AGU PUBLICATIONS

Geophysical Research Letters

RESEARCH LETTER
10.1002/2017GL073688

Key Points:
- Novel computationally efficient Bayesian hierarchical approach for detection and attribution
- Uncertainty from covariance estimation incorporated in regression coefficient estimates
- A range of EOF truncations is considered and optimally weighted using Bayesian model averaging

Supporting Information:
- Supporting Information S1

Correspondence to:
D. Hammerling,
dorith@ucar.edu

Citation:
Katzfuss, M., D. Hammerling, and R. L. Smith (2017), A Bayesian hierarchical model for climate change detection and attribution, *Geophys. Res. Lett.*, *44*, 5720–5728, doi:10.1002/2017GL073688.

Received 23 DEC 2016
Accepted 22 MAY 2017
Accepted article online 24 MAY 2017
Published online 11 JUN 2017

A Bayesian hierarchical model for climate change detection and attribution

Matthias Katzfuss[1], Dorit Hammerling[2], and Richard L. Smith[3,4]

[1]Department of Statistics, Texas A&M University, College Station, Texas, USA, [2]Institute for Mathematics Applied to Geosciences, National Center for Atmospheric Research, Boulder, Colorado, USA, [3]Statistical and Applied Mathematical Sciences Institute, Research Triangle Park, North Carolina, USA, [4]Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

Abstract Regression-based detection and attribution methods continue to take a central role in the study of climate change and its causes. Here we propose a novel Bayesian hierarchical approach to this problem, which allows us to address several open methodological questions. Specifically, we take into account the uncertainties in the true temperature change due to imperfect measurements, the uncertainty in the true climate signal under different forcing scenarios due to the availability of only a small number of climate model simulations, and the uncertainty associated with estimating the climate variability covariance matrix, including the truncation of the number of empirical orthogonal functions (EOFs) in this covariance matrix. We apply Bayesian model averaging to assign optimal probabilistic weights to different possible truncations and incorporate all uncertainties into the inference on the regression coefficients. We provide an efficient implementation of our method in a software package and illustrate its use with a realistic application.

1. Introduction

The subject of climate change detection and attribution, also referred to as optimal fingerprinting, has continued to take a prominent role in the Assessment Reports of the Intergovernmental Panel on Climate Change (IPCC) [e.g., *Hegerl et al.*, 2007; *Bindoff et al.*, 2013] as the most common way to quantitatively assess if and how the climate has changed as a result of human activity. Broadly speaking, the goal of climate change detection and attribution methods is to differentiate if contributions to observed changes are climate internal or result from external forcings [*Hegerl and Zwiers*, 2011]. From a methodological point of view, this subject has mainly been addressed with regression methods, where the observed change is the dependent variable and the external forcing scenarios are the regressors or predictors. The goal is inference on the regression coefficients, whose estimated values and uncertainty ranges determine if a change has been detected and to which (combination of) scenarios it can be attributed.

Since its early development [*Hegerl et al.*, 1996], assuming a standard linear model with a single source of noise, the methodology has continued to evolve. *Allen and Stott* [2003] allowed for errors in the forcings signals and, assuming they have the same structure as the internal variability, applied a total least squares solution, which is still commonly used [e.g., *Lott et al.*, 2013]. This was further advanced by *Huntingford et al.* [2006], who relaxed the assumption of identical covariance structures. However, *Hannart et al.* [2014] argued that the approaches discussed in *Huntingford et al.* [2006] were not actually suitable to solve the problem as stated. *Hannart et al.* [2014] addressed this shortcoming and developed an iterative inference procedure based on likelihood optimization. This procedure successfully derives estimates of the regression coefficients and their asymptotic confidence intervals in the more general case of different covariance structures on observations and forcing signals. Yet *Hannart et al.* [2014] still assume that these covariance structures are known in the inferential procedure, stating that "While this [covariance] estimation step has been shown to critically influence the end result, it is in general handled preliminarily and rather independently from the regression inference which is at stake here." If covariances are estimated outside the regression, it is difficult to account for uncertainties associated with the covariance estimation in the inference on the regression coefficients. In recently published work, *Hannart* [2016] proposed a hierarchical regression approach, which accounts for uncertainty in the covariance matrix based on an inverse-Wishart prior.

Due to the high-dimensional nature of observations and model output (even when aggregated to relatively coarse grid cells), estimation of the climate variability is challenging. Typically, the covariance matrix of the climate variability is estimated based on a small to moderate number of "control runs" of a climate model without any forcings. To regularize the resulting estimate, several previous approaches [e.g., *Santer et al.*, 1993; *Ribes et al.*, 2009] have represented the covariance matrix using a small number of empirical orthogonal functions (EOFs) or principal components of the sample covariance matrix. This approach is effective for regularization, but it ignores the uncertainty in the shape and number of EOFs and the values of the corresponding eigenvalues in the subsequent analysis.

In this work, we consider an empirical Bayesian hierarchical framework that addresses some of the open methodological questions within the realm of regression-based climate change detection and attribution. We integrate out the true mean forcing signals under the assumption of a noninformative prior, which is more consistent with Bayesian inference than the standard practice of profiling (i.e., maximizing) them out [e.g., *Ribes et al.*, 2009; *Hannart et al.*, 2014; *Hannart*, 2016]. Another novel aspect to our work is the accounting for uncertainty in the observations as represented by an ensemble. Hence, our Bayesian framework allows us to incorporate the uncertainty in the observations and in the climate model variability explicitly in the regression procedure and to propagate these uncertainty components to the regression coefficient estimates. Unlike the related approach of *Hannart* [2016], we stay within the commonly used framework of empirical orthogonal functions (EOFs). We address the issue of using a specific truncation of the number of EOFs by choosing a reasonable range for the number of EOFs and applying Bayesian model averaging (BMA) to infer the individual probabilities among this range of truncations. Assessing a large number, or all possible numbers, of EOFs is feasible, because the implementation of our method is computationally very efficient and the BMA framework lends itself to parallelization.

2. The Model

2.1. Regression-Based Climate Change Detection and Attribution

From a statistical point of view, climate change detection and attribution can be viewed as a multivariate spatial or spatiotemporal regression problem. We consider temperatures over the entire globe, divided into fairly large grid cells (e.g., a $5^\circ \times 5^\circ$ grid, resulting in $72 \times 36 = 2592$ grid cells). Let $\mathbf{y} = (y_1, \ldots, y_n)'$ denote a vector representing the true temperature changes in the n grid cells. For example, $\mathbf{y}$ could contain the slope coefficients in linear regressions of temperature on time for some time period under investigation. Quantifying temperature change through the slope coefficient in a linear regression model is common practice, because it is the simplest and therefore most defensible way of characterizing the overall change. When the time period under consideration is long enough, cyclical climate effects (e.g., El Niño) are smoothed over and can be ignored.

Corresponding to the vector $\mathbf{y}$ of true temperature changes, we have vectors $\mathbf{x}_1, \ldots, \mathbf{x}_m$ representing the (true) temperature changes that would have happened under m different forcing scenarios. Let $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_m)$. Also, denote by $\mathbf{C}$ an $n \times n$ covariance matrix characterizing the internal climate variability (without any forcing).

We write the commonly assumed linear regression model in the form of a conditional distribution,

$$\mathbf{y}|\mathbf{X}, \boldsymbol{\beta}, \mathbf{C} \sim \mathcal{N}_n\left(\sum_{j=1}^{m} \beta_j \mathbf{x}_j, \mathbf{C}\right), \tag{1}$$

where $\mathcal{N}_n$ denotes an n variate normal or Gaussian distribution. In this regression framework, climate change detection consists, in essence, of deciding whether each of the β_j is equal to 0 or not. We assume that $\mathbf{x}_1$ corresponds to the anthropogenic forcing, and so the conclusion that $\beta_1 \neq 0$ means that climate change (in terms of linear temperature change over the time period of interest) due to human activity has been detected. Attribution goes a step further by requiring that the observations are consistent with the responses to a specific combination of forcings assuming an additive response [e.g., *Ribes et al.*, 2013]. This is equivalent to testing if the β_j are equal to unity if the mean responses for each forcing have been removed. However, there is an understanding in the literature that while the response patterns are essential, the magnitude of the response can be incorrect. Therefore, if the estimated uncertainty range does not include unity, the model response can be rescaled to match the observations if physically reasonable [e.g., *Hegerl and Zwiers*, 2011]. Note that a normal distribution is implicitly assumed in the generalized least squares algorithms used by many climate scientists for regression-based detection and attribution, because the maximum likelihood estimate for $\boldsymbol{\beta}$ under

this assumption coincides with the generalized least squares estimate. Here we take a Bayesian approach and assume independent standard normal priors for the β_j. The use of this (vaguely) informative prior is necessary for the Bayesian model averaging in section 3.3, but our simulation experiments showed that it has little effect on the posterior distribution of $\boldsymbol{\beta}$ relative to a noninformative prior.

The regression problem of inferring $\boldsymbol{\beta}$ based on the model in (1) seems very straightforward. The challenge is that in practice, $\mathbf{y}$, $\mathbf{X}$, $\boldsymbol{\beta}$, and $\mathbf{C}$ are all unknown. Because it is not sufficient to simply "plug in" point estimators for those quantities in (1) without accounting for uncertainty in their estimation [see, e.g., *Allen and Stott*, 2003], we opt for a Bayesian hierarchical model in which the components of the model are random. Such a setup allows the uncertainties from the estimation of the different model components to be propagated to the uncertainty estimates of the β_j.

2.2. Uncertainty in the "Data"

We cannot directly observe the true temperatures on which $\mathbf{y}$ is based [cf. *Berliner et al.*, 2000]. Instead, we assume that we have available an ensemble of reconstructed temperatures, from each of which we can calculate a vector of time trend coefficients corresponding to $\mathbf{y}$. Then, for an N member ensemble, we assume that

$$\mathbf{y}^{(i)} | \mathbf{y}, \mathbf{W} \overset{iid}{\sim} \mathcal{N}_n(\mathbf{y}, \mathbf{W}), \quad i = 1, \ldots, N, \tag{2}$$

where $\mathbf{W}$ is a covariance matrix describing the variability of the ensemble members around the true temperature change $\mathbf{y}$.

For simplicity and to avoid nonidentifiability issues, we let $\mathbf{W}$ be a fixed diagonal matrix, with its jth diagonal element $(\mathbf{W})_{j,j} = var(\{y_j^{(i)} : i = 1, \ldots, N\})$ equal to the sample variance of the jth elements of the ensemble vectors $\{\mathbf{y}^{(i)}\}$.

Similarly, while we do not know the true temperature change $\mathbf{x}_j$ under the jth forcing scenario, we assume that each global circulation model (GCM) output $\mathbf{x}_j^{(l)}$ can be written as the true mean temperature change due to forcing, $\mathbf{x}_j$, plus internal climate variability with covariance matrix $\mathbf{C}$, which under a normal assumption can be written as

$$\mathbf{x}_j^{(l)} | \mathbf{x}_j, \mathbf{C} \overset{ind}{\sim} \mathcal{N}_n(\mathbf{x}_j, \mathbf{C}), \qquad l = 1, \ldots, L_j, \quad j = 1, \ldots, m, \tag{3}$$

where L_j is the number of GCM runs under the jth forcing scenario. At least for $L_j = 1$, this part of our model is similar to the errors-in-variable model used by *Allen and Stott* [2003], which was extended to the multiple-GCM case by *Huntingford et al.* [2006], and for which *Hannart et al.* [2014] developed a maximum likelihood approach. For our Bayesian model, we assume independent uniform prior distributions on $\mathbb{R}$ for all elements of $\mathbf{X}$. The implications of this prior assumption will be further explored in section S2 in the supporting information.

In summary, our observations are given by the vectors

$$\mathcal{Y} = (\mathbf{y}^{(1)\prime}, \ldots, \mathbf{y}^{(N)\prime})' \qquad \text{and} \qquad \mathcal{X} = \left(\mathbf{x}_1^{(1)\prime}, \ldots, \mathbf{x}_1^{(L_1)\prime}, \ldots, \mathbf{x}_m^{(1)\prime}, \ldots, \mathbf{x}_m^{(L_m)\prime}\right)', \tag{4}$$

which are linked to the quantities in the regression model (1) via (2) and (3), respectively.

2.3. The Model for the Climate Variability

The covariance matrix $\mathbf{C}$ in (1) and (3) characterizes the internal climate variability and the variation between different GCMs. This matrix is typically estimated from "control runs," which are GCM runs without any external forcing, and thus only represent internal climate variability and GCM variability. Given L_0 control runs $\mathbf{x}_0^{(1)}, \ldots, \mathbf{x}_0^{(L_0)}$, let $\hat{\mathbf{C}}$ be the corresponding sample covariance matrix; that is, $\hat{\mathbf{C}} = \frac{1}{L_0 - 1} \sum_{l=1}^{L_0} \left(\mathbf{x}_0^{(l)} - \bar{\mathbf{x}}_0\right)\left(\mathbf{x}_0^{(l)} - \bar{\mathbf{x}}_0\right)'$, where $\bar{\mathbf{x}}_0 = \frac{1}{L_0} \sum_{l=1}^{L_0} \mathbf{x}_0^{(l)}$.

Because typically $L_0 < n$, this estimate of $\mathbf{C}$ is highly unstable, and some regularization is required. The most commonly used approach is to estimate $\mathbf{C}$ using so-called empirical orthogonal functions (EOFs) [e.g., *Santer et al.*, 1993; *Ribes et al.*, 2009] as $\mathbf{BKB}'$, where $\mathbf{B}$ is an $n \times r$ basis function matrix consisting of the first r principal components of $\hat{\mathbf{C}}$, $\mathbf{K} = \text{diag}\{e^{\hat{\lambda}_1}, \ldots, e^{\hat{\lambda}_r}\}$ is a diagonal matrix containing the corresponding eigenvalues, and typically $r \ll n$. However, the matrix $\mathbf{BKB}'$ in (5) is singular, implying that the multivariate distributions in (1) and (3) would be degenerate for $\mathbf{C} = \mathbf{BKB}'$.

To obtain a nonsingular covariance matrix, we modify this approach slightly and assume

$$\mathbf{C} = \mathbf{BKB}' + \sigma^2\mathbf{I}_n, \tag{5}$$

where $\mathbf{K} = \text{diag}\{e^{\lambda_1}, \ldots, e^{\lambda_r}\}$, and $\mathbf{I}_n$ is the $n \times n$ identity matrix. Note that the regularized estimator, $\gamma\hat{\mathbf{C}} + \sigma^2\mathbf{I}_n$, proposed by *Ribes et al.* [2009], is a special case of (5) for $r = L_0$.

We account for some of the uncertainty in estimating $\mathbf{C}$ by letting $\lambda_1, \ldots, \lambda_r$ and σ be random unknown parameters, with independent prior distributions $\lambda_j \sim \mathcal{N}(\hat{\lambda}_j, 1)$ and $\log\sigma \sim \mathcal{N}(m_r, 1)$, where $m_r = \log\sqrt{\frac{tr(\hat{\mathbf{C}}) - \sum_j \hat{\lambda}_j}{n}}$.

Usually, it is not clear how to choose r, the number of EOFs, and there is a whole range of possible values, say, $\{r_{\min}, \ldots, r_{\max}\}$. The results of the analysis might differ depending on which value of r is chosen. We explicitly take this uncertainty into account, by modeling r as an additional unknown in the model with a discrete uniform prior distribution on $\{r_{\min}, \ldots, r_{\max}\}$.

3. Inference

Given the observations $\mathcal{Y}$ and the GCM output $\mathcal{X}$ from (4), we carry out Bayesian inference on the unknown quantities in the model using Markov chain Monte Carlo (MCMC). Specifically, our interest is in obtaining the posterior distribution of the unknown parameters $\theta_r = (\boldsymbol{\beta}', \sigma, \lambda_1, \ldots, \lambda_r)'$ given the data $\mathcal{Y}$ and $\mathcal{X}$. Note that the number of parameters in θ_r depends on r, the number of EOFs, which makes inference challenging. Hence, we first describe inference on θ_r for a particular fixed value of r (i.e., conditional on r), and then we combine the results for different values of r in section 3.3. Using Bayes's Theorem, we have the conditional posterior

$$[\theta_r|\mathcal{Y}, \mathcal{X}, r] \propto [\mathcal{Y}, \mathcal{X}|\theta_r, r]\,[\theta_r|r]. \tag{6}$$

3.1. The Integrated Likelihood

Note that the "unobservable data" $\mathbf{X}$ and $\mathbf{y}$ do not appear in (6). These two quantities are often high dimensional (consisting of nm and n elements, respectively), and so it is crucial for efficient inference to analytically integrate out (i.e., marginalize over) these two quantities. Hence, our first task is to obtain the so-called integrated likelihood,

$$[\mathcal{Y}, \mathcal{X}|\theta_r, r] = [\mathcal{Y}|\mathcal{X}, \theta_r, r]\,[\mathcal{X}|\theta_r, r], \tag{7}$$

where

$$[\mathcal{X}|\theta_r, r] = \int \prod_{j=1}^{m}\prod_{l=1}^{L_j} \mathcal{N}_n\left(\mathbf{x}_j^{(l)}|\mathbf{x}_j, \mathbf{C}\right)[\mathbf{x}_j]\mathrm{d}\mathbf{X} \propto \prod_{j=1}^{m}\prod_{l=1}^{L_j}\int \mathcal{N}_n\left(\mathbf{x}_j|\mathbf{x}_j^{(l)}, \mathbf{C}\right)\mathrm{d}\mathbf{x}_j = 1.$$

To obtain $[\mathcal{Y}|\mathcal{X}, \theta_r, r]$ in (7), first note that we have

$$\text{vec}(\mathbf{X}) \mid \mathcal{X}, \theta_r, r \sim \mathcal{N}_{nm}\left(\text{vec}(\bar{\mathbf{X}}), \mathbf{L}^{-1} \otimes \mathbf{C}\right),$$

where the vec() operator stacks the columns of a matrix into a vector, $\mathbf{L} = \text{diag}\{L_1, \ldots, L_m\}$, $\otimes$ denotes the Kronecker product, and $\bar{\mathbf{X}}$ is the matrix with jth column $\sum_{l=1}^{L_j}\mathbf{x}_j^{(l)}/L_j$. Then, one can integrate out $\mathbf{X}$ by applying the laws of total expectation and variance to obtain

$$\mathbf{y} \mid \mathcal{X}, \theta_r, r \sim \mathcal{N}_n\left(\bar{\mathbf{X}}\boldsymbol{\beta}, g(\boldsymbol{\beta})\mathbf{C}\right),$$

where $g(\boldsymbol{\beta}) = 1 + \sum_{j=1}^{m}\beta_j^2/L_j$. Finally, if we also integrate out $\mathbf{y}$, we have

$$\mathcal{Y} \mid \mathcal{X}, \theta_r, r \sim \mathcal{N}_{nN}\left((\mathbf{1}_N \otimes \bar{\mathbf{X}})\boldsymbol{\beta}, \boldsymbol{\Sigma}_{\mathcal{Y}}\right), \tag{8}$$

where $\boldsymbol{\Sigma}_{\mathcal{Y}} = \mathbf{J}_N \otimes g(\boldsymbol{\beta})\mathbf{C} + \mathbf{I}_N \otimes \mathbf{W}$, $\mathbf{J}_N = \mathbf{1}_N\mathbf{1}_N'$, and $\mathbf{1}_N$ is an N-dimensional vector of ones.

3.2. MCMC for Fixed Number of EOFs

Based on the integrated likelihood in (7), we then sample from the posterior of θ_{r_i} in (6) for a particular value $r = r_i$ using Markov chain Monte Carlo (MCMC), resulting in samples $\theta_{r_i}^{(1)}, \ldots, \theta_{r_i}^{(M)}$. In section S1 [*Haario et al.*, 2001; *Harville*, 1997; *Sherman and Morrison*, 1950; *Woodbury*, 1950; *Henderson and Searle*, 1981], we describe the details of the algorithm, including how to quickly evaluate the integrated likelihood. While this likelihood involves the very large $nN \times nN$ matrix $\boldsymbol{\Sigma}_{\mathcal{Y}}$, we derive an expression of the likelihood that only involves determinants and inverses of (small) $r \times r$ and diagonal $n \times n$ matrices.

3.3. Bayesian Model Averaging Over the Number of EOFs

Once samples from the posterior for each value of $r \in \{r_{\min}, \ldots, r_{\max}\}$ have been obtained as described in section 3.2, we can perform Bayesian model averaging (BMA) [e.g., *Hoeting et al.*, 1999] to average the posterior results for different values of r using weights (i.e., probabilities) automatically chosen by the data.

The posterior of $\boldsymbol{\beta}$ averaged over the posterior of r (i.e., taking the uncertainty about the value of r into account) is given by

$$[\boldsymbol{\beta}|\mathcal{Y}, \mathcal{X}] = \sum_{i=\min}^{\max} [\boldsymbol{\beta}|r_i, \mathcal{Y}, \mathcal{X}]\, P(r = r_i|\mathcal{Y}, \mathcal{X}),$$

where $[\boldsymbol{\beta}|r_i, \mathcal{Y}, \mathcal{X}]$ is the posterior distribution of $\boldsymbol{\beta}$ for the model with fixed $r = r_i$ from section 3.2. Further, due to the discrete uniform prior on r, the posterior probability of $r = r_i$ is given by $P(r = r_i|\mathcal{Y}, \mathcal{X}) \propto [\mathcal{Y}, \mathcal{X}|r_i]\, P(r = r_i) \propto [\mathcal{Y}, \mathcal{X}|r_i]$.

Hence, the weights for each r_i are determined solely by the so-called marginal likelihood $[\mathcal{Y}, \mathcal{X}|r_i]$, which integrates out the parameters $\boldsymbol{\theta}_{r_i}$. Fortunately, it turns out that an estimate of this quantity [e.g., *Newton and Raftery*, 1994] can be obtained using the evaluations of the likelihood already performed in the MCMC procedure (i.e., without any significant additional calculations), as

$$[\mathcal{Y}, \mathcal{X}|r_i] = \left(\frac{1}{M} \sum_{j=1}^{M} \frac{1}{[\mathcal{Y}, \mathcal{X}|r_i, \boldsymbol{\theta}_{r_i}^{(j)}]} \right)^{-1}$$

where $\boldsymbol{\theta}_{r_i}^{(j)}$ is the jth MCMC sample (posterior draw) of the parameter vector from section 3.2, and $[\mathcal{Y}, \mathcal{X}|r_i, \boldsymbol{\theta}_{r_i}^{(j)}]$ is the likelihood in (7) for $r = r_i$.

3.4. Residual Consistency Test

A residual consistency test is often used to determine whether the assumed model is consistent with the observations, or whether the observations cannot be fully explained by the forcings under consideration. Such a test arises naturally here as a Bayesian goodness-of-fit test. Consider the quadratic form

$$q(\boldsymbol{\theta}_r, r) = \mathrm{vec}(\tilde{\mathbf{Y}})' \boldsymbol{\Sigma}_{\mathcal{Y}}^{-1} \mathrm{vec}(\tilde{\mathbf{Y}}),$$

where $\widetilde{\mathbf{Y}}$ is an $n \times N$ matrix with kth column $\mathbf{y}^{(i)} - \bar{\mathbf{X}}\boldsymbol{\beta}$. Then, under the null hypothesis that the observations actually do come from the assumed model, we have $q(\boldsymbol{\theta}_r, r) \sim \chi^2_{nN}$ when evaluated at a random draw from the posterior distribution $[\boldsymbol{\theta}_r, r|\mathcal{Y}, \mathcal{X}]$ [*Jun et al.*, 2014]. As we have already obtained $q(\boldsymbol{\theta}_{r_i}^{(j)}, r_i)$ as part of our MCMC procedure (see section S1), we can obtain a posterior distribution of test statistics for our goodness-of-fit test at virtually no additional cost. Straightforwardly converting each test statistic to a p value, we obtain a posterior distribution of p values. If this distribution is close to 0, there is evidence of model misfit. If it is close to 1, no such evidence exists.

4. Illustration

We applied our method to simulated and real data. The results of the simulation are summarized in section S2. Here we illustrate our methodology on monthly means of zonally averaged tropospheric temperature (TLT) from a previously investigated data set [*Santer et al.*, 2013a, 2013b].

4.1. Observational and Climate Model Data

We use observational data from satellite-based microwave sounding units (MSUs) from Remote Sensing Systems (RSS) for the 27 year period from January 1979 to December 2011. To account for various uncertainties associated with the sampling and the retrieval, the data are provided as a 396-member ensemble of observations [*Mears et al.*, 2011]. We consider this ensemble to be noisy samples of the unknown true temperatures and use it to represent the observational uncertainty as in (2).

For climate model data, we selected historical runs with natural-only (NAT) and anthropogenic-only (ANT) forcings from the Coupled Model Intercomparison Project Phase 5 (CMIP5) archive. Some of the CMIP5 simulations were vertically weighted to be similar to the MSU observations [*Santer et al.*, 2013a], and we based our selection on the availability of those zonally averaged MSU-comparable temperatures and earlier scientific findings to exclude problematic runs [*Santer et al.*, 2013b]. This resulted in four natural forcings-only model runs (bcc csm1, can esm2, giss e2 r p1, and giss e2 r p3) and three anthropogenic forcings-only runs (gfdl, giss e2 r p1, and giss e2 r p3), shown in Figures S1 and S2, respectively. As controls, we use the preindustrial control

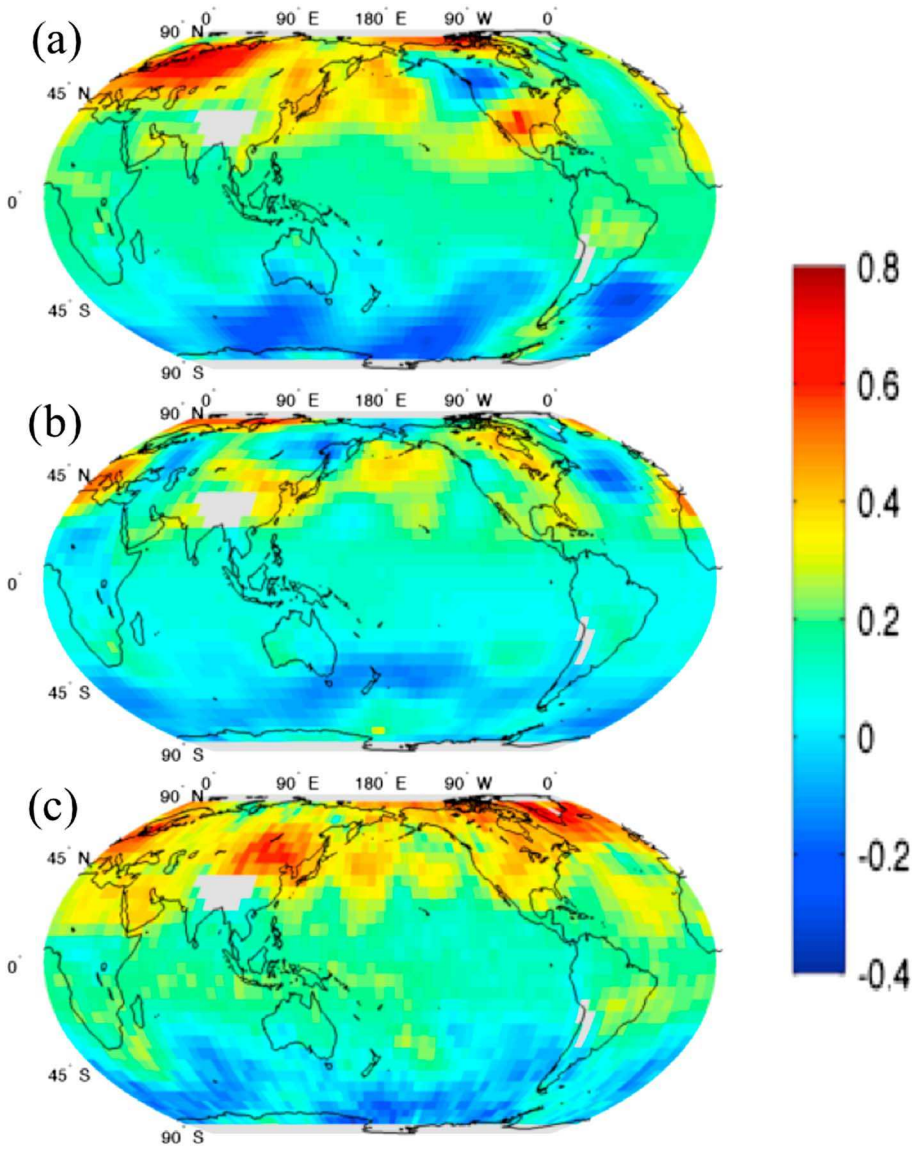

Figure 1. Linear tropospheric temperature trend (in °C per decade) for the period 1979–2005 for the giss e2 r p1 model run using (a) only anthropogenic forcing and (b) only natural forcing, and (c) the average of 396 observational ensemble members from RSS. Areas in grey were excluded from the analysis due to the absence of satellite observations.

run giss e2 r p1 from the CMIP5 archive. This run spans 550 years, and we split the run in 27 year segments, which resulted in 20 such segments.

We use the built-in MATLAB function "`geoloc2grid`" to convert all climate model and observational data from their native resolution to a 5° × 5° grid. We limit the domain to 80° north and −70° south, and we further exclude locations with an altitude higher than 3000 m as there is no satellite data coverage for these locations. The resulting fields consist of $n = 2107$ grid cells, for which we calculate the linear slope coefficients used in the analysis. Figure 1 shows the spatial field of coefficients for one example of the anthropogenic forcings-only run, one example of the natural forcings-only run, and the mean field of the observational ensemble.

4.2. Implementation and Results

We implement the model in MATLAB 2014b using 20,000 MCMC iterations for each value of $r \in \{1, \dots, 19\}$, 5000 of which are considered burn-in. This setup provides good convergence properties of the Markov chains. Figure 2a shows the posterior distributions of β_1 (anthropogenic forcing) and β_2 (natural forcing) for each truncation, $r \in \{1, \dots, 19\}$, individually, while Figure 2c shows the corresponding Bayesian model averaged estimates. According to standard interpretation, both forcings are detected, as the 95% credible intervals do

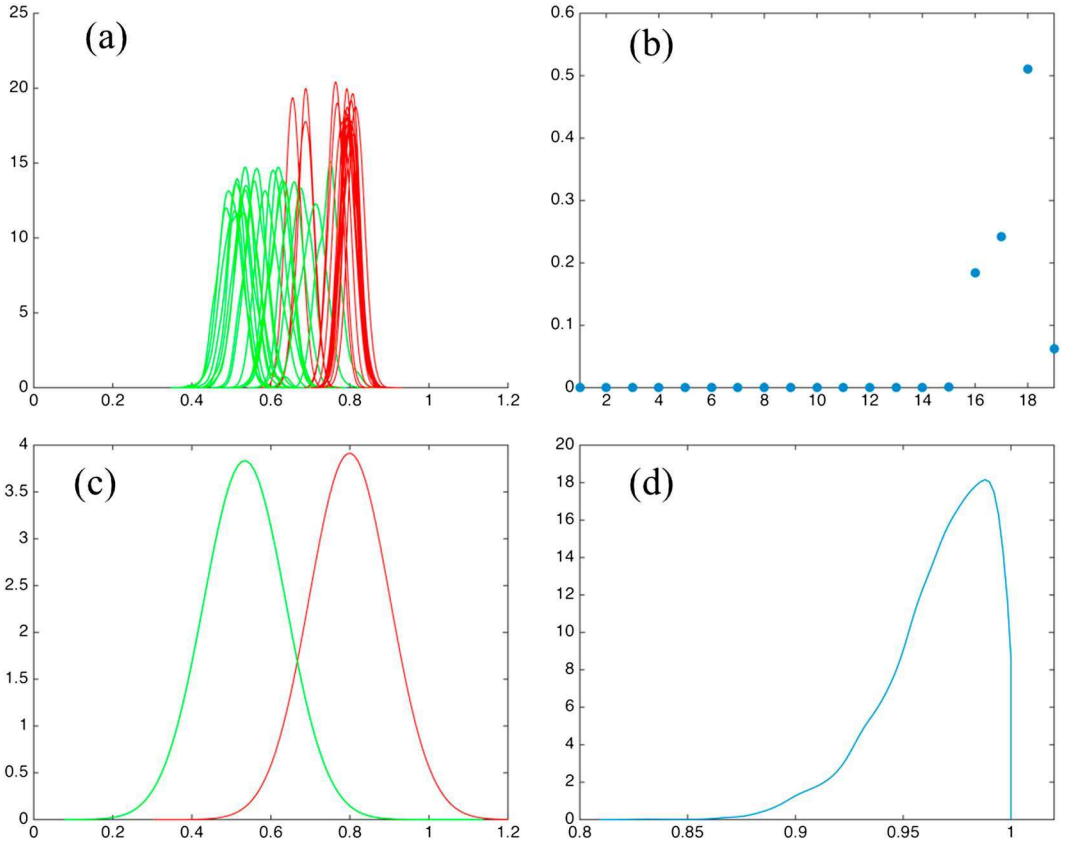

Figure 2. Results for the tropospheric temperature data analysis. (a) Posterior distributions of β_1 (anthropogenic forcing) in red and β_2 (natural forcing) in green for each EOF truncation $r \in \{1, \ldots, 19\}$. (b) Posterior probabilities for different values of r. (c) Marginal posterior distributions of β_1 and β_2 obtained by Bayesian model averaging over r. (d) Posterior distribution of p values for the residual consistency test.

not contain zero. The coefficient corresponding to the anthropogenic forcing is of larger magnitude and hence potentially stronger than the natural forcing. The averaged estimates are obtained using the posterior probabilities for each truncation option shown in Figure 2b. For this application, most of the weight, slightly more than 50%, is placed on the truncation using 18 EOFs. The difference between the widths of the posterior distributions for the individual truncations (Figure 2a) and the distributions based on BMA (Figure 2c) illustrates how taking the uncertainty about the value of r into account leads to broader, and arguably more realistic, credible intervals on the coefficients. Applying the residual consistency test (described in section 3.4), there is no evidence of a model misfit as the distribution of p values is close to 1 (Figure 2d).

4.3. Computational Considerations

One of the challenges with Bayesian methods, due to their requirement for repeated, iterative sampling, is often their computational speed, which can be prohibitively slow for some applications. For our model, the derivations in section 3.2 result in short computation times for the MCMC for each possible EOF truncation r. In addition, these MCMC analyses for different r can be executed completely in parallel. On a moderate laptop, a MacBook Pro with an Intel quad-core 2.6 GHz i7 processor and 8 GB of memory, using MATLAB 2014b and four cores, the algorithm takes less than 8 min to evaluate all 19 truncation options and perform Bayesian model averaging to obtain the results shown in Figure 2. On clusters or high-performance computing systems, where a large number of cores is available, the algorithm can easily be scaled up to a higher number of EOF truncations if longer control runs are available.

5. Discussion and Conclusion

We have introduced a Bayesian hierarchical model that addresses some of the open methodological questions in regression-based detection and attribution. Specifically, we take into account the uncertainties related to

observations and climate signals under different forcing scenarios. Further, we incorporate the uncertainty associated with estimating the climate variability covariance matrix into the inference on the regression coefficients. Instead of choosing a specific truncation of the number of empirical orthogonal functions (EOFs) in this covariance matrix, we apply Bayesian model averaging to assign optimal probabilistic weights to a range of possible truncations. The method is computationally efficient and scalable due to its parallel nature, which we have demonstrated with a substantive application.

A broader issue that we have not addressed, but which is not specific to our methodology, is the sensitivity to the choice of control runs. Our inferential scheme is still conditional on a given set of control runs, and we are not accounting for the uncertainty stemming from choosing these runs. In particular, the estimation of the EOF vectors based on a small number of control runs can be unstable and sensitive to the specific choice of runs. One avenue for further research would be to regularize the EOF estimation [e.g., *Wang and Huang*, 2017] or to treat the EOFs as random fields and account for their uncertainty [e.g., *Suarez and Ghosal*, 2017].

We have further assumed a relatively simple form for the covariance matrix associated with the observation ensemble, namely, a fixed diagonal matrix with the empirical variances as its elements. We could allow this matrix to be of a more general form (e.g., based on a Gaussian Markov random field) with potentially unknown (random) parameters. We have also assumed the different climate model runs to be independent replicates of the same random quantity; for situations where multiple ensemble members are available from different models, we could improve the setup by modeling intermodel and intramodel variability.

In additional future work, we plan on conducting a comprehensive performance evaluation using extensive simulations. This evaluation will extend our preliminary simulations in section S2, for example, by considering the effect of the number of control runs and shrinkage of the regression coefficients for small grid size.

Acknowledgments

This material was based upon work partially supported by the National Science Foundation under grant DMS-1127914 to the Statistical and Applied Mathematical Sciences Institute and by the U.S. National Science Foundation (NSF) Research Network on Statistics in the Atmosphere and Ocean Sciences (STATMOS) through grant DMS–1106862. Katzfuss was also partially supported by NASA's Earth Science Technology Office AIST-14 program and by NSF grant DMS–1521676 and NSF CAREER grant DMS–1654083. Smith was also partially supported by NSF grant DMS–1242957. We would like to acknowledge high-performance computing support from Yellowstone (ark:/85065/d7wd3xhc) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation. We would like to thank Ben Santer for generously providing the data and many helpful comments. We would further like to thank Peter Thorne, Philippe Naveau, James Long, and two anonymous reviewers for helpful comments and discussions that greatly improved the manuscript. All the data and software used can be freely obtained by emailing the corresponding author.

References

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

Berliner, L. M., R. A. Levine, and D. J. Shea (2000), Bayesian climate change assessment, *J. Clim.*, *13*, 3805–3820.

Bindoff, N., et al. (2013), *Detection and Attribution of Climate Change: From Global to Regional*, book section 10, pp. 867–952, Cambridge Univ. Press, Cambridge, U. K., and New York, doi:10.1017/CBO9781107415324.022.

Haario, H., E. Saksman, and J. Tamminen (2001), An adaptive Metropolis algorithm, *Bernoulli*, *7*(2), 223–242.

Hannart, A. (2016), Integrated optimal fingerprinting: Method description and illustration, *J. Clim.*, *29*(6), 1977–1998, doi:10.1175/JCLI-D-14-00124.1.

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

Harville, D. A. (1997), *Matrix Algebra From a Statistician's Perspective*, Springer, New York.

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

Hegerl, G., F. W. Zwiers, P. Braconnot, N. Gillett, Y. Luo, J. Marengo Orsini, N. Nicholls, J. Penner, and P. Stott (2007), *Understanding and Attributing Climate Change*, pp. 663–745, book section 9, Cambridge Univ. Press, Cambridge, U. K., and New York.

Hegerl, G. C., H. von Storch, K. Hasselmann, B. D. Santer, U. Cubasch, and P. D. Jones (1996), Detecting greenhouse-gas-induced climate change with an optimal fingerprint method, *J. Clim.*, *9*(10), 2281–2306.

Henderson, H., and S. Searle (1981), On deriving the inverse of a sum of matrices, *Soc. Industrial Appl. Math. Rev.*, *23*(1), 53–60.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999), Bayesian model averaging: A tutorial, *Stat. Sci.*, *14*(4), 382–417.

Huntingford, C., P. A. Stott, M. R. Allen, and F. H. Lambert (2006), Incorporating model uncertainty into attribution of observed temperature change, *Geophys. Res. Lett.*, *33*(5), L05710, doi:10.1029/2005GL024831.

Jun, M., M. Katzfuss, J. Hu, and V. E. Johnson (2014), Assessing fit in Bayesian models for spatial processes, *Environmetrics*, *25*(8), 584–595, doi:10.1002/env.2315.

Lott, F. C., P. A. Stott, D. M. Mitchell, N. Christidis, N. P. Gillett, L. Haimberger, J. Perlwitz, and P. W. Thorne (2013), Models versus radiosondes in the free atmosphere: A new detection and attribution analysis of temperature, *J. Geophys. Res. Atmos.*, *118*, 2609–2619, doi:10.1002/jgrd.50255.

Mears, C. A., F. J. Wentz, P. Thorne, and D. Bernie (2011), Assessing uncertainty in estimates of atmospheric temperature changes from MSU and AMSU using a Monte-Carlo estimation technique, *J. Geophys. Res.*, *116*, D08112, doi:10.1029/2010JD014954.

Newton, M. A., and A. E. Raftery (1994), Approximate Bayesian inference with the weighted likelihood bootstrap, *J. R. Stat. Soc. B*, *56*(1), 3–48.

Ribes, A., J.-M. Azaïs, and S. Planton (2009), Adaptation of the optimal fingerprint method for climate change detection using a well-conditioned covariance matrix estimate, *Clim. Dyn.*, *33*, 707–722, doi:10.1007/s00382-009-0561-4.

Ribes, A., S. Planton, and L. Terray (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.

Santer, B., T. Wigley, and P. Jones (1993), Correlation methods in fingerprint detection studies, *Clim. Dyn.*, *8*, 265–276.

Santer, B. D., et al. (2013a), Identifying human influences on atmospheric temperature, *Proc. Natl. Acad. Sci. U.S.A.*, *110*(1), 26–33, doi:10.1073/pnas.1210514109.

Santer, B. D., et al. (2013b), Human and natural influences on the changing thermal structure of the atmosphere, *Proc. Natl. Acad. Sci. U.S.A.*, *110*(43), 17,235–17,240, doi:10.1073/pnas.1305332110.

Sherman, J., and W. Morrison (1950), Adjustment of an inverse matrix corresponding to a change in one element of a given matrix, *Ann. Math. Stat.*, *21*(1), 124–127.

Suarez, A. J., and S. Ghosal (2017), Bayesian estimation of principal components for functional data, *Bayesian Anal.*, *12*(2), 311–333.
Wang, W.-T., and H.-C. Huang (2017), Regularized principal component analysis for spatial data, *J. Comput. Graph. Stat.*, *26*(1), 14–25, doi:10.1080/10618600.2016.1157483.
Woodbury, M. (1950), Inverting modified matrices, Memorandum Rep. 42, Statistical Research Group, Princeton Univ.