and attribution based on temperature and precipitation

The detection procedure described in section 2 differs from that of previous studies in several ways (e.g., the data are represented in terms of Laplacian eigenvectors instead of EOFs, the response patterns are obtained by maximizing detectability, and temporal smoothing is based on Legendre polynomials instead of decadal averages). Therefore, it may be instructive to illustrate the procedure by analyzing a single case in detail. Accordingly, we perform a detection analysis of AA and noAA forced responses using joint temperature and precipitation data. The detection analysis is performed on a single ensemble from a historical simulation, which serves as a surrogate for observations, thereby avoiding complications due to model error and observational error. There are 10 such ensemble members for CSIRO and 6 such ensemble members for IPSL. We first consider a state vector based on the first Laplacian eigenvector (i.e., the spatially uniform pattern) for temperature and precipitation. At one time step, this vector contains only two elements: global average temperature and precipitation. The response vectors $\mathbf{p}_{AA}$ and $\mathbf{p}_{noAA}$ that maximize detectability with respect to internal noise are determined. DelSole et al. (2015, manuscript submitted to *J. Climate*) examine this case in detail and show that the response vector effectively measures hydrological sensitivity. These two-dimensional response vectors are then used in optimal fingerprinting to estimate the respective time series from a single historical simulation. For comparison, we also consider response vectors using 20 spherical harmonics for temperature and 20 spherical harmonics for precipitation.

The estimated time series of global mean temperature change using five Legendre polynomials are shown in Fig. 3. Different curves of the same color show results for different ensemble members of the historical run (10 for CSIRO and 6 for IPSL). Figure 3 also shows global mean temperature change estimated from the AA and noAA simulations separately without temporal smoothing (A34). The top and bottom rows of Fig. 3 show results derived based on 1 and 20 spherical harmonics, respectively, while the left and right columns show results derived from CSIRO and IPSL, respectively. As is evident from the figure, the global mean temperature changes for AA and noAA are estimated very well in this
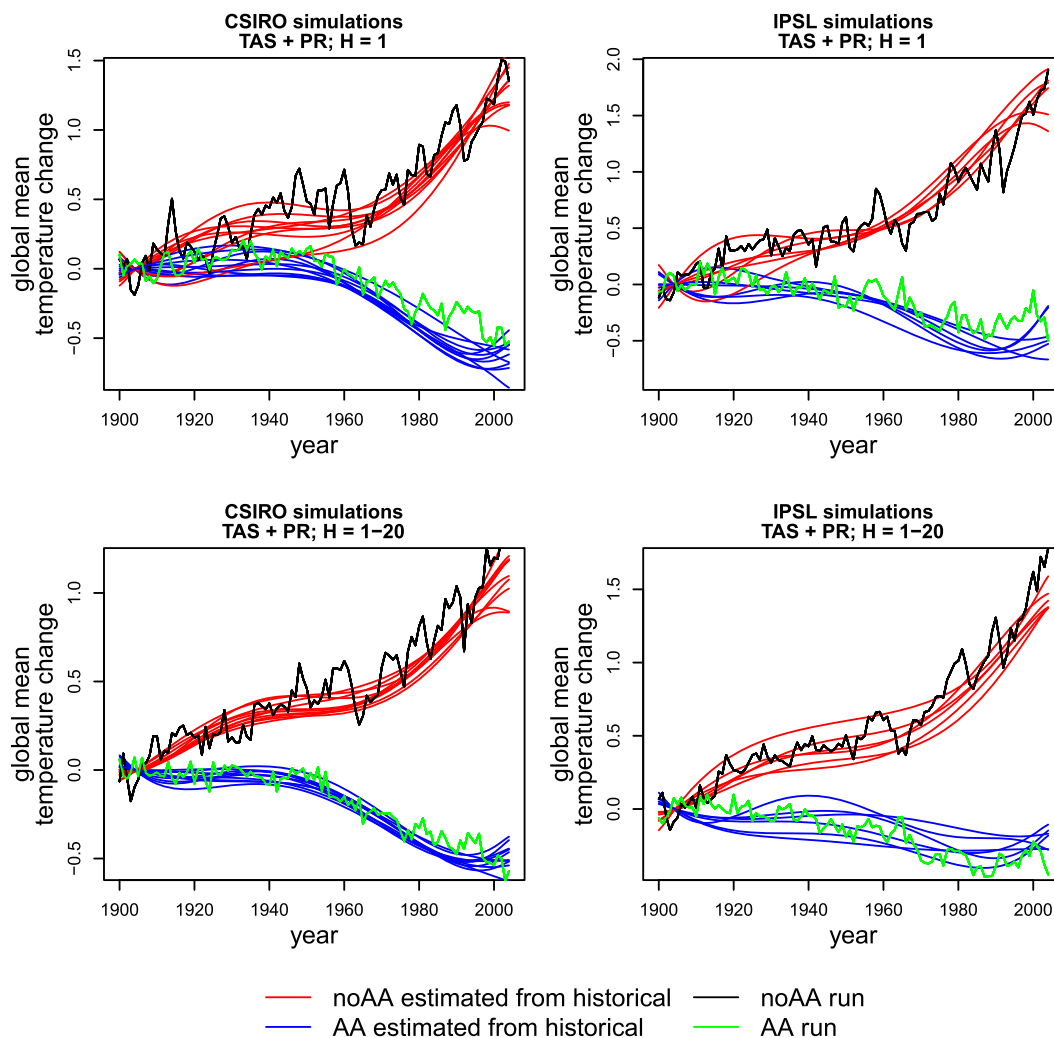
FIG. 3. Time series of global mean temperature change calculated from historical simulations (blue and red) and from single-forcing AA runs (green) and noAA runs (black), using the joint annual mean temperature–precipitation response vector. The left and right columns show results for CSIRO and IPSL, respectively, while the top and bottom rows show results for 1 and 20 spherical harmonics, respectively. Time series from historical simulations are based on using five Legendre polynomials, whereas time series based on individual forcing runs are based on year-by-year estimation. The response vector is derived from only one ensemble member of the individual forced runs and then used to estimate global mean temperature change time series for each ensemble member from the historical run.

"perfect model" case. In fact, the time series of global mean temperature change estimated from historical runs look essentially like smoothed versions of those obtained from the AA and noAA runs, even though the two time series were estimated independently. Moreover, global mean temperature change time series for different ensemble members are relatively close to each other, indicating that the estimates are not sensitive to sampling errors.

The impact of adding spatial gradient information in the forced response patterns is shown in the bottom row of Fig. 3. Based on 20 harmonics for temperature and 20 harmonics for precipitation, the time series estimated

from the CSIRO historical run (Fig. 3, bottom left) are generally closer together (hence less sensitive to internal variability) and closer to the single-forcing runs (hence more accurate). In contrast, time series for IPSL have changed relatively little, suggesting that detectability is not strongly improved by adding spatial gradient information. Thus, the importance of spatial gradient information for separating forced responses is model dependent.

So far in this paper, optimal fingerprinting has been applied in the context of a perfect model world, in the sense that the same model generates the data and generates the forced response used in fingerprinting. In
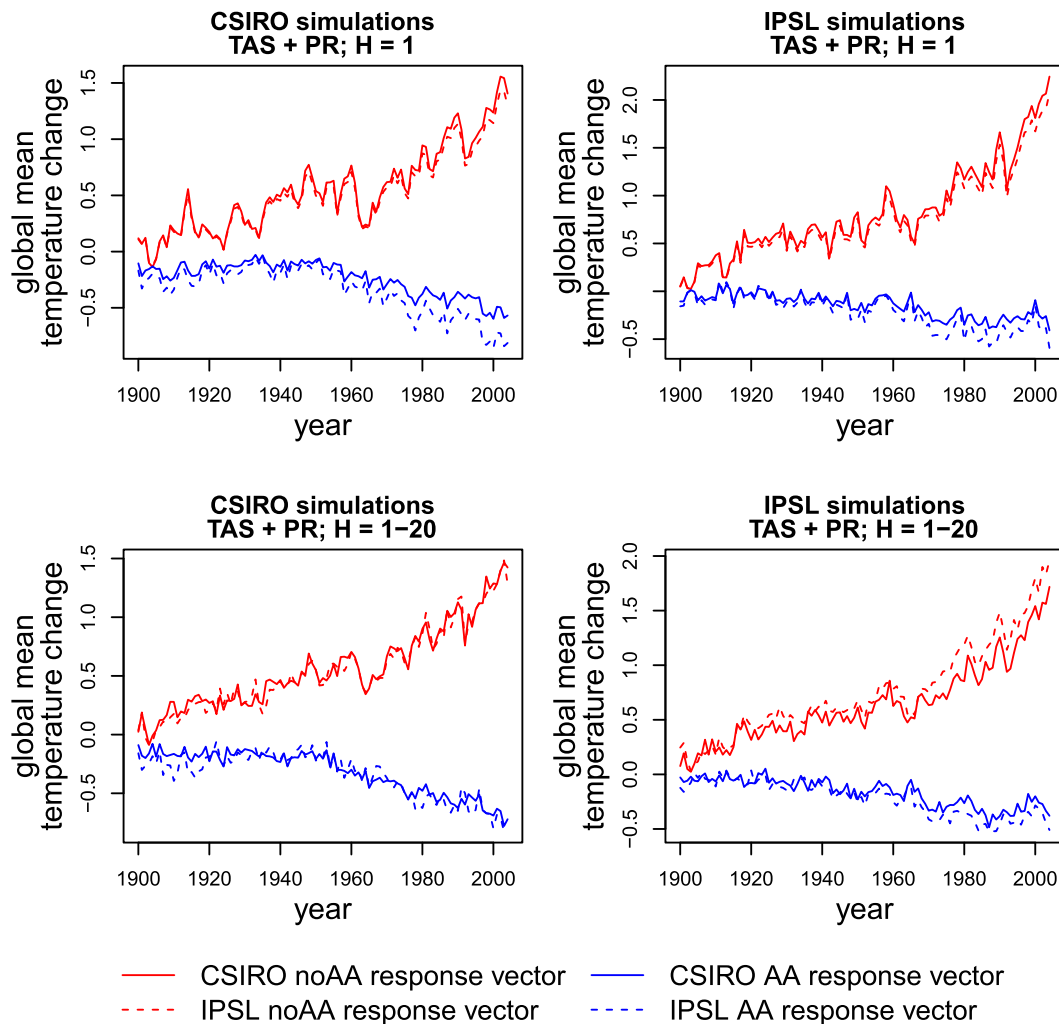
FIG. 4. Time series of global mean temperature change obtained from single-forcing AA runs (blue) and noAA runs (red), using the joint annual mean temperature–precipitation response vector derived from the same model (solid) or a different model (dashed). The left and right columns show results for CSIRO and IPSL, respectively, while the top and bottom rows show results for 1 and 20 spherical harmonics, respectively.

practice, the forced response vectors used in finger-printing differ from the "true" responses that influence observations. To assess model error, one often examines the sensitivity of estimated responses to response vectors derived from different models. Accordingly, we show in Fig. 4 the change in global mean temperature in single-forcing simulations attributable to AA and noAA forced response vectors derived from different models. The top and bottom rows in Fig. 4 show results based on 1 and 20 spherical harmonics, respectively, while the left and right columns show results derived from CSIRO and IPSL single-forcing runs, respectively. Figure 4 shows that the change in global mean temperature is relatively insensitive to the model response vector. Thus, while response vectors differ between models, these differences do not lead to significantly

different conclusions about climate changes attribut-able to AA and noAA.

The response patterns $\mathbf{p}_{AA}$ and $\mathbf{p}_{noAA}$ used to derive time series in the bottom panels of Figs. 3 and 4 are shown in Fig. 5. Recall that this pattern is the only component with statistically significant signal-to-noise ratio. Because the state vector depends on the joint temperature–precipitation fields, the response pattern is characterized by a pair of fields, shown along each row. Multiplying these patterns by their respective time series gives the spatiotemporal response to AA and to noAA forcing in each model. The models agree that noAA forcing tends to warm the globe and AA forcing tends to cool the globe. However, CSIRO suggests that the noAA warming is approximately equally distributed between the Northern and Southern Hemispheres, whereas IPSL suggests that
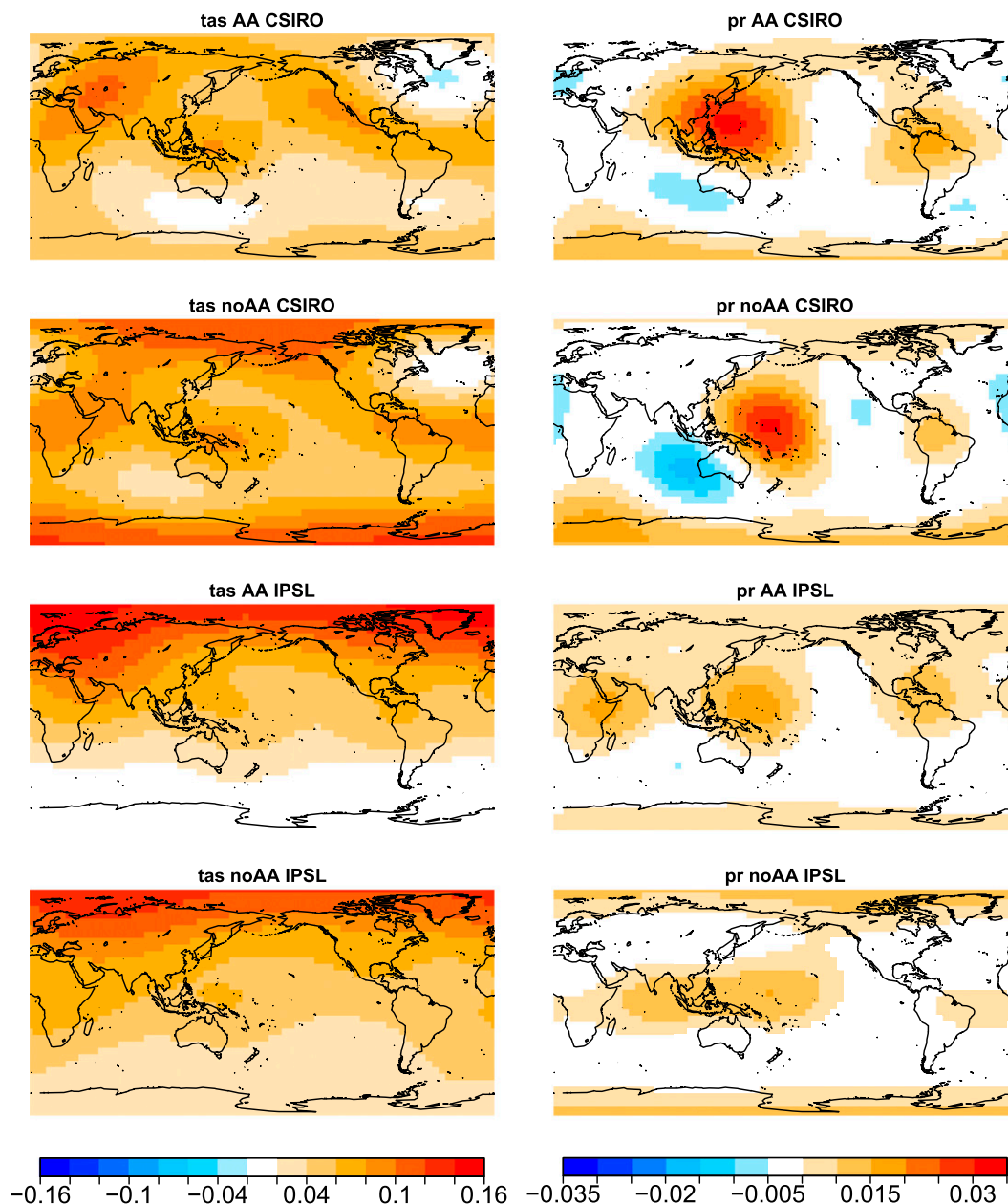
FIG. 5. Response vectors that maximize detectability of AA forcing and noAA forcing in two different models (CSIRO and IPSL) based on 20 Laplacian eigenvectors and annual mean fields. The response vector depends on a joint temperature–precipitation relation and therefore is characterized by a pair of patterns, shown in each row: (left) temperature fields and (right) precipitation fields. The corresponding forcing and model is indicated in the title of each figure. These response patterns were used to estimate the time series shown in Fig. 4.

noAA warming is dominated in the Northern Hemisphere. The models also agree that noAA will tend to enhance precipitation and AA will tend to decrease precipitation over the western Pacific, near the Maritime Continent, and over the Atlantic and Pacific centered near Central America. However, the models disagree in the direction of precipitation changes over the Indian Ocean. The fact that

the AA and noAA response patterns are similar implies that the patterns have a large squared pattern correlation ρ^2. Despite this large pattern correlation, the total-to-noise ratio is sufficiently strong to overcome multicollinearity problems to produce stable time series estimates.

Confidence intervals for the polynomial coefficients (in the perfect model case) are shown in Fig. 6. A forced
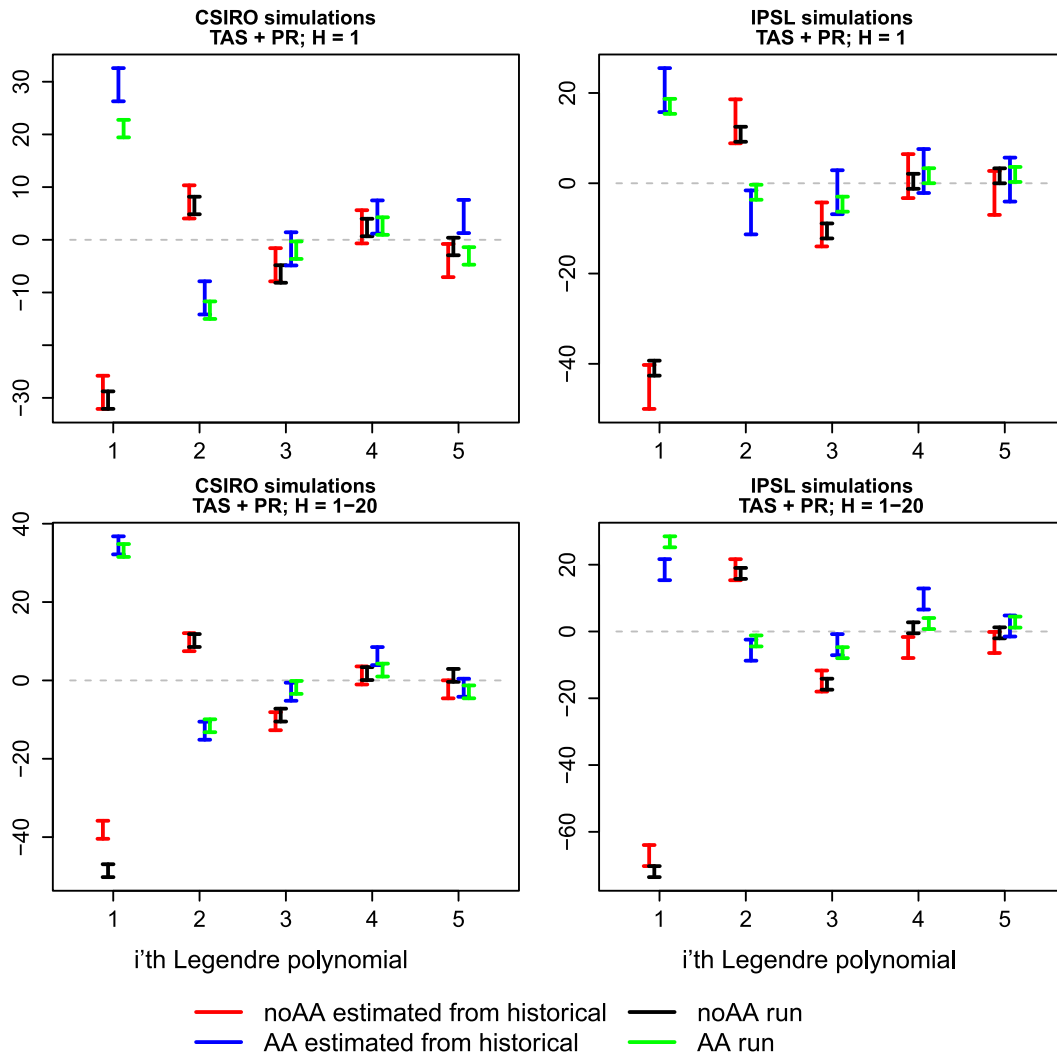
FIG. 6. The 2.5% and 97.5% confidence limits for the coefficients of the five Legendre polynomials (in time) estimated from (left) CSIRO and (right) IPSL, using (top) 1 Laplacian eigenvector and (bottom) 20 Laplacian eigenvectors. The dashed line indicates zero. Only the perfect model case, in which all parameters and data come from the same model, is shown.

response is said to be detected when the interval derived from the historical run does not include zero and is said to be attributable to AA or noAA forcing when the interval from the historical run overlaps with the appropriate interval from the single-forcing run. Almost all forced responses are attributable to the appropriate forcing as expected because the data come from a perfect model experiment. For both CSIRO and IPSL, the linear trend and the quadratic growth are detected (i.e., coefficients associated with the first two Legendre polynomials differ significantly from zero). Based on all variable combinations that we have analyzed, adding more Laplacian eigenvectors tends to improve the ability to detect and attribute changes in higher-order polynomials in time. However, as the order of the

polynomial increases, so too does the number of fitted parameters and the likelihood of overfitting. We have found that 20 Laplacian eigenvectors produces stable results in a perfect model scenario. Since multidecadal variability is associated with polynomials of degree 3–6, these results suggest that detecting and attributing multidecadal variability to forcings in a perfect model requires spatial gradient information.

b. Detectability under different combinations of variables, spatial structures, and seasons

We now explore detectability based on various combinations of variables. To summarize the results, we show the potential detectability measure ϕ_m^2 in (7) using the total-to-noise ratios derived from the single-forcing
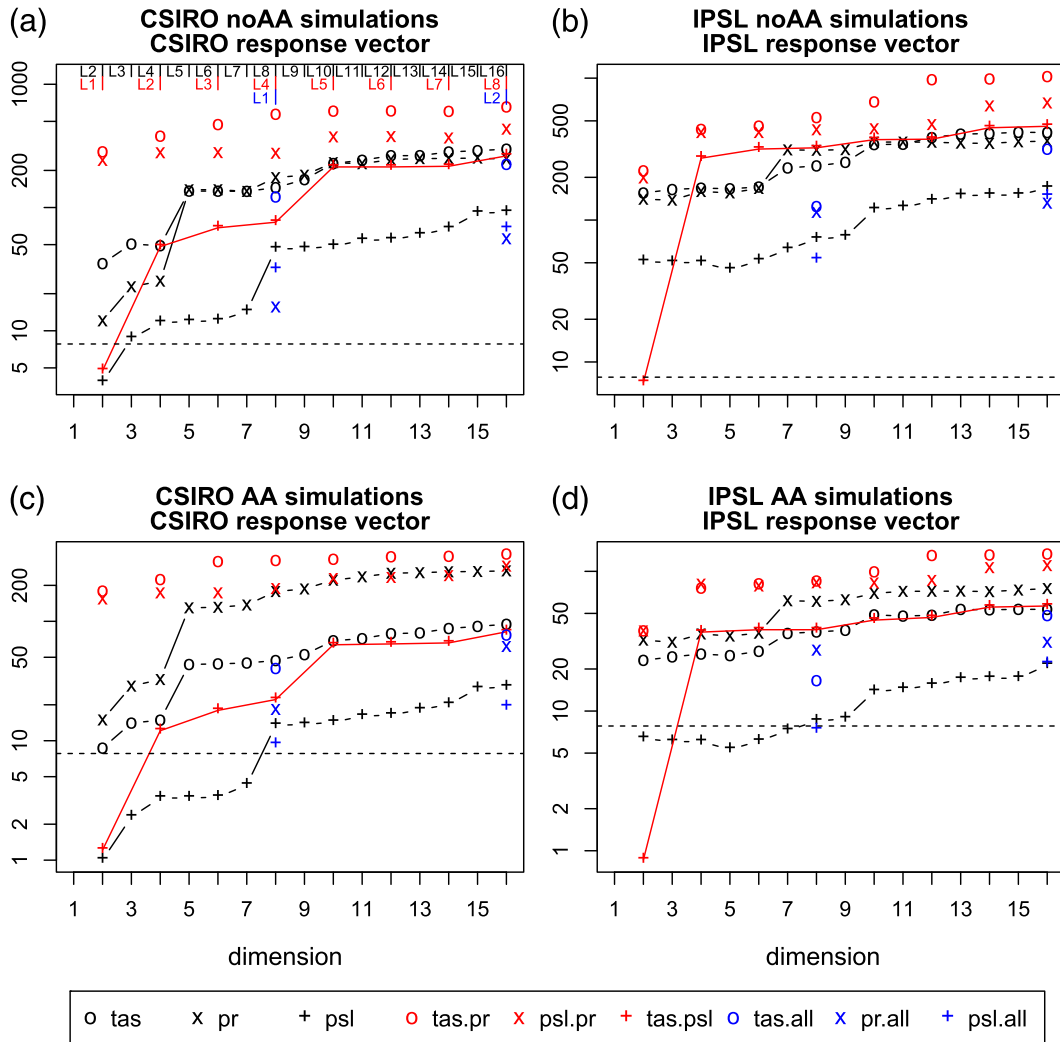
FIG. 7. Potential detectability of (a) noAA in CSIRO, (b) noAA in IPSL, (c) AA in CSIRO, and (d) noAA in IPSL. Black symbols are for single variables (TAS, PR, and PSL). Red symbols are for variable pairs (TAS–PR, PSL–PR, and TAS–PSL). Blue symbols are for single variables using variations between seasons and between land and sea. The horizontal axis shows the dimension of the state vector. The precise state vectors are described in the text. The number of Laplacians of each variable used to construct different variable combinations is indicated by the marks at the top of the *x* axis in (a): black for single variables, red for pairs of variables, blue for variables incorporating information of both seasonal variation and land–sea contrast. Dashed line indicates the threshold for detectability at the 5% significance level.

runs. To precisely state the state vector, let $Y_{v,l}^{\Delta T,\text{domain}}$ denote the *v*th physical variable—TAS, PR, or PSL—for the *l*th Laplacian eigenvector for time averaging period ΔT and geographic "domain." For a variable *v* on a global domain averaged over a year (ann), the state vector is of the following form:

$$\mathbf{p} = (Y_{v,1}^{\text{ann,globe}} \quad Y_{v,2}^{\text{ann,globe}} \quad Y_{v,3}^{\text{ann,globe}} \quad \ldots)^{\text{T}}. \quad (9)$$

In this case, as the dimension of the vector increases, more spatial gradient information is included through higher-order spherical harmonics. The vector is calculated by maximizing total-to-noise ratio in the first ensemble member of the appropriate forced run from each model. The potential detectability ϕ_m^2 as a function of vector dimension using 3 Legendre polynomials are shown as black symbols in Fig. 7 (the other symbols will be discussed shortly).

Potential detectability slightly above the significance threshold implies that the signal is marginally above the noise level for detection. While such values are desirable because they meet the criterion for detection, in practice,

we are interested in response vectors whose potential detectability is several times higher than the threshold because these vectors can detect a signal that is several times larger than the noise level. To be clear, we would not select the vector that maximizes detectability in the training sample because of overfitting.

Figure 7 shows that, in general, TAS and PR lead to stronger detection than PSL. PR gives stronger detection of AA than any other single variable. PSL cannot detect AA forced responses if too few spherical harmonics (<8) are included in the state vector. In CSIRO, the potential detectability jumps after including the second and fifth spherical harmonic, suggesting that the hemispheric gradient and pole-to-equator gradient provides critical information for separating AA and noAA forced responses. Such jumps are not evident in IPSL, indicating that the importance of spatial gradient information to separating AA and noAA forced responses

is model dependent. Note that AA or noAA forcing individually produce spatial gradients. The above result implies that the degree to which such spatial gradients discriminate between forcings is model dependent.

Recall that the potential detectability ϕ^2 is proportional to the product of the total-to-noise ratio and multicollinearity measure $(1 - \rho^2)$. These individual terms are shown in Fig. 8. For the single-variable case (i.e., black symbols), Fig. 8 shows that PSL has relatively low total-to-noise ratio and high degree of collinearity compared to TAS or PR. This explains why PSL is less effective than other variables at detecting AA. PR has the least collinearity between AA and noAA responses than any other single variable.

We next consider potential detectability for pairs of variables. Specifically, for a pair of physical variables i and j on a global domain averaged over a year, the state vector is of the following form:

$$\mathbf{p} = (Y_{i,1}^{\text{ann,globe}} \quad Y_{j,1}^{\text{ann,globe}} \quad Y_{i,2}^{\text{ann,globe}} \quad Y_{j,2}^{\text{ann,globe}} \quad Y_{i,3}^{\text{ann,globe}} \quad Y_{j,3}^{\text{ann,globe}} \quad \dots)^{\text{T}}. \quad (10)$$

For instance, if $i = $ TAS and $j = $ PR, then the first two elements give the projection of these variables on the first spherical harmonic, the next two elements give the projection on the second spherical harmonic, and so on. The potential detectability for this case is shown as red symbols in Fig. 7. Note that in this case the vector dimension does not equal the number of spherical harmonics since the vector is formed by concatenating pairs of variables. Accordingly, the dimension of the state vector is shown in the lower x axis while the number of spherical harmonics is shown on the upper x axis of Fig. 7. Figure 7 reveals that state vectors based on TAS–PR variable pairs give the strongest potential detectability relative to other vectors of the same dimension. Interestingly, potential detectability based on TAS–PR

does not increase strongly beyond the second spherical harmonic. The strong potential detectability using TAS–PR is because the associated response vectors are less collinear than other vectors (see Fig. 8). For example, other variables in CSIRO can have larger total-to-noise ratios but lower potential detectability due to high collinearity.

As mentioned in the introduction, many studies suggest that seasonal variations and land–sea gradients are important for detection. We have examined state vectors based on this type of information, but potential detectability often is worse than just using spatial gradients. As an example, we select state vectors that include both seasonal variations and land–sea gradients of the following form:

$$\mathbf{p} = (Y_{v,1}^{\text{djf,land}} \quad Y_{v,1}^{\text{mam,land}} \quad Y_{v,1}^{\text{jja,land}} \quad Y_{v,1}^{\text{son,land}} \quad Y_{v,1}^{\text{djf,ocean}} \quad Y_{v,1}^{\text{mam,ocean}} \quad \dots)^{\text{T}}, \quad (11)$$

where the superscripts indicate the obvious 3-month mean and the domain over which Laplacian eigenvectors are computed. For this vector, the first eight elements correspond to the first Laplacian eigenvector (i.e., the spatial average over the domain), and the second eight elements correspond to the second Laplacian eigenvector in the respective domain. Potential detectability based on these vectors is shown as blue symbols in Fig. 7. For temperature and precipitation, this vector leads to less potential detectability than a

vector of the same dimension based on spherical harmonics (especially for IPSL). The total-to-noise ratios for AA forcing are especially low for vectors based on (11) (see Fig. 8). These results suggest that seasonal variations or land–sea contrast information does not significantly enhance potential detectability relative to global, annual mean indices.

The above conclusions pertain to a perfect model analysis. The question arises as to whether our conclusions hold in an imperfect model setting. Figure 9 shows
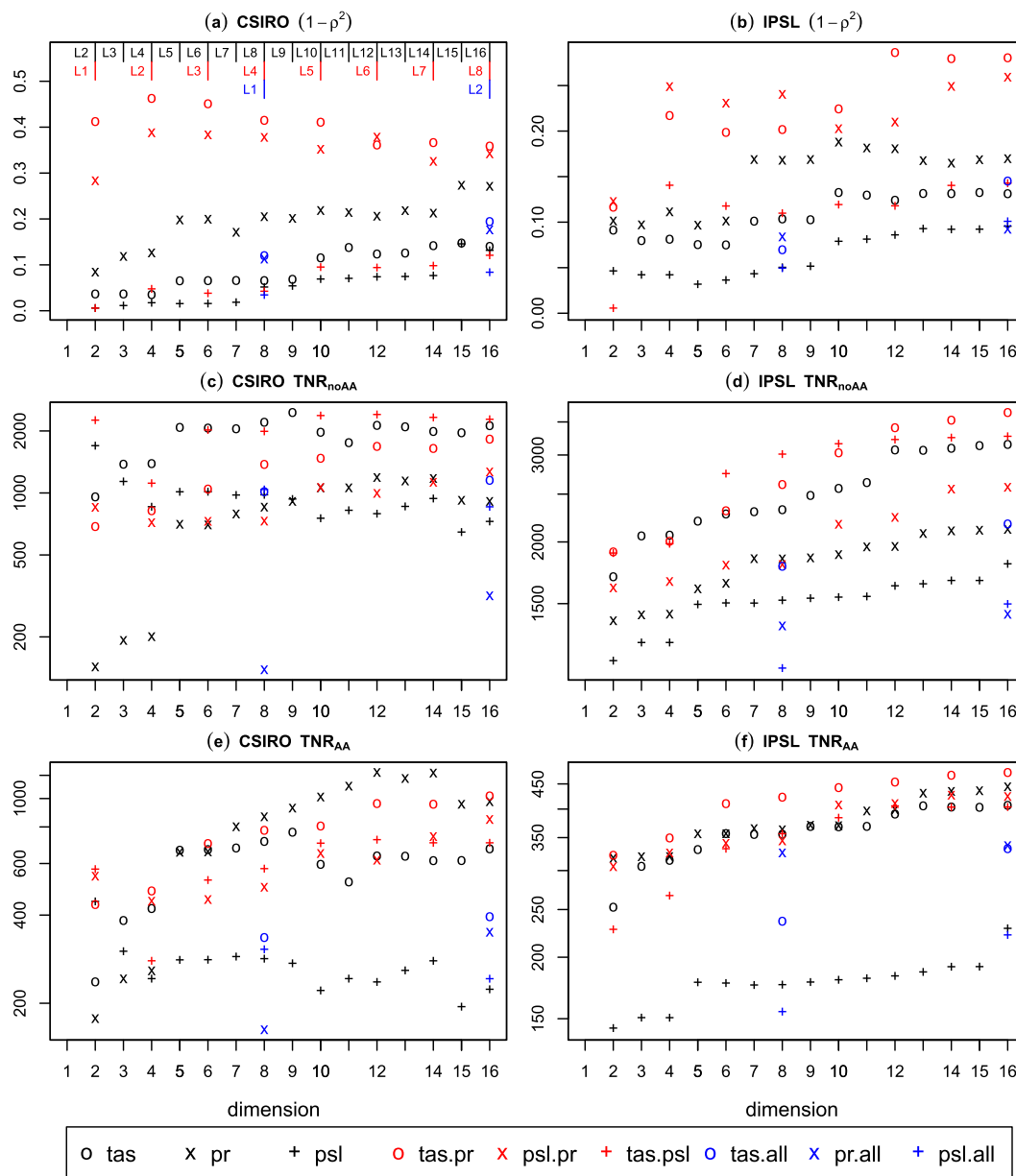
FIG. 8. (top) Multicollinearity measure, (middle) the total-to-noise ratio for noAA, and (bottom) the total-to-noise ratio of AA, for (left) CSIRO and (right) IPSL, for the variable combinations indicated by colored symbols corresponding to Fig. 7.

results similar to those in Fig. 7, except for the imperfect model case in which response vectors and noise covariances are estimated from a model that differs from the model that generated the historical single-forcing (AA and noAA) runs. As in the perfect model case, precipitation still gives the strongest potential detectability of AA, and combining precipitation with either temperature or sea level pressure gives the strongest potential detectability of all. In contrast, jumps in potential detectability occur at the second and fifth Laplacian

in IPSL rather than CSIRO, implying that the jumps are due to resolving gradient information in the response vector and the noise covariance matrix rather than the single-forcing runs used. The use of seasonal variations or land–sea contrast information is more mixed in the imperfect model case; such information can slightly enhance potential detectability relative to vectors of the same dimension using annual mean information and spherical harmonics, but these increases are generally small, whereas the losses of potential detectability as a
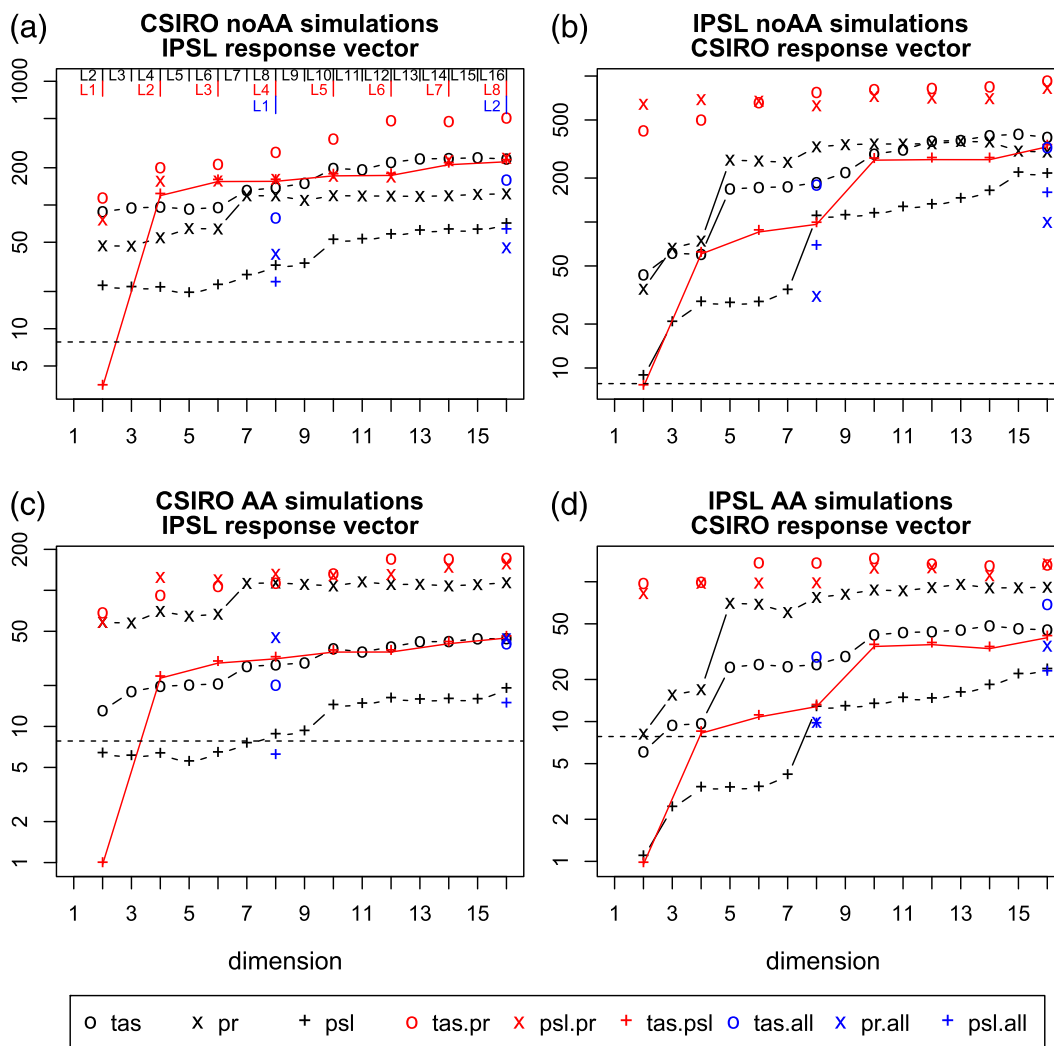
FIG. 9. As in Fig. 7, but for imperfect model cases, in which response vectors and noise covariances are estimated from a model that differs from the model that generated the historical single-forcing runs, as follows: (a) detecting noAA with response pattern from CSIRO and noise covariance matrix from IPSL, (b) detecting noAA with response pattern from IPSL and noise covariance matrix from CSIRO, (c) detecting AA with response pattern from CSIRO and noise covariance matrix from IPSL, and (d) detecting AA with response pattern from IPSL and noise covariance matrix from CSIRO.

result of seasonal variations or land–sea contrast information can be substantial (especially for precipitation).

Note that TAS–PSL produces weaker potential detectability than other variable pairs. PSL is atmospheric pressure extrapolated to sea level and is computed partly from temperature data. The question arises as to whether PSL enhances potential detectability beyond TAS. In some cases, TAS and PSL are highly correlated, as shown in Fig. 10 for the global mean, which explains why the AA response is not detectable using only one Laplacian eigenvector of TAS and PSL. In all cases shown in Fig. 7, TAS always has stronger potential detectability than PSL when based on the same number of Laplacian eigenvectors. In rare cases in Fig. 9, PSL has

only marginally larger potential detectability than TAS. In general, then, PSL does not enhance potential detectability significantly beyond that already available from the temperature data used to compute PSL.

The potential detectabilities shown above are computed from single-forcing runs without actually analyzing data in which both forcings are present. Realized detectability also can be computed from (7) using data from historical simulations, in which both forcings are present. A comparison between the potential detectability and realized detectability for the different vectors is shown in Fig. 11. In general, potential detectability tends to overestimate the realized detectability. This bias is presumably related to the neglect of sampling
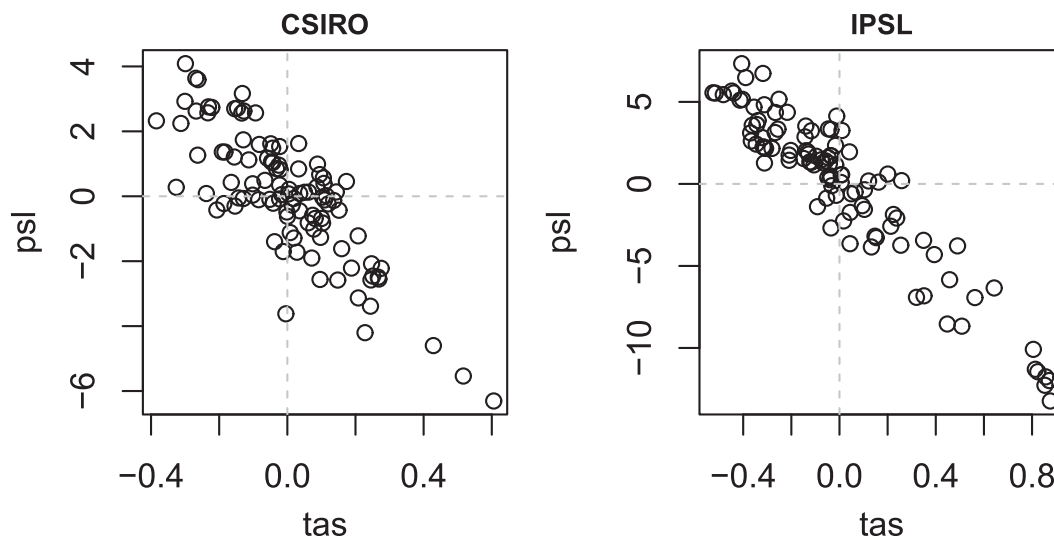
FIG. 10. Scatterplot of annual mean global average temperature and sea level pressure over the period 1900–2004.

variability in the forced response patterns, an artifact of least squares that can be avoided using the error-in-variables methodology (Allen and Stott 2003). Nevertheless, the two measures tend to agree more strongly as the realized detectability increases beyond the significance threshold.

5. Summary and discussion

This paper showed that joint temperature–precipitation information over a global domain provides the more accurate estimate of aerosol-forced responses in climate models compared to using temperature, precipitation, or sea level pressure individually or in combination. This fact is demonstrated using a new quantity called potential detectability, which measures the extent to which a forced response can be detected in a model. Significant potential detectability is a necessary condition for detection in observations—if a forced response cannot be detected in a perfect model scenario, then it cannot provide a legitimate basis for detection and attribution in observations. Importantly, potential detectability can be evaluated from single-forcing runs. Consequently, potential detectability allows the detectability of a forced response in a model to be compared across a variety of variables so that the best choice of variables for detecting a given forced response can be identified prior to analyzing observations. When the forcing is partitioned into two response vectors, potential detectability is proportional to the product of the total-to-noise ratio of the forced response multiplied by a measure of multi-collinearity between the two response vectors. This result substantially clarifies the relation between detectability, signal-to-noise ratio, and multicollinearity.

Detectability is generally enhanced by selecting response vectors that maximize signal-to-noise ratio. Such vectors can be identified using the method of Jia and DelSole (2012). Typically, only one vector is detectable for any given forcing and model. To clarify the importance of spatial versus temporal information, we use only spatial structure as input to fingerprinting and infer the associated time series from fingerprinting (which subsumes the traditional scaling factor). To do this, the data were represented in terms of a few Laplacian eigenvectors instead of EOFs. Laplacian eigenvectors enhance physical interpretation (since they often are of the form of monopoles, dipoles, tripoles, and so forth) and are data independent and hence can be compared across models. Also, the time evolution is assumed to be low frequency, in the sense that it is representable by a low-order polynomial in time. Since no temporal information is included in the response vector, the results clarify how detectability depends on spatial structure and covariability with other variables.

To illustrate the above methodology, we used it to estimate the change in global average temperature in historical simulations due to different forcings using joint temperature–precipitation response vectors. The inferred time series closely matched the time series computed from single-forcing runs, even if only the first spherical harmonic was used and even if the response vector of one model is used to infer the change in a different model. In some models, the accuracy of the estimates improves as more spherical harmonics are included in the response vector. A third-order polynomial in time was found to be sufficient for capturing most of the response over a century.
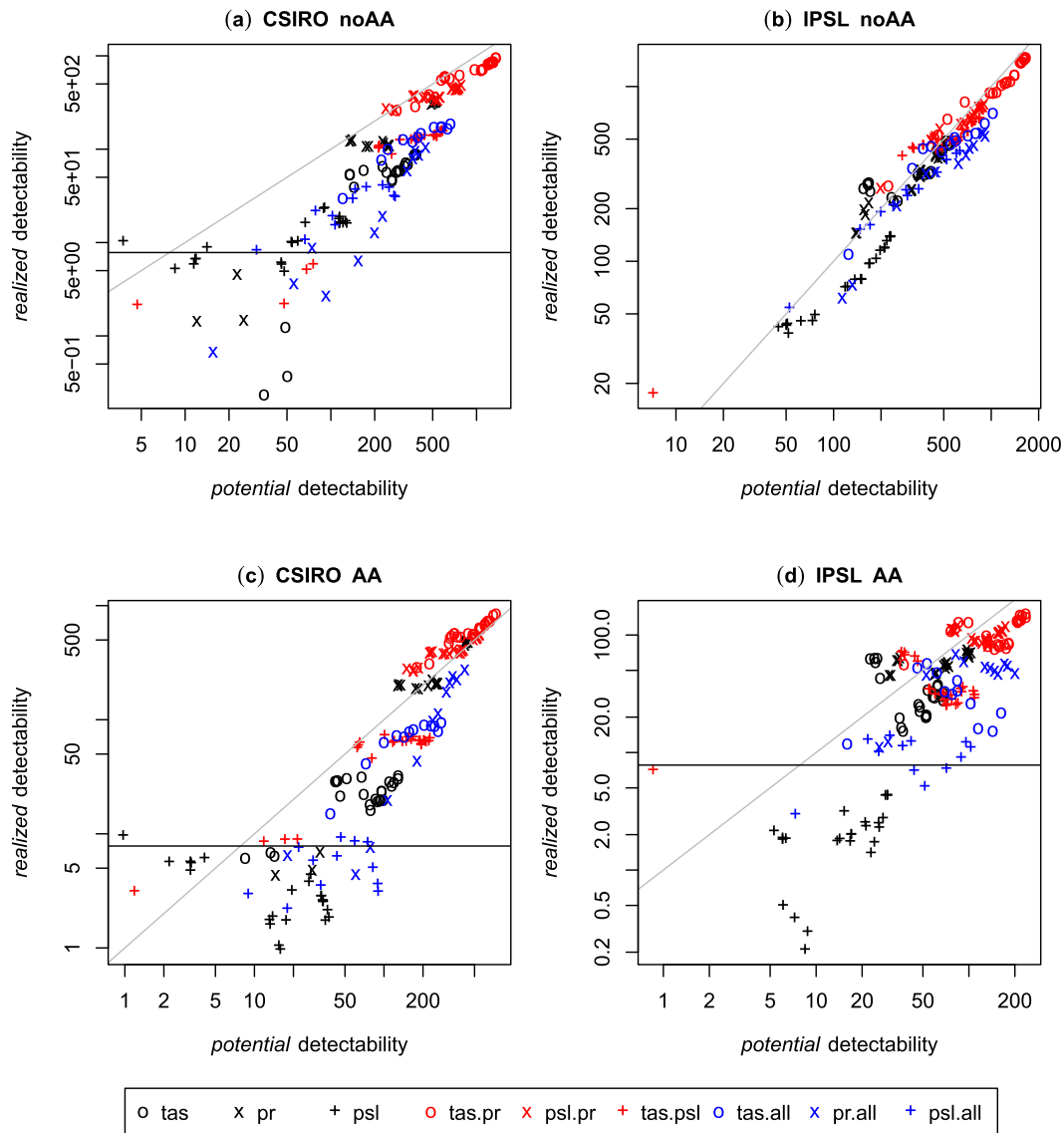
FIG. 11. Comparison of potential detectability and realized detectability for (top) noAA forcing and (bottom) AA forcing and for (left) CSIRO and (right) IPSL. Potential detectability is computed using the TNR ratio estimated from single-forcing run, and realized detectability is computed using the TNR ratio estimated from the historical run. The colored symbols correspond to the same cases as in Fig. 7. Gray lines have slope of 1 and cross the origin. Black horizontal lines indicate the threshold of 5% significance level of realized detectability.

The above methodology was applied to identify the more effective variables in a model for estimating the cooling attributable to anthropogenic aerosols. We explored three different variables—surface temperature, precipitation, and sea level pressure—and various combinations of spatiotemporal information. We found that joint temperature–precipitation response vectors yielded the strongest detection of aerosol cooling, even in the absence of spatial and temporal gradient information. The best single variable for detecting anthropogenic aerosols is annual mean precipitation. We

call attention to the fact that sea level pressure is computed partly from temperature data and hence is not independent of temperature. We find no evidence that sea level pressure provides any significant gain in detectability beyond that already available in the surface temperature data from which sea level pressure is derived. We also used state vectors that included seasonal variations and that separated land and ocean domains, but these vectors generally produced less detectability than vectors of the same dimension based on annual mean variables of Laplacian eigenvectors. These conclusions

hold even in the imperfect model setting, in which response vectors and noise covariances are estimated from a model that differs from the model that generated the historical single-forcing runs. In one model, jumps in detectability occurred after the second and fifth spherical harmonics were included, which measure, respectively, the north–south hemispheric gradient and equator-to-pole gradient. Further comparison between perfect and imperfect model analyses indicate that these jumps are due to resolving gradient information in the response vectors and the noise covariance matrix.

As emphasized throughout this paper, our results are derived from model experiments in which errors in the forced response and observations can be neglected. There is no guarantee that response vectors that work well in models also will work well in observations. In fact, DelSole et al. (2015, manuscript submitted to *J. Climate*) applied optimal fingerprinting to observations using global mean joint temperature and precipitation information, but observed, global, long-term precipitation was inconsistent with models and even between datasets. Thus, detection analysis based on observed precipitation turns out to be difficult, despite the fact that model precipitation is the best single detection variable in a model setting. Nevertheless, the methodology illustrated here is expected to provide a useful starting point for deciding which variables to use for detection and attribution analysis of observations.

We investigated AA and noAA simulations in this paper because we were interested primarily in the response to anthropogenic aerosols. The methodology presented in this paper is not limited to the combination of single forcings of AA and noAA. In general, the potential detectability measure can be computed for any problem that can be formulated as optimal fingerprinting.

APPENDIX

Generalization of Optimal Fingerprinting and Derivation of Potential Detectability

This appendix shows how optimal fingerprinting can be generalized to estimate time series for forced responses and formally derives the potential detectability measure. We begin by first summarizing generalized least squares (GLS), which is a standard framework in linear regression and then discuss how optimal fingerprinting relates to GLS. This material is standard (Allen and Tett 1999; Hegerl et al. 2007; Hegerl and Zwiers 2011). Then, we show how our modified regression problem can be framed as a GLS problem and then derive potential detectability.

Generalized least squares assumes that an observed K-dimensional vector $\mathbf{y}$ can be modeled as follows:

$$\underset{K \times 1}{\mathbf{y}} = \underset{K \times M}{\mathbf{X}} \; \underset{M \times 1}{\boldsymbol{\beta}} + \underset{K \times 1}{\boldsymbol{\epsilon}}, \quad \text{(A1)}$$

where $\mathbf{X}$ is a matrix of known "explanatory variables," $\boldsymbol{\beta}$ is a vector of unknown regression coefficients to be estimated from data, and $\boldsymbol{\epsilon}$ is a random noise term. Ordinary least squares (OLS) gives estimates of $\boldsymbol{\beta}$ when elements of the noise term $\boldsymbol{\epsilon}$ are independent and identically distributed. In contrast, GLS gives estimates of the regression coefficients when elements of the noise term are correlated, or more precisely, have covariance matrix $\tilde{\boldsymbol{\Sigma}}_\varepsilon$:

$$\text{cov}[\boldsymbol{\epsilon}] = \tilde{\boldsymbol{\Sigma}}_\varepsilon. \quad \text{(A2)}$$

A standard fact is that a variable transformation can adjust the covariance matrix so that OLS can be used. The end result is that the GLS estimates of the regression coefficients is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}} \tilde{\boldsymbol{\Sigma}}_\varepsilon^{-1} \mathbf{X})^{-1} \mathbf{X}^{\mathrm{T}} \tilde{\boldsymbol{\Sigma}}_\varepsilon^{-1} \mathbf{y}, \quad \text{(A3)}$$

where the caret ̂ denotes a sample estimate. The estimate in (A3) is unbiased, in the sense that $E[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$ and has the following covariance matrix:

$$\text{cov}[\hat{\boldsymbol{\beta}}] = (\mathbf{X}^{\mathrm{T}} \tilde{\boldsymbol{\Sigma}}_\varepsilon^{-1} \mathbf{X})^{-1}. \quad \text{(A4)}$$

If the noise is normally distributed, then $\hat{\boldsymbol{\beta}}$ also is normally distributed, and a 95% confidence interval for the mth regression coefficient is

$$(\hat{\beta}_m - 1.96\{(\mathbf{X}^{\mathrm{T}} \tilde{\boldsymbol{\Sigma}}_\varepsilon^{-1} \mathbf{X})^{-1}\}_{mm}^{1/2},$$
$$\hat{\beta}_m + 1.96\{(\mathbf{X}^{\mathrm{T}} \tilde{\boldsymbol{\Sigma}}_\varepsilon^{-1} \mathbf{X})^{-1}\}_{mm}^{1/2}). \quad \text{(A5)}$$

Optimal fingerprinting assumes a model of the form of (1), repeated below:

$$\mathbf{o} = \sum_{m=1}^{M} \mathbf{f}_m a_m + \boldsymbol{\eta}. \quad \text{(A6)}$$

In this paper, we assume that the forced response $\mathbf{f}_m$ is represented as (2), where the response pattern $\mathbf{p}_m$ has

been derived by maximizing the signal-to-noise ratio [i.e., by solving (3)], and use GLS to estimate the time series $\mathbf{r}_m$ (which subsumes the scaling factor a_m). In addition, we assume the response time series $\mathbf{r}_m$ can be represented by a low-order polynomial in time, as in (4). Under these assumptions, the forced response vector $\mathbf{f}_m$ at the sth spatial location and nth time step is

$$(\mathbf{f}_m)_{s,n} = p_{s,m} r_{n,m} = \sum_{j=1}^{J} p_{s,m} c_{m,j} L_{n,j}. \qquad (A7)$$

To express the final regression model in the form in (A1), it is convenient to "stack" the spatial locations first and then loop over time, leading to the following identifications:

$$y_{s+S(n-1)} = o_{s,n}, \qquad (A8)$$

$$X_{s+S(n-1),m+M(j-1)} = p_{s,m} L_{n,j}, \qquad (A9)$$

$$\beta_{m+(j-1)} = c_{m,j}, \quad \text{and} \qquad (A10)$$

$$\varepsilon_{s+S(n-1)} = \eta_{s,n}. \qquad (A11)$$

As discussed in the main text, we assume internal variability is Gaussian white noise. Under the white noise assumption, the covariance matrix of the noise term has the following block diagonal structure:

$$\tilde{\boldsymbol{\Sigma}}_\varepsilon = \begin{pmatrix} \boldsymbol{\Sigma}_\eta & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_\eta & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\Sigma}_\eta \end{pmatrix}, \qquad (A12)$$

where $\boldsymbol{\Sigma}_\eta$ is the single-time, $S \times S$ covariance matrix of internal variability. The covariance matrix of the estimated coefficients is the inverse of the following matrix:

$$(\mathbf{X}^T \tilde{\boldsymbol{\Sigma}}_\varepsilon^{-1} \mathbf{X})_{m+(j-1),m'+M(j'-1)}$$

$$= p_{s,m} L_{n,j} (\boldsymbol{\Sigma}_\eta^{-1})_{s,s'} \delta_{n,n'} p_{s',m'} L_{n',j'}, \quad \text{and} \quad (A13)$$

$$= (\mathbf{L}^T \mathbf{L})_{j,j'} (\mathbf{p}_m^T \boldsymbol{\Sigma}_\eta^{-1} \mathbf{p}_m)_{m,m'}, \qquad (A14)$$

where we employ the familiar summation convention in which indices that appear twice are summed over all values of that index. Because Legendre polynomials are orthogonal,

$$\mathbf{L}^T \mathbf{L} = \mathbf{I}. \qquad (A15)$$

This orthogonality relation implies that estimates for different coefficients are independent of each other and that the matrix (A13) has the following block diagonal form:

$$\mathbf{X}^T \tilde{\boldsymbol{\Sigma}}_\varepsilon^{-1} \mathbf{X} = \begin{pmatrix} \mathbf{P}^T \boldsymbol{\Sigma}_\eta^{-1} \mathbf{P} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{P}^T \boldsymbol{\Sigma}_\eta^{-1} \mathbf{P} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{P}^T \boldsymbol{\Sigma}_\eta^{-1} \mathbf{P} \end{pmatrix}, \qquad (A16)$$

where $\mathbf{P}$ is the $S \times M$ matrix containing the forced response patterns

$$\mathbf{P} = (\mathbf{p}_1 \quad \mathbf{p}_2 \quad \dots \quad \mathbf{p}_M). \qquad (A17)$$

Similar considerations yield

$$(\mathbf{X}^T \tilde{\boldsymbol{\Sigma}}_\varepsilon^{-1} \mathbf{y})_{m+(j-1)} = L_{n,j} p_{s,m} (\boldsymbol{\Sigma}_\eta^{-1})_{s,s'} \delta_{n,n'} \mathbf{O}_{s',n'} \qquad (A18)$$

$$= (\mathbf{p}_m^T \boldsymbol{\Sigma}_\eta^{-1} \mathbf{O} \mathbf{L})_j. \qquad (A19)$$

Substituting (A19) and (A16) into (A3) gives the following coefficient estimates:

$$\hat{c}_{m,j} = \{(\mathbf{P}^T \boldsymbol{\Sigma}_\eta^{-1} \mathbf{P})^{-1} \mathbf{P}^T \boldsymbol{\Sigma}_\eta^{-1} \mathbf{O} \mathbf{L}\}_{m,j}. \qquad (A20)$$

Let $\hat{\mathbf{C}}$ denote the Legendre coefficient matrix defined above, and let $\hat{\mathbf{c}}_j$ denote the M-dimensional vector of coefficients for the jth Legendre polynomial, so that

$$\hat{\mathbf{C}} = (\hat{\mathbf{c}}_1 \hat{\mathbf{c}}_2 \dots \hat{\mathbf{c}}_J). \qquad (A21)$$

Then, the covariance matrix of the "stacked" coefficients

$$\text{vec}[\hat{\mathbf{C}}] = \begin{pmatrix} \hat{\mathbf{c}}_1 \\ \hat{\mathbf{c}}_2 \\ \vdots \\ \hat{\mathbf{c}}_J \end{pmatrix} \qquad (A22)$$

is (A26). The block diagonal structure of the covariance matrix (A16) implies that $\hat{\mathbf{c}}_i$ and $\hat{\mathbf{c}}_j$ are independent for $i \neq j$. Moreover, the block diagonals are equal, so the covariance matrix of the estimates for the jth Legendre polynomial are independent of j and given by

$$\text{cov}[\hat{\mathbf{c}}_j] = (\mathbf{P}^T \boldsymbol{\Sigma}_\eta^{-1} \mathbf{P})^{-1}. \qquad (A23)$$

Thus, a 95% confidence interval for the jth Legendre polynomial for the mth forced response is

$$(\hat{c}_{m,j} - 1.96\sqrt{\{(\mathbf{P}^T \boldsymbol{\Sigma}_\eta^{-1} \mathbf{P})^{-1}\}_{mm}},$$

$$\hat{c}_{m,j} + 1.96\sqrt{\{(\mathbf{P}^T \boldsymbol{\Sigma}_\eta^{-1} \mathbf{P})^{-1}\}_{mm}}). \qquad (A24)$$

The null hypothesis is that the forced response vanishes (i.e., $\mathbf{C} = \mathbf{0}$). Under this null hypothesis, the statistic given by

$$\phi_m^2 = \frac{\sum_{j=1}^{J} \hat{\mathbf{c}}_{m,j}^2}{\left\{ (\mathbf{P}^{\mathrm{T}} \boldsymbol{\Sigma}_\eta^{-1} \mathbf{P})^{-1} \right\}_{mm}} \qquad (A25)$$

has a chi-squared distribution with J degrees of freedom. The quantity ϕ_m^2 measures the detectability of the mth forced response and must exceed a critical value for a forced response to be detected.

If only one forced response vector (i.e., $M = 1$) is considered, the covariance matrix for the Legendre coefficients reduces to the scalar $\mathrm{cov}[\hat{\mathbf{c}}_j] = (\mathbf{p}_m^{\mathrm{T}} \boldsymbol{\Sigma}_\eta^{-1} \mathbf{p}_m)^{-1}$. Then the detectability in (A25) of one forced response vector becomes

$$\phi_m^2[M = 1] = \left(\sum_{j=1}^{J} \hat{c}_{m,j}^2 \right) \left(\mathbf{p}_m^{\mathrm{T}} \boldsymbol{\Sigma}_\eta^{-1} \mathbf{p}_m \right). \qquad (A26)$$

Note that this expression is of the form of a total-to-noise ratio:

$$\phi_m^2[M = 1] = \frac{\sum_{j=1}^{J} \hat{c}_{m,j}^2}{\mathrm{cov}[\hat{\mathbf{c}}_j]} = \mathrm{TNR}_m, \qquad (A27)$$

where TNR_m denotes for the total-to-noise ratio of the forced response relative to the internal noise.

Now we consider the case of two forced response patterns $\mathbf{p}_1$ and $\mathbf{p}_2$. In this case,

$$\mathbf{P}^{\mathrm{T}} \boldsymbol{\Sigma}_\eta^{-1} \mathbf{P} = \begin{pmatrix} \mathbf{p}_1^{\mathrm{T}} \boldsymbol{\Sigma}_\eta^{-1} \mathbf{p}_1 & \mathbf{p}_1^{\mathrm{T}} \boldsymbol{\Sigma}_\eta^{-1} \mathbf{p}_2 \\ \mathbf{p}_2^{\mathrm{T}} \boldsymbol{\Sigma}_\eta^{-1} \mathbf{p}_1 & \mathbf{p}_2^{\mathrm{T}} \boldsymbol{\Sigma}_\eta^{-1} \mathbf{p}_2 \end{pmatrix}. \qquad (A28)$$

The inverse of this matrix can be computed analytically as

$$(\mathbf{P}^{\mathrm{T}} \boldsymbol{\Sigma}_\eta^{-1} \mathbf{P})^{-1}$$
$$= \frac{1}{\det[\mathbf{P}^{\mathrm{T}} \boldsymbol{\Sigma}_\eta^{-1} \mathbf{P}]} \left\{ \begin{matrix} (\mathbf{p}_2^{\mathrm{T}} \boldsymbol{\Sigma}_\eta^{-1} \mathbf{p}_2) & -(\mathbf{p}_1^{\mathrm{T}} \boldsymbol{\Sigma}_\eta^{-1} \mathbf{p}_2) \\ -(\mathbf{p}_2^{\mathrm{T}} \boldsymbol{\Sigma}_\eta^{-1} \mathbf{p}_1) & (\mathbf{p}_1^{\mathrm{T}} \boldsymbol{\Sigma}_\eta^{-1} \mathbf{p}_1) \end{matrix} \right\}. \qquad (A29)$$

The determinant is

$$\det[\mathbf{P}^{\mathrm{T}} \boldsymbol{\Sigma}_\eta^{-1} \mathbf{P}] = (\mathbf{p}_1^{\mathrm{T}} \boldsymbol{\Sigma}_\eta^{-1} \mathbf{p}_1)(\mathbf{p}_2^{\mathrm{T}} \boldsymbol{\Sigma}_\eta^{-1} \mathbf{p}_2)(1 - \rho^2), \qquad (A30)$$

where

$$\rho = \frac{\mathbf{p}_1^{\mathrm{T}} \boldsymbol{\Sigma}_\eta^{-1} \mathbf{p}_2}{\sqrt{\mathbf{p}_1^{\mathrm{T}} \boldsymbol{\Sigma}_\eta^{-1} \mathbf{p}_1} \sqrt{\mathbf{p}_2^{\mathrm{T}} \boldsymbol{\Sigma}_\eta^{-1} \mathbf{p}_2}}. \qquad (A31)$$

The parameter ρ measures the cosine of the angle between the response vectors, which in turn can be interpreted as a generalized spatial correlation. Consolidating the above results into the detection statistic in (A25) gives

$$\phi_m^2[M = 2] = \left(\sum_{j=1}^{J} \hat{c}_{m,j}^2 \right) (\mathbf{p}_m^{\mathrm{T}} \boldsymbol{\Sigma}_\eta^{-1} \mathbf{p}_m)(1 - \rho^2). \qquad (A32)$$

A comparison of (A26) and (A32) shows that the detectability of the response when two forced responses are present is equal to the detectability of the response associated with one response vector scaled by a collinearity term $(1 - \rho^2)$:

$$\phi_m^2 = \mathrm{TNR}_m (1 - \rho^2). \qquad (A33)$$

For future reference, we note that if the time series were estimated for each year independently, as we do in the case of a single forced response, then the generalized least squares estimate of the time series is

$$\hat{r}_{m,y} = \{ (\mathbf{p}_m^{\mathrm{T}} \boldsymbol{\Sigma}_\eta^{-1} \mathbf{p}_m)^{-1} \mathbf{p}_m^{\mathrm{T}} \boldsymbol{\Sigma}_\eta^{-1} \mathbf{O} \}_{m,y}. \qquad (A34)$$

REFERENCES

Allen, M. R., and S. F. B. Tett, 1999: Checking for model consistency in optimal fingerprinting. *Climate Dyn.*, **15**, 419–434, doi:10.1007/s003820050291.

——, and P. A. Stott, 2003: Estimating signal amplitudes in optimal fingerprinting, part I: Theory. *Climate Dyn.*, **21**, 477–491, doi:10.1007/s00382-003-0313-9.

Allen, R. J., J. R. Norris, and M. Kovilakam, 2014: Influence of anthropogenic aerosols and the Pacific decadal oscillation on tropical belt width. *Nat. Geosci.*, **7**, 270–274, doi:10.1038/ngeo2091.

Barnett, T. P., and Coauthors, 2008: Human-induced changes in the hydrology of the western United States. *Science*, **319**, 1080–1083, doi:10.1126/science.1152538.

Bindoff, N. L., and Coauthors, 2013: Detection and attribution of climate change: From global to regional. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 867–952.

Booth, B. B. B., N. J. Dunstone, P. R. Halloran, R. Andrews, and N. Bellouin, 2012: Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability. *Nature*, **484**, 228–232, doi:10.1038/nature10946.

Boucher, O., and Coauthors, 2013: Clouds and aerosols. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 571–657.

Chang, C. Y., J. C. H. Chiang, M. F. Wehner, A. R. Friedman, and R. Ruedy, 2011: Sulfate aerosol control of tropical Atlantic climate over the twentieth century. *J. Climate*, **24**, 2540–2555, doi:10.1175/2010JCLI4065.1.

DelSole, T., and M. K. Tippett, 2007: Predictability: Recent insights from information theory. *Rev. Geophys.*, **45**, RG4002, doi:10.1029/2006RG000202.

——, and ——, 2015: Laplacian eigenfunctions for climate analysis. *J. Climate*, **28**, 7420–7436, doi:10.1175/JCLI-D-15-0049.1.

Dong, L., and T. Zhou, 2014: The Indian Ocean sea surface temperature warming simulated by CMIP5 models during the twentieth century: Competing forcing roles of GHGs and anthropogenic aerosols. *J. Climate*, **27**, 3348–3362, doi:10.1175/JCLI-D-13-00396.1.

Gettelman, A., D. Shindell, and J. Lamarque, 2015: Impact of aerosol radiative effects on 2000–2010 surface temperatures. *Climate Dyn.*, **45**, 2165–2179, doi:10.1007/s00382-014-2464-2.

Hegerl, G., and F. Zwiers, 2011: Use of models in detection and attribution of climate change. *Wiley Interdisc. Rev.: Climate Change*, **2**, 570–591, doi:10.1002/wcc.121.

——, and Coauthors, 2007: Understanding and attributing climate change. *Climate Change 2007: The Physical Science Basis*, S. Solomon et al., Eds., Cambridge University Press, 663–745.

Iles, C. E., G. C. Hegerl, A. P. Schurer, and X. Zhang, 2013: The effect of volcanic eruptions on global precipitation. *J. Geophys. Res. Atmos.*, **118**, 8770–8786, doi:10.1002/jgrd.50678.

Jia, L., and T. DelSole, 2012: Optimal determination of time-varying climate change signals. *J. Climate*, **25**, 7122–7137, doi:10.1175/JCLI-D-11-00434.1.

Jones, G. S., S. F. B. Tett, and P. A. Stott, 2003: Causes of atmospheric temperature change 1960–2000: A combined attribution analysis. *Geophys. Res. Lett.*, **30**, 1228, doi:10.1029/2002GL016377.

——, P. A. Stott, and N. Christidis, 2013: Attribution of observed historical near-surface temperature variations to anthropogenic and natural causes using CMIP5 simulations. *J. Geophys. Res. Atmos.*, **118**, 4001–4024, doi:10.1002/jgrd.50239.

Lamarque, J.-F., and Coauthors, 2010: Historical (1850–2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: Methodology and application. *Atmos. Chem. Phys.*, **10**, 7017–7039, doi:10.5194/acp-10-7017-2010.

Lawley, D. N., 1956: Tests of significance for the latent roots of covariance and correlation matrices. *Biometrika*, **43**, 128–136, doi:10.1093/biomet/43.1-2.128.

Lean, J. L., and D. H. Rind, 2009: How will Earth's surface temperature change in future decades? *Geophys. Res. Lett.*, **36**, L15708, doi:10.1029/2009GL038932.

Marvel, K., and C. Bonfils, 2013: Identifying external influences on global precipitation. *Proc. Natl. Acad. Sci. USA*, **110**, 19 301–19 306, doi:10.1073/pnas.1314382110.

——, G. A. Schmidt, D. Shindell, C. Bonfils, A. N. LeGrande, L. Nazarenko, and K. Tsigaridis, 2015: Do responses to different anthropogenic forcings add linearly in climate models? *Environ. Res. Lett.*, **10**, 104010, doi:10.1088/1748-9326/10/10/104010.

Myhre, G., and Coauthors, 2013: Anthropogenic and natural radiative forcing. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 659–740.

Pierce, D. W., P. J. Gleckler, T. P. Barnett, B. D. Santer, and P. J. Durack, 2012: The fingerprint of human-induced changes in the ocean's salinity and temperature fields. *Geophys. Res. Lett.*, **39**, L21704, doi:10.1029/2012GL053389.

Ridley, H. E., and Coauthors, 2015: Aerosol forcing of the position of the intertropical convergence zone since AD 1550. *Nat. Geosci.*, **8**, 195–200, doi:10.1038/ngeo2353.

Rotstayn, L. D., and U. Lohmann, 2002: Tropical rainfall trends and the indirect aerosol effect. *J. Climate*, **15**, 2103–2116, doi:10.1175/1520-0442(2002)015<2103:TRTATI>2.0.CO;2.

——, S. J. Jeffrey, M. A. Collier, S. M. Dravitzki, A. C. Hirst, J. I. Syktus, and K. K. Wong, 2012: Aerosol- and greenhouse gas-induced changes in summer rainfall and circulation in the Australasian region: A study using single-forcing climate simulations. *Atmos. Chem. Phys.*, **12**, 6377–6404, doi:10.5194/acp-12-6377-2012.

Schnur, R., and K. Hasselmann, 2005: Optimal filtering for Bayesian detection and attribution of climate change. *Climate Dyn.*, **24**, 45–55, doi:10.1007/s00382-004-0456-3.

Shindell, D. T., A. Voulgarakis, G. Faluvegi, and G. Milly, 2012: Precipitation response to regional radiative forcing. *Atmos. Chem. Phys.*, **12**, 6969–6982, doi:10.5194/acp-12-6969-2012.

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

Stern, D. I., and R. K. Kaufmann, 2000: Detecting a global warming signal in hemispheric temperature series: A structural time series analysis. *Climatic Change*, **47**, 411–438, doi:10.1023/A:1005672231474.

Stott, P. A., and G. S. Jones, 2009: Variability of high latitude amplification of anthropogenic warming. *Geophys. Res. Lett.*, **36**, L10701, doi:10.1029/2009GL037698.

——, J. F. B. Mitchell, M. R. Allen, T. L. Delworth, J. M. Gregory, G. A. Meehl, and B. D. Santer, 2006: Observational constraints on past attributable warming and predictions of future global warming. *J. Climate*, **19**, 3055–3069, doi:10.1175/JCLI3802.1.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, doi:10.1175/BAMS-D-11-00094.1.

Xie, S.-P., B. Lu, and B. Xiang, 2013: Similar spatial patterns of climate responses to aerosol and greenhouse gas changes. *Nat. Geosci.*, **6**, 828–832, doi:10.1038/ngeo1931.

Zhou, J., and K.-K. Tung, 2013: Deducing multidecadal anthropogenic global warming trends using multiple regression analysis. *J. Atmos. Sci.*, **70**, 3–8, doi:10.1175/JAS-D-12-0208.1.