An Objective Bayesian Improved Approach for Applying Optimal Fingerprint Techniques to Estimate Climate Sensitivity*

NICHOLAS LEWIS

Bath, United Kingdom

(Manuscript received 23 July 2012, in final form 15 March 2013)

ABSTRACT

A detailed reanalysis is presented of a ''Bayesian'' climate parameter study (as exemplified by Forest et al.) that estimates climate sensitivity (ECS) jointly with effective ocean diffusivity and aerosol forcing, using optimal fingerprints to compare multidecadal observations with simulations by the Massachusetts Institute of Technology 2D climate model at varying settings of the three climate parameters. Use of improved methodology primarily accounts for the 90% confidence bounds for ECS reducing from 2.1–8.9 K to 2.0–3.6 K. The revised methodology uses Bayes's theorem to derive a probability density function (PDF) for the whitened (made independent using an optimal fingerprint transformation) observations, for which a uniform prior is known to be noninformative. A dimensionally reducing change of variables onto the parameter surface is then made, deriving an objective joint PDF for the climate parameters. The PDF conversion factor from the whitened variables space to the parameter surface represents a noninformative joint parameter prior, which is far from uniform. The noninformative prior prevents more probability than data uncertainty distributions warrant being assigned to regions where data respond little to parameter changes, producing better-constrained PDFs. Incorporating 6 years of unused model simulation data and revising the experimental design to improve diagnostic power reduces the best-fit climate sensitivity. Employing the improved methodology, preferred 90% bounds of 1.2–2.2 K for ECS are then derived (mode and median 1.6 K). The mode is identical to those from Aldrin et al. and [using the same Met Office Hadley Centre Climate Research Unit temperature, version 4 (HadCRUT4), observational dataset] from Ring et al. Incorporating nonaerosol forcing and observational surface temperature uncertainties, unlike in the original study, widens the 90% range to 1.0–3.0 K.

1. Introduction

Key climate system parameters—in particular equilibrium climate sensitivity (S_{eq}), effective vertical deep-ocean diffusivity (K_v), and total aerosol forcing (F_{aer})—are often estimated by studies (usually formulated in Bayesian terms) that compare simulations by adjustable parameter climate models with observations. Examples of such studies include Forest et al. (2000, 2001, 2002, 2006, 2008, hereafter F00, F01, F02, F06, and F08, respectively; collectively the Forest studies), Andronova and Schlesinger (2001), Frame et al. (2005), Hegerl et al. (2006), Knutti et al. (2002), Sansó et al. (2008, hereafter SFZ08), Sansó and Forest (2009, hereafter SF09), Aldrin et al. (2012), and Ring et al. (2012). These studies provided six of the eight probability density functions (PDFs) given in the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4) for equilibrium climate sensitivity inferred from observed changes in climate [Hegerl et al. (2007), see their appendix 9.B for an explanation of such studies].

The Forest studies are excellent examples since they used a wide spread of instrumental observations and avoided dependence on existing ill-constrained estimates of K_v and F_{aer} by jointly estimating them with S_{eq}. Using for this reason F06 as a starting point, this paper derives and implements an improved, objective Bayesian, methodology, which provides better defined PDFs for S_{eq} in particular. Further, by taking advantage of the final 6 years of model simulation data, unused in F06, and revising the experimental design to improve diagnostic power, an updated closely constrained estimate for S_{eq} is obtained.

F06 and similar studies involve comparisons between observed temperatures at various spatiotemporal

* Supplemental information related to this paper is available at the Journals Online website: http://dx.doi.org/10.1175/JCLI-D-12-00473.s1.

Corresponding author address: Nicholas Lewis, Walden, Widcombe Hill, Bath, United Kingdom.
E-mail: nhlewis@btinternet.com

DOI: 10.1175/JCLI-D-12-00473.1

coordinates and climate model simulations. The models have adjustable calibrated parameters controlling key climate properties and are more suited than atmosphere–ocean general circulation models (AOGCMs) for exploring the entire parameter space and running multiple simulations at varying parameter settings.

F06 used the Massachusetts Institute of Technology (MIT) 2D climate model (2DCM) (Sokolov and Stone 1998; F06). Forcings from all greenhouse gases (specified explicitly), sulfate aerosols (calculated from emissions), stratospheric and tropospheric ozone, land use, solar irradiance, and volcanism (specified as stratospheric aerosol optical depth) (collectively referred to as GSOLSV) were included: see the F06 auxiliary material for details. To place F06 in context, it represented an update of F02 using more comprehensive forcings; F08 used the same methods and data as F06 except for its model simulations, which employed a later version of the MIT 2DCM, while Libardoni and Forest (2011) investigated sensitivities to the surface temperature dataset. Drignei et al. (2008) used a statistical model as a nonlinear regression surrogate to estimate the same parameters using F02's data and alternative correlation structures. SFZ08 and SF09 used F06 data but employed more complex hierarchical Bayesian methods, unlike the approaches used in F06 and this paper, which differ principally in the prior distribution used. The SFZ08 and SF09 posterior PDFs for S_{eq} and $\sqrt{K_v}$ substantially reflected the chosen priors and, using uniform priors, were poorly constrained.

F06 used three "diagnostics" (groups of variables whose observed values are compared to model simulations):

- Surface-air temperatures [surface (sfc)]: These are four equal-area latitude averages for each of the 5 decades comprising 1946–95, referenced to 1905–95 climatology (Jones et al. 1999).
- Deep-ocean temperatures [deep ocean (do)]: This is the trend in global mean 0–3-km-deep-layer pentadal averages ending in 1959–95 (Levitus et al. 2005).
- Upper-air temperatures [upper air (ua)]: These are the differences between 1986–95 and 1961–80 averages at eight standard pressure levels from 850 to 50 hPa on a 5° grid (Parker et al. 1997).

Simulations were run from 1860 to 2001 using 499 parameter combinations, with S_{eq} ranging from 0.5 to 15 K, K_v from 0 to 64 cm^2 s^{-1}, and F_{aer} from -1.5 to $+0.5$ W m^{-2}. (Units for these parameters are generally omitted from here on.) Cases with $S_{\mathrm{eq}} > 10$ were discarded. For estimation, K_v was parameterized as its square root (Sokolov et al. 2003), ocean heat uptake being proportional thereto. The term F_{aer} represents net forcing (direct and indirect) during the 1980s relative to pre-1860 levels and implicitly includes omitted forcings with patterns similar to those of sulfate aerosols. Means of four-member initial condition ensembles were used at each parameter combination to reduce the impact of internal variability. The diagnostics ended in 1995, matching F02 to enable the effects of including a more complete set of forcings to be illustrated, so the final six simulation years were not used. Reference should be made to F06 and F02 for a fuller description of the MIT2DCM and simulation runs, applied climate forcings, and methods used.

F06 uses an optimal fingerprint method involving comparing the modeled $\mathbf{T}_m(\boldsymbol{\theta}_m)$ and observed $\tilde{\mathbf{T}}_o$ spatiotemporal patterns of temperature change for each diagnostic, where $\boldsymbol{\theta} = \{S_{\mathrm{eq}}, \sqrt{K_v}, F_{\mathrm{aer}}\}$ represents the three climate parameters being estimated. These are idealized parameters and, in practice, S_{eq} changes somewhat with time as additional feedbacks are activated. Simulation variability and model error are ignored.

AOGCM control run data provide an estimate of the natural variability (climate noise) covariance matrix $\mathbf{C}_N$ for each diagnostic. The estimated noise covariance matrix $\hat{\mathbf{C}}_N$ will be inaccurate, since AOGCM control runs are of limited length and imperfectly simulate natural variability. Regularization is applied to obtain an approximate but more robust inverse covariance matrix estimate. The usual truncated eigen-decomposition regularization method is employed in F06, with only the largest κ eigenfunctions (EOFs), or modes of variability, being retained in the estimate, $\hat{\mathbf{C}}_N(\kappa)$. Sensitivity to the truncation parameter is a known problem with optimal fingerprint methods (Allen and Tett 1999, hereafter AT99).

Measurement error for the surface and upper-air diagnostics is small compared to estimated internal climate variability and is therefore ignored. For the univariate deep-ocean diagnostic, neither observational measurement error nor control run variance dominates. The two variances are added to give $\hat{\mathbf{C}}_N$, κ being irrelevant here since $\hat{\mathbf{C}}_N$ has only one element.

For each diagnostic, goodness-of-fit statistics $r^2(\boldsymbol{\theta}_m, \tilde{\mathbf{T}}_o) = [\mathbf{T}_m(\boldsymbol{\theta}_m) - \tilde{\mathbf{T}}_o]^{\mathrm{T}} \hat{\mathbf{C}}_N^{-1}(\kappa)[\mathbf{T}_m(\boldsymbol{\theta}_m) - \tilde{\mathbf{T}}_o]$ are computed. From r^2 F06 compute a joint likelihood $p(\tilde{\mathbf{T}}_o \mid \boldsymbol{\theta}_m)$ for each $\mathbf{T}_m(\boldsymbol{\theta}_m)$, representing the relative probability of that diagnostic's observations as a function of the candidate parameter value $\boldsymbol{\theta}_m$. The likelihood function is based on Δr^2, the excess of $r^2(\boldsymbol{\theta}, \tilde{\mathbf{T}}_o)$ over the minimum r^2 value, having a $mF_{m,\nu}$ distribution: m being the number of parameters being estimated and ν the degrees of freedom (DF) available for estimating $\mathbf{C}_N$ (see supplemental material for additional discussion). The minimum r^2 is checked for consistency with the errors being generated by internal variability.

F06 uses a Bayesian paradigm, whereby probability distributions can be estimated for unknown parameters.

Applying Bayes's theorem, F06 derives a joint posterior PDF for the parameter vector $\boldsymbol{\theta}$ as the normalized product of the likelihood function from one diagnostic and a ''prior'' probability distribution consisting of the product of separate uniform priors for each parameter. Bayes's theorem is then applied twice more, each time multiplying the previous posterior PDF for $\boldsymbol{\theta}$, used as the prior, by the likelihood function from another diagnostic to obtain an updated posterior PDF. Marginal PDFs for individual climate system parameters are obtained by integrating out the other parameters from the final joint parameter posterior PDF. Although such PDFs may appear to provide precise probabilistic information, they are perhaps better viewed as indicating how likely it is that any chosen range brackets the parameter value.

Inferences as to climate system parameter values will be affected by the data used. The choice of data (and parameterizations) may be guided by the physics of the climate system but remains somewhat subjective. There is merit in seeking data that well constrain parameter values, but there may be a trade-off with data quality and coverage. Surface, upper-air, and/or deep-ocean temperature data spanning much of the instrumental period are typically used. Using other data types or periods may give rise to significantly different parameter estimates, as can detailed data processing choices. The selection of data types to compare to model simulations is a key decision; in this paper we work from the F06 data choices.

2. Objective Bayesian inference

In order for Bayesian inference to reflect, insofar as possible, only the data from which it is derived—as is appropriate when reporting objectively stand-alone scientific results—a noninformative prior must be used (Bernardo and Smith 1994; Kass and Wasserman 1996). If the prior is informative, and not overwhelmed by the data, a Bayesian posterior density is unlikely to approximate the density arising if, hypothetically, the experiment(s) concerned were to be repeated indefinitely, so valid frequentist confidence intervals cannot be estimated. Since the available data are insufficient to constrain climate parameters narrowly, an informative prior exerts strong influence.

Typically, scientific studies not expressed in Bayesian terms implicitly use a noninformative prior when considered from a Bayesian perspective. That occurs when observables are sampled according to their probability distributions. Gregory et al. (2002) estimated the change in global mean temperature between two periods and the corresponding change in forcing net of ocean heat uptake. They repeatedly sampled the probability distributions for those estimated changes, assuming independent normal error distributions, to derive a PDF for the ratio of the changes, which gives S_{eq}. The resulting PDF effectively embodies a noninformative prior. Similarly, Forster and Gregory (2006) diagnosed S_{eq} by ordinary least squares regression of changes in forcing net of radiative flux on changes in surface temperature, all errors being assumed Gaussian. They stated that this was equivalent to assuming uniform priors in the data (observables), which is noninformative given Gaussian errors.

Explicitly Bayesian climate sensitivity studies have commonly used uniform priors, or sometimes deliberatively informative ''expert'' priors, for the parameters being estimated. Frame et al. (2005) advocated sampling a flat prior distribution in S_{eq} if that is the target of the estimate and did not mention noninformative priors, but nevertheless derived PDFs for S_{eq} using various different sampling methods. Their uniform sampling-of-observables method, which they stated was appropriate for and gave an objective range for S_{eq} relevant to twenty-first-century warming forecasts, is very similar to the one we propose, although their implementation requires equal numbers of observables and parameters, and likelihood skewness may mean the prior involved was only approximately noninformative. Pueyo (2012) asserted that the problems of estimating S_{eq} and its reciprocal, the climate feedback parameter, were equivalent, and hence their priors should have the same form, implying a uniform-in-log(S_{eq}) prior. However, it is not clear that in practice the two problems have the same characteristics. In any event, Pueyo's arguments are not applicable where, as here, the prior is for jointly estimating S_{eq} and other parameters and is a function of all those parameters.

When data affected by random errors bear strongly nonlinear relationships to parameters upon which the (hypothetical) true data values depend, as with S_{eq} and K_v in particular, a uniform prior is only noninformative if applied to the true *data*, the likelihood functions for which are of known form, centered on the observations (Box and Tiao 1973). Applied to the model *parameters*, as in F06, a uniform prior will be informative and may lead to substantially erroneous estimated parameter PDFs. Many parameterizations are possible. In this paper, a uniform prior is applied to the true data, a parameterization in which a uniform prior is noninformative.

Correct inference about parameters that have nonlinear relationships with data is impossible when, as in F06, only a sum of squared whitened differences (r^2) is computed. The sum of squares suffices to derive a probability density in data space but not to map that density into a density in parameter space. To estimate objectively a joint probability density for the parameters, the relationship between natural volume elements in data space and in parameter space must be computed. Determining a noninformative prior for the parameters is intimately bound up with that metric relationship (Kass 1989). Computing it requires information on how each whitened difference changes with model parameter values. Without such information, there is no way of correctly allocating probability mass between different locations in parameter space with identical sums of squared differences. For example, if the position of a point in 3D relative to an origin is measured with error in Cartesian coordinates (x, y, z) but the parameters to be estimated are distance, azimuth, and elevation (r, ϕ, θ) (spherical coordinates), knowing $(x^2 + y^2 + z^2)$ provides no information as to ϕ or θ, and the relationship between the data space and parameter space volume elements $dxdydz$ and $drd\phi d\theta$ varies with location.

Furthermore, Bayesian updating cannot provide objective inference when the data used in deriving the posterior to be updated (forming the prior) and the data from which the updating likelihood function is derived have differing nonlinear relationships with the parameters, as the sets of diagnostic data in F06 do. In that case, the noninformative parameter priors required for objective Bayesian inference from the two datasets individually would differ. Using the appropriate individually noninformative prior, Bayesian updating would produce a different result according to the order in which Bayes's theorem was applied to the two datasets (see supplemental material for additional discussion). That noninformative priors and Bayesian updating conflict is a known problem (Kass and Wasserman 1996).

Noninformative priors vary with the experiment involved; they cannot be directly interpreted in probabilistic terms (Bernardo and Smith 1994). In order for the posterior to be dominated to the greatest possible extent by the data, however weak, and thereby for the prior to convey no particular knowledge as to the parameters, a noninformative prior must differ according to what nonlinear relationships the data has with the parameters.

3. Data

a. Data sources

The raw surface and upper-air observational datasets employed in F06 were revised subsequent to use in F06 and are no longer online. Only partial AOGCM control run data still exist. We were initially unable to obtain original data for F06 and instead obtained archives of processed diagnostic data for two related studies: Curry et al. (2005, hereafter CSF05) and SFZ08. F06 stated that CSF05 used its data, while SFZ08 stated that it used F06 data. An archive of F06's computer code and partially processed annual and decadal data (GRL06_reproduce) was subsequently made available, with model simulation data verified against partially extant raw data. Analysis showed that the SFZ08 surface and upper-air diagnostic data were essentially identical to that generated by GRL06_reproduce and that the significantly different CSF05 data were misprocessed.

The GRL06_reproduce MIT model data ended in November 2001, with the last 6 years' data being discarded since the F06 diagnostics ended in 1995, but the surface observational data used ran to August 1996. Dr. Forest has confirmed (C. E. Forest 2012, personal communication) that there is thus a 9-month discrepancy between the F06/SFZ08 model simulation and observational surface diagnostic data and also that the forcings used were valid through to 2001. The timing mismatch has little impact on results.

We present results using the F06/SFZ08 5-decade to 1995/96 surface diagnostic data so as to provide accurate comparisons with the F06 results. We also present results using a revised extended surface diagnostic with correctly matched model and observational data for the 6 decades to 2001, using a 9-decade climatology to 1991 to compute temperature anomalies. Doing so substantially improves the constraint provided by the surface diagnostic. With a surface diagnostic comprising 5 decades, all included in the climatology, greater model simulation temperature increases at higher S_{eq} settings are more heavily diluted by deduction of increased model simulation climatological means. This effect may be illustrated by calculating, for both surface diagnostics, what proportion of the difference between the global mean model-simulated temperatures in the key final diagnostic decade at settings of $S_{\text{eq}} = 1$ and $S_{\text{eq}} = 6$, before deduction of the climatological mean, remains after its deduction. Settings of $\sqrt{K_v} = 1$ and $F_{\text{aer}} = -0.5$ are used. For the revised diagnostic, that proportion is over 1.3 times as great as it is for the F06 diagnostic. Over the lower half of the F_{aer} range, the F06 surface diagnostic is unable to provide any significant constraint on S_{eq} when $\sqrt{K_v} > 1$, while the deep-ocean diagnostic cannot do so until $\sqrt{K_v}$ is higher—greatly higher when $F_{\text{aer}} < -0.75$ (see Fig. S2 of the supplemental material). Use of the revised surface diagnostic remedies this weakness in constraining S_{eq}.

For surface observational data to 2001, to which the original Met Office Hadley Centre Climate Research Unit temperature (HadCRUT) dataset does not extend, we use the latest version, HadCRUT4 (Morice et al. 2012). Otherwise, we use GRL06_reproduce data (for the F06 surface and upper-air diagnostics, the essentially identical SFZ08 data), save for using third climate configuration of the Met Office Unified Model (HadCM3; Gordon et al. 2000) control run deep-ocean data.

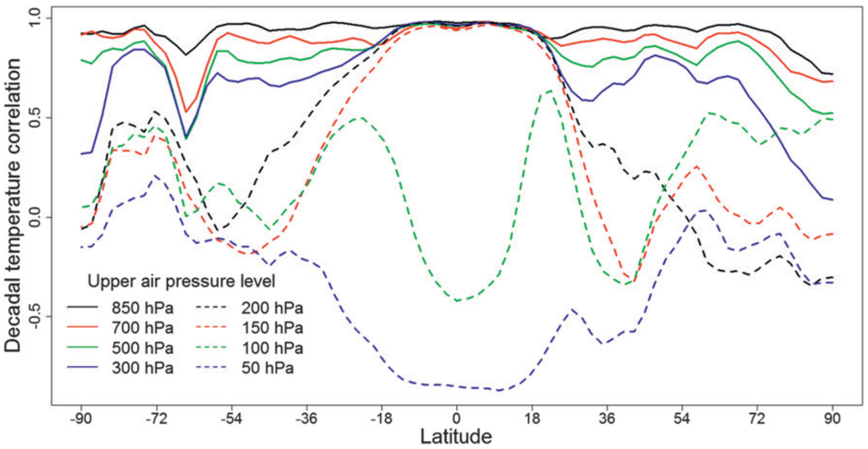

FIG. 1. Correlation of upper-air temperatures with surface-air temperatures based on decadal mean HadCM2 control run data by pressure level. Only data from years 1 to 300 and 781 to 900 have been used, being those periods for which uncorrupted HadCM2 control run upper-air data were obtainable. Data are averaged over all longitudes; latitude resolution is 2.5°.

When employing the surface diagnostic extending to 2001, we use the full Levitus et al. (2005) deep-ocean observational dataset, matching the model simulation data for 40 pentads ending 1998 rather than 37 ending in 1995. The periods covered by the various diagnostics did not match in F06 and need not do so, but it is preferable that they be broadly similar.

b. *Issues with the upper-air diagnostic*

The Bayesian inference in both our and F06's method involves multiplying probability densities relating to the three diagnostics' whitened differences. Doing so will only be valid if errors in those differences are independent. If the errors are positively correlated, multiplying the diagnostic likelihood functions will overstate statistical significance. Covariance of the temperature change variables within the individual diagnostics is addressed by regularized whitening, resulting in a reduced number of variables, which should be independent provided AOGCM control run simulations represent natural variability sufficiently accurately. This issue is discussed in AT99, where a consistency criterion is proposed based (in the F06 case and our case) on r^2_{min}, using truncation parameter κ, lying within the 5%–95% points of the $(\kappa - 3)F_{\kappa-3,\nu}$ or, more cautiously, $X^2_{\kappa-3}$ distribution.

However, whitening does not address interdiagnostic correlations. Nonindependence of observational variability is not a particularly serious concern regarding the deep-ocean diagnostic, where natural variability is only moderately correlated with surface temperature variability and most variance comes from measurement/analysis error. However, it is a concern between the surface and upper-air diagnostics, where no dilution of correlations by added measurement/analysis error occurs. Because of linkage via the lapse rate, fluctuations in surface and tropospheric temperatures are likely to be highly correlated. In the tropics, the tropopause is generally above the 150-hPa level. Figure 1 shows that the decadal-scale correlation of natural variability of upper air with surface temperatures, as simulated by the second climate configuration of the Met Office Unified Model (HadCM2; Johns et al. 1997) control run, is close to one from 20°S–20°N except at 100- and 50-hPa pressure levels. Outside the tropics, these correlations remain generally high for the 850–300-hPa levels.

Even ignoring the correlation issue, inferences from the upper-air diagnostic are problematic. Parameter inference from the upper-air diagnostic varies greatly as κ_{ua} and/or the weighting of levels is changed and is also somewhat sensitive to the smoothness of interpolation. Mass weighting of upper-air data was employed in F06, treating each pressure level as extending halfway toward the next, and toward 1000 hPa (surface pressure) at the bottom and 30 hPa at the top. If alternatively the top (50 hPa) level is treated as extending halfway toward zero pressure—increasing its weight modestly, from 4.0% to 5.6%—parameter inference in the S_{eq}–K_{υ} plane from the fit between the model and observational upper-air data changes dramatically when $\kappa_{\mathrm{ua}} = 14$, as in F06. The F06 weighting is perhaps preferable, but sensitivity to minor reweighting is worrying. The possibility of inference being sensitive to fairly arbitrary choices of weightings and truncation parameters is inherent in optimal fingerprint methods.

Figure 2a replicates the AT99 statistical consistency test on the sum of squares of whitened differences

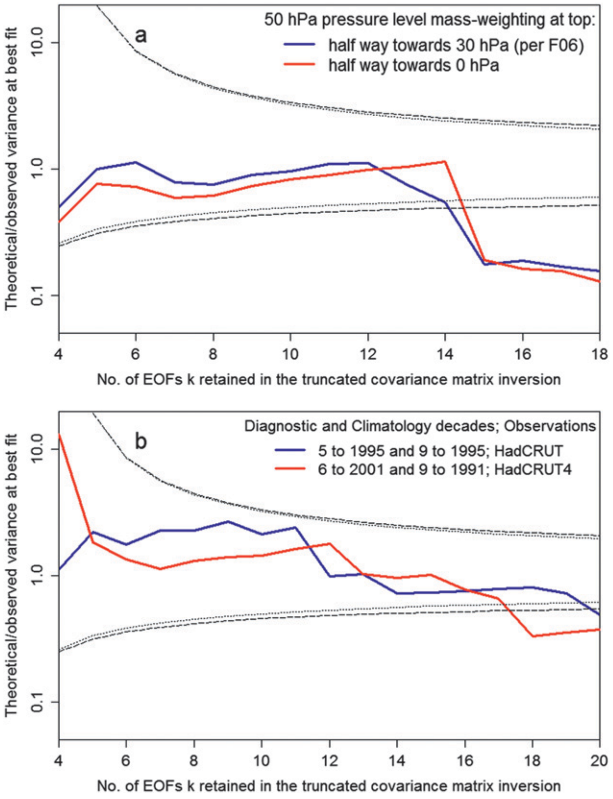

FIG. 2. Consistency of (a) upper-air diagnostic and (b) surface diagnostic with the statistical model: theoretical sum of squared whitened model − observation differences at best-fit parameter settings relative to actual sum, by diagnostic data and truncation parameter (κ_{ua} or κ_{sfc}). Consistent regions lie between black lines. In (a) the dashed and dotted black lines represent respectively 5% and 95% points of $F_{\kappa_{\mathrm{ua}}-3,39}$ and (more stringently) $X^2_{\kappa_{\mathrm{ua}}-3}$ distributions. In (b) those lines represent respectively 5% and 95% points of $F_{\kappa_{\mathrm{sfc}}-3,49}$ and (more stringently) $X^2_{\kappa_{\mathrm{sfc}}-3}$ distributions.

between best-fit modeled and observed upper-air diagnostic temperatures. At F06's selection of $\kappa_{\mathrm{ua}} = 14$, the minimum r^2 with the top level weighted toward 0 hPa is 10.2, easily satisfying the AT99 consistency test; whereas with the F06 weightings it is 20.2, failing the stricter $X^2_{\kappa-3}$ test version. This high r^2_{min} value is not apparent from the interpolated r^2 values produced by the GRL06_reproduce code, which do not accurately reflect the F06 method and have a much lower and incorrect minimum of 11.4. On both weightings, at $\kappa_{\mathrm{ua}} = 12$ the AT99 consistency test is well satisfied, and parameter inference is similar to that at $\kappa_{\mathrm{ua}} = 14$ with the alternative mass weighting. This (along with it failing the AT99 $X^2_{\kappa-3}$ test) strongly suggests that it is the F06 weighting–truncation combination that produces invalid inference. We validated the weighting–truncation combinations by recomputing r^2_{min} values using leave-one-out subsampling, omitting each row of the control data matrix in turn. At $\kappa_{\mathrm{ua}} = 14$, over half these subsampled r^2_{min} values failed the AT99 $X^2_{\kappa-3}$ consistency test when using F06 50-hPa-level weighting, and over a quarter failed when using the alternative weighting. At $\kappa_{\mathrm{ua}} = 12$, all of the r^2_{min} values passed it, irrespective of weighting. We therefore prefer $\kappa_{\mathrm{ua}} = 12$, but retain F06's 50-hPa-level weighting.

Given high correlations and sensitivity to weightings, the validity of inference based on inclusion of information from the upper-air diagnostic is questionable. Fortunately, the upper-air diagnostic is considerably less informative than the surface diagnostic in constraining parameter values. The results using the extended surface diagnostic avoid the problematic upper-air diagnostic by employing only surface and deep-ocean diagnostics; adding the upper-air diagnostic, using the preferred $\kappa_{\mathrm{ua}} = 12$, hardly changes parameter inference.

4. Method

a. Introduction

We retain the F06 approach of whitening diagnostic variables, using the same regularized inverse climate noise covariance matrix estimates. As discussed in section 3, whitening does not remove interdiagnostic noise correlations, which appear substantial as regards the upper-air diagnostic but tolerably low between the surface and deep-ocean diagnostics. We remove dependence on parameter surface flatness by working with the full set of whitened variables $\tilde{\mathbf{w}}_c$, a vector of length $\kappa_{\mathrm{sfc}} + \kappa_{\mathrm{ua}} + \kappa_{\mathrm{do}}$ (where $\kappa_{\mathrm{do}} = 1$), rather than using the F06 Δr^2 method—the mathematical basis of which requires the whitened variables to be linear functions of the parameters, which they are not. Accordingly, the F distribution we use relates to r^2, not Δr^2. We adjust the F distribution for the geometry involved; the unadjusted density goes to zero at the best-fit point (see supplemental material for additional discussion). This step appears to have been omitted in F06: its effects increase with dimensionality and are modest when using the Δr^2 method. We undertake only a single Bayesian step, in whitened variables space, at which point the likelihood function is a radially symmetric multidimensional Gaussian and a joint uniform prior—corresponding to uniform sampling over the whitened variables $\tilde{\mathbf{w}}$—is the so-called Jeffreys' prior (Jeffreys 1946) and is known to be noninformative. We then convert the thus-derived joint PDF for the "true" whitened diagnostic variables (hypothetical observations excluding climate noise) $\mathbf{w}$ into a joint PDF for the true model parameters $\boldsymbol{\theta}_t$, which, assuming model accuracy, equal the true climate system parameters.

Although working with full sets of whitened differences, rather than just their sum of squares, is much more computationally demanding than the F06 method, given the large number of parameter combinations involved, we thereby retain the information required to derive a PDF conversion factor [$\pi(\boldsymbol{\theta}_t)$] equating to a noninformative joint parameter prior. F06's choice of uniform parameter priors was made in the absence of such information. We avoid use of Bayesian updating, with its known incompatibility with noninformative priors, by working with a single set $\tilde{\mathbf{w}}_c$ of combined whitened differences from all the diagnostics.

b. Derivation and interpretation of the PDF conversion factor

Mathematical details of F06's $mF_{m,\nu}$ method and of the objective Bayesian method, including derivation of the conversion factor $\pi(\boldsymbol{\theta}_t) = |\mathbf{D}^{\mathrm{T}}\mathbf{D}|^{1/2}$ from probability density in whitened observation space to that in parameter space, are given in appendixes A and B respectively. (The term $\mathbf{D}$ is the Jacobian matrix of partial derivatives of the whitened modeled diagnostic temperatures with respect to the model parameters.) The resulting inferential equation is

$$p(\boldsymbol{\theta}_t \,|\, \tilde{\mathbf{w}}_c) \propto p(\tilde{\mathbf{w}}_c \,|\, \boldsymbol{\theta}_t)\pi(\boldsymbol{\theta}_t). \qquad (1)$$

This implies that $\pi(\boldsymbol{\theta}_t)$ can alternatively be viewed as the joint parameter prior when applying Bayes's theorem to derive a posterior joint parameter PDF from the combined diagnostics likelihood, being $p(\tilde{\mathbf{w}}_c \,|\, \boldsymbol{\theta}_t)$ as a function of $\boldsymbol{\theta}_t$. This will be a noninformative Jeffreys' prior, since the prior used when applying Bayes's theorem to the whitened temperatures was a noninformative Jeffreys' prior, and Jeffreys' priors are invariant under reparameterization.

The standard Jeffreys' noninformative prior for parameters related to normally distributed whitened variables, derived in Jewson et al. (2009), is indeed identical to $\pi(\boldsymbol{\theta}_t)$. An alternative derivation of the same posterior using a result in a differential geometry approach to conditional probability (Mosegaard and Tarantola 2002) is outlined in the supplemental material. A joint posterior can also be derived using our method on the assumption, implicit in F06, that the parameter surface is flat. On that assumption, the combined likelihood for each diagnostic agrees to that using the F06 $mF_{m,\nu}(\Delta r^2)$ method, where $m = 3$, after the geometrical volume correction, which converts a PDF for Δr^2 into a joint PDF for the three underlying error variables (see supplemental material for additional discussion).

The detailed analysis has been undertaken with uncertainty in the variance of the whitened differences ignored. On that basis, noting from (B6) that the right-hand side of (B8) is proportional to $p(\tilde{\mathbf{w}}_c|\boldsymbol{\theta}_t)$,

$$\begin{aligned} p(\tilde{\mathbf{w}}_c \,|\, \boldsymbol{\theta}_t) &\propto \exp[-(\tilde{\mathbf{s}}_c(\boldsymbol{\theta}_m = \boldsymbol{\theta}_t))^{\mathrm{T}}(\tilde{\mathbf{s}}_c(\boldsymbol{\theta}_m = \boldsymbol{\theta}_t))/2] \\ &\propto [\exp(-r_{\mathrm{sfc}}^2/2)\exp(-r_{\mathrm{ua}}^2/2)\exp(-r_{\mathrm{do}}^2/2)]|_{\boldsymbol{\theta}_m=\boldsymbol{\theta}_t}. \end{aligned} \qquad (2)$$

This represents the product of likelihood functions for the individual diagnostics, its form for each of them equating, after multiplying by a $(r^2)^{\kappa/2-1}$ geometric volume adjustment (see supplemental material for additional discussion), to a X_κ^2 distribution for the sum of squares r^2 of the κ whitened differences. To allow for variance uncertainty, we replace X_κ^2 distributions with $\kappa F_{\kappa,\nu}$ distributions (a t distribution for the deep-ocean diagnostic), as in the F06 method, giving the combined likelihood as

$$p(\tilde{\mathbf{w}}_c|\boldsymbol{\theta}_t) \propto [F_{\kappa_{\mathrm{sfc}},\nu_{\mathrm{sfc}}}(r_{\mathrm{sfc}}^2/\kappa_{\mathrm{sfc}})/r_{\mathrm{sfc}}^{\kappa_{\mathrm{sfc}}-2} F_{\kappa_{\mathrm{ua}},\nu_{\mathrm{ua}}}(r_{\mathrm{ua}}^2/\kappa_{\mathrm{ua}})/r_{\mathrm{ua}}^{\kappa_{\mathrm{sfc}}-2} t_{\nu_{\mathrm{do}}}(r_{\mathrm{do}})]|_{\boldsymbol{\theta}_m=\boldsymbol{\theta}_t}. \qquad (3)$$

To simplify the mathematics, the derivation of the PDF conversion factor from whitened variables to parameter space—which allocates probability according to relative volumes in those spaces—ignores uncertainty in the variance of the whitened differences arising from $\mathbf{C}_N$ being estimated with only ν DF. Such variance uncertainty has little effect on the calculation involved, and the conversion factor has much less influence than the joint likelihood (where variance uncertainty is allowed for). Variance uncertainty is small for the surface diagnostic since ν_{sfc} is large. We take $\nu_{\mathrm{sfc}} = 49$ and 40 for respectively the 5- and 6-decade surface diagnostics, reflecting the numbers of samples with nonoverlapping diagnostic periods obtainable from the control data, increased by 50% for overlapping samples (see AT99). Variance uncertainty for the upper-air diagnostic has little influence on the conversion factor. Uncertainty in the deep-ocean trend variance estimate, which is large and estimated with fewer DF, is explicitly allowed for. A more detailed justification for the approach adopted, including for the deep-ocean diagnostic, is set out in the supplemental material.

5. Implementation

a. Interpolation

F06 computed r^2 for each of the modeled parameter combinations and interpolated it to a fine parameter grid spanning their range in two stages. Interpolation was first to a spacing of $\Delta S_{\mathrm{eq}} = 0.1\,\mathrm{K}$ and $\Delta\sqrt{K_v} = 0.1\,\mathrm{cm\,s^{-1/2}}$ and then to $\Delta F_{\mathrm{aer}} = 0.05\,\mathrm{W\,m^{-2}}$. Our method requires separate interpolation of all model diagnostic variables. Interpolation of model data is preferable as it is less nonlinear in the parameters than is r^2, and model simulation variability can be allowed for.

We follow Curry (2007) in using a thin plate spline (TPS) single-stage interpolation. For convenience we extrapolate S_{eq} into the range 0–0.5; as likelihood is extremely low there, this has negligible impact. The interpolation fits a TPS, separately for each variable, so as to best match actual values at all model run parameter combinations. The total squared misfits are minimized subject to a smoothness constraint, which by giving influence to all actual values in a neighborhood restricts the impact of distortions in individual values caused by model simulation variability. The F06 method is less able to reduce the effects of model simulation variability both because it uses two steps and because the relationship between the quantity being interpolated and model simulation variability is indirect.

Since we compute derivatives of the whitened interpolated variables with respect to the parameters—and from that (Jacobian) matrix the PDF conversion factor/noninformative prior—by differencing across adjacent fine grid cells, reasonably smooth interpolation is preferable. We achieve this by restricting the DF used in interpolation to below that resulting from the default smoothness constraint. With 256 DF for all diagnostics, the noninformative prior remains somewhat artifacted, particularly at model run locations. Using 128 DF instead for all diagnostics produced less artifaction, while still departing only modestly from model ensemble values.

However, the interpolation should depart from model ensemble values on account of model simulation variability (Curry 2007). Using 128 DF, interpolation error at model run locations averages 60%–100% of estimated model ensemble variability for the surface diagnostic, confirming that choice is reasonable. Similar comparisons suggest that 64 DF would be appropriate for the upper-air diagnostic interpolation and 256 DF for the deep-ocean diagnostic. These values are used for all results. The parameter marginal posterior PDFs using our method are very similar with 64, 128, or 256 DF interpolation.

When using the F06 method, parameter PDFs exhibit more sensitivity to the DF used in interpolation, principally for the upper-air diagnostic. TPS upper-air interpolation with unrestricted DF results in a flattening—starting at S_{eq} approaching 4—appearing in the S_{eq} PDF. Investigation reveals that the flattening is only noticeable when unrestricted DF interpolation is used for the 0°–5°S latitude band, where it appears that some data misprocessing may have occurred. The even more limited reduction of model simulation variability with the F06 interpolation method may account for such flattening being more pronounced in the published F06 S_{eq} PDF.

b. Whitening the diagnostic variables

We follow F06's use of surface and upper-air data from the 1691-yr HadCM2 control run, masking for observational availability. F06 found that PDFs resulting from use of surface control run data from different AOGCMs did not differ qualitatively. Before generating the new 6-decade surface diagnostic, we verified that, using the F06 diagnostic periods, our code accurately generated the F06/SFZ08 surface diagnostic from model, observational, and control data extracted from the GRL06_reproduce archive.

Figure 2b replicates the AT99 consistency test on the sum of squares of whitened differences between best-fit modeled and observed surface diagnostic temperatures. Using the GRL06_reproduce HadCM2 control data, the 5- and 6-decade diagnostic data fail the AT99 consistency test beyond $\kappa_{\mathrm{sfc}} = 19$ and 17, respectively. Using the 5-decade to 1995/96 diagnostic, inferred parameter PDFs are almost identical between $\kappa_{\mathrm{sfc}} = 16$ and $\kappa_{\mathrm{sfc}} = 19$ and differ only slightly at $\kappa_{\mathrm{sfc}} = 15$. Using the revised 6-decade to 2001 diagnostic, parameter PDFs differ little between $\kappa_{\mathrm{sfc}} = 15, 16$, or 17. We therefore emphasize use of $\kappa_{\mathrm{sfc}} = 16$, as in F06, for both the original F06 and the revised surface diagnostics. Stability was tested by recomputing r^2_{min} values with varying segments of the control data omitted. For both diagnostic versions the AT99 test was satisfied in all cases at $\kappa_{\mathrm{sfc}} = 16$.

In view of the significant difference in inference as to S_{eq} arising from use of the revised surface diagnostic and the danger that this might arise from a particular EOF correlating with some noise pattern in the observations, we have examined the effects of excluding one EOF at a time. Since the AT99 consistency test is only satisfied up to $\kappa_{\mathrm{sfc}} = 17$ for the revised surface diagnostic, we took the first 17 EOFs, calculated r^2 values with the contribution thereto from each EOF in turn removed, and derived parameter PDFs for each case, using the same 40-yr revised deep-ocean diagnostic throughout. No major differences emerged, with the mode of all the PDFs for S_{eq} being within ±0.2 K of that at $\kappa_{\mathrm{sfc}} = 16$ with no EOFs excluded. Similar variation occurred using the original diagnostic.

As shown in Fig. 3, using the revised diagnostics, parameter PDFs become less well constrained at $\kappa_{\mathrm{sfc}} = 14$, but not materially so when using the objective Bayesian method. Below $\kappa_{\mathrm{sfc}} = 14$, parameter PDFs become increasingly poorly constrained using either the original or

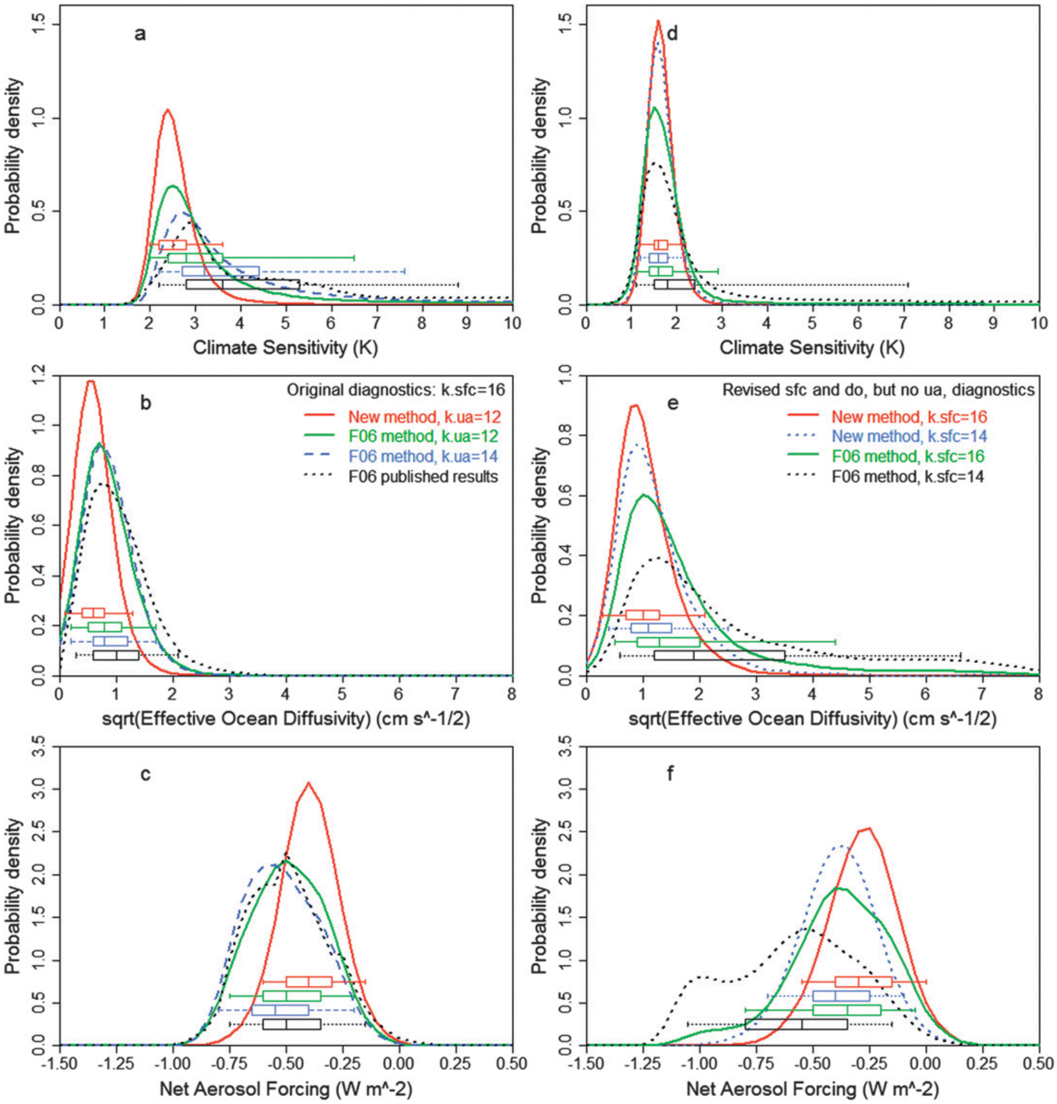

FIG. 3. Marginal posterior PDFs for the three climate system parameters by diagnostics employed, method used, and either (a),(b),(c) upper-air diagnostic EOF truncation parameter with the F06 diagnostics or (d),(e),(f) surface diagnostic EOF truncation parameter with revised surface and deep-ocean diagnostics and no upper-air diagnostic. Panels show marginal posterior PDFs for, from the top, S_{eq}, K_{υ}, and F_{aer}. In (a),(b), and (c), the solid red lines show marginal PDFs using the new objective Bayesian method and $\kappa_{\mathrm{ua}} = 12$, while marginal PDFs using the F06 method are shown at $\kappa_{\mathrm{ua}} = 12$ (solid green lines) and at $\kappa_{\mathrm{ua}} = 14$, as used in F06 (dashed blue lines). The dotted black lines show the published F06 PDFs. In (d),(e), and (f), marginal PDFs using the new objective Bayesian method are shown with $\kappa_{\mathrm{sfc}} = 16$ (solid red lines) and $\kappa_{\mathrm{sfc}} = 14$ (dotted blue lines), while corresponding marginal PDFs using the F06 method are shown with solid green and dotted black lines, respectively. The box plots indicate boundaries, to the nearest fine grid value, for the percentiles 5–95 (vertical bar at ends), 25–75 (box ends), and 50 (vertical bar in box). Parameter ranges are given by plot boundaries.

revised surface and deep-ocean diagnostics, as insufficient information is then retained to discriminate adequately between differing regions in parameter space.

We use the F06/SFZ08 HadCM2 upper-air control data matrix, containing 40 nonoverlapping samples, and take $\nu_{\mathrm{ua}} = 39$, one DF being lost in estimating a mean.

Whitening the deep-ocean diagnostic differences involves no truncation but necessitates aggregating climate noise and estimated uncertainty in the observational temperature trend. Weighted linear least squares regression is used, as in F06, giving observational trend estimates for the 37 years to 1995 (40 years to 1998) of 0.70 (0.68) mK yr^{-1}, with unadjusted standard error (SE) of 0.08 (0.07) mK yr^{-1}. However, the regression residuals are highly autocorrelated, largely because of the 80% overlap of adjacent pentads. Adjusted, approximately twice as high, SEs estimated from regressions on nonoverlapping pentadal data were therefore used. This ignores the previously discussed gain in effective DF resulting from using overlapping samples. Most remaining autocorrelation is probably because of slow natural variability in ocean temperature, which is accounted for by adding climate noise rather than measurement error correlation. To test sensitivity to observational trend SEs, variants of the main results using 50% higher SEs were generated. These variants also serve to test the impact of climate noise correlation between the surface and deep-ocean diagnostics since the higher SEs reduce (to below 15%) the noise contribution to total deep-ocean whitening variance. In each case, the effective DF $\nu_{\rm do}$ used for the t distribution was adjusted to reflect the adjustments made to the SEs.

We obtained deep-ocean control data from the 6100-yr HadCM3 control run, whereas F06 used years 1–900 of the Geophysical Fluid Dynamics Laboratory (GFDL) R30 spectral resolution control run (Delworth et al. 2002). Trend variability in the first one-third of these control runs greatly exceeds that in the middle and final thirds, with the models appearing to be adjusting nonmonotonically toward dynamic equilibrium during the first one-third, perhaps because of dynamically inconsistent initial conditions. We therefore use the final two-thirds of the HadCM3 data; the corresponding segments of GFDL R30 data yield a similar estimate. Our deep-ocean climate noise estimate is accordingly only about half F06's. Conversely, our observational trend SE is about double F06's since F06 made no autocorrelation adjustment. The aggregate climate noise and observational estimate SE deep-ocean trend variance used, without a further 50% increase in SE, is close to that in F06.

c. *Off-grid probability mass*

We adopt the F06 approach of assigning zero probability to off-grid regions and normalizing to unit total probability. The joint parameter posterior PDFs are low at all grid boundaries except (when using the objective Bayesian method with the original F06 diagnostics) to a modest extent at the $\sqrt{K_v} = 0$ boundary—which the true $\sqrt{K_v}$ value must lie above—for below-average $S_{\rm eq}$ values. If substantial parts of the likelihood lay in infeasible parameter regions, parameter inference would be problematic.

6. Results

The left-hand panels in Fig. 3 show marginal posterior PDFs for $S_{\rm eq}$, K_v, and $F_{\rm aer}$ obtained with the original diagnostics, at the preferred $\kappa_{\rm sfc} = 16$ and $\kappa_{\rm ua} = 12$, using both the F06 method with uniform priors and our new objective Bayesian method. For the F06 method, PDFs are also shown using $\kappa_{\rm ua} = 14$. PDFs using the new method are insensitive to the choice of $\kappa_{\rm ua}$. Identical diagnostic whitened differences are used for both methods. F06 employed different DF for the three diagnostics, but on both methods results are insensitive to adopting the DF used in F06.

Using the F06 method, PDFs for $S_{\rm eq}$ are much worse constrained than when using our new method, particularly at the deprecated $\kappa_{\rm ua} = 14$. The $S_{\rm eq}$ PDF using our method has a shape approximating that expected theoretically (Roe and Baker 2007): when converted into a PDF for the climate feedback parameter, which is reciprocally related to $S_{\rm eq}$, its distribution is close to normal. The $S_{\rm eq}$ PDFs using the F06 method, when so converted, are much less symmetric.

The PDFs for $\sqrt{K_v}$ are also better constrained using our method and the already well constrained PDFs for $F_{\rm aer}$ become more so.

The F06 main result PDFs based on uniform priors are shown for comparison. They differ slightly from the PDFs we compute using the F06 method and $\kappa_{\rm ua} = 14$, partly because of differences in interpolation and partly because of various errors in F06's implementation of its method. In addition to its upper-air r^2 values not according with its method, as already mentioned, the GRL_reproduce code shows that in computing likelihoods the F distribution's cumulative distribution function (CDF), was erroneously used, rather than its PDF, and the univariate deep-ocean diagnostic was treated as trivariate, its r^2 value being wrongly divided by 3.

At $\kappa_{\rm ua} = 12$, using our method the 5%–95% bounds for $S_{\rm eq}$, $\sqrt{K_v}$, and $F_{\rm aer}$ are respectively 2.0–3.6 K, 0.1–1.3 cm s$^{-1/2}$, and −0.6 to −0.15 W m^{-2}. Using the F06 method (in parentheses: as reported in F06) the corresponding ranges are 2.0–6.5 (2.1–8.9) K, 0.2–1.7 (0.2–2.0) cm s$^{-1/2}$, and −0.75 to −0.2 (−0.74 to −0.14) W m^{-2}. The modes using our and the F06 method are respectively 2.4 and 2.5 (2.9) K for $S_{\rm eq}$, 0.6 and 0.7 (0.8) cm s$^{-1/2}$ for $\sqrt{K_v}$, and −0.4 and −0.5 (−0.5) W m^{-2} for $F_{\rm aer}$.

Imposing the assumption implicit in the F06 Δr^2 method that the parameter surface is flat, but otherwise using our new method, reduces the upper 95% bound on $S_{\rm eq}$ but only by 0.3 K even in the 50% higher than standard deep-ocean observational trend SE case. This confirms that assuming flatness is inappropriate, although the effects are modest here.

The right-hand panels in Fig. 3 show PDFs corresponding to those in the left-hand panels but using the revised diagnostics: longer 6-decade to 2001 surface diagnostic, 40-yr to 1998 deep-ocean diagnostic, and no upper-air diagnostic. The shapes of the $\kappa_{\rm sfc} = 16$ PDFs are broadly similar to those using the original F06 diagnostics, but those for S_{eq} are narrower and have lower modes, while those for $\sqrt{K_v}$ and F_{aer} are wider, particularly when using the F06 method. Varying $\sqrt{K_v}$ or F_{aer} has only a small effect on model-simulated surface temperatures and hence on the diagnostic fit when model climate sensitivity is low. Therefore, looser constraint on $\sqrt{K_v}$ and F_{aer} is a counterpart of S_{eq} being tightly constrained at lower levels.

As when using the F06 diagnostics, we emphasize results using $\kappa_{\mathrm{sfc}} = 16$ where consistency with the statistical model, per the test in AT99, is good. We show PDFs at $\kappa_{\mathrm{sfc}} = 14$ for comparison. Those using our new method differ relatively little from PDFs at $\kappa_{\mathrm{sfc}} = 16$, save for the F_{aer} PDF shifting slightly. However, PDFs using the F06 method become substantially less well constrained.

At $\kappa_{\mathrm{sfc}} = 16$, using our method the 5%–95% bounds for S_{eq}, $\sqrt{K_v}$, and F_{aer} are respectively 1.2–2.2 K, 0.3–2.1 cm s$^{-1/2}$, and −0.55 to 0.0 W m^{-2}. Using the F06 method the corresponding ranges are 1.1–2.9 K, 0.5–4.4 cm s$^{-1/2}$, and −0.8 to −0.05 W m^{-2}. The modes (medians) using our and the F06 method are, respectively, 1.6 and 1.5 K (both 1.6 K) for S_{eq}, 0.9 and 1.0 (1.0 and 1.3) cm s$^{-1/2}$ for $\sqrt{K_v}$, and −0.25 and −0.4 (−0.3 and −0.35) W m^{-2} for F_{aer}.

Imposing a 50% increase in the estimated deep-ocean observational trend SE (at $\kappa_{\mathrm{sfc}} = 16$) marginally increases the widths of the parameter marginal PDFs when using our method, with the 95% bound on S_{eq} rising to 2.3 K. Using the F06 method they increase somewhat more: most notably, the 95% bound for S_{eq} becomes 4.4 K.

Figure 4 shows the computed PDF conversion factor from whitened difference to parameter space (or noninformative joint parameter prior) used to generate the new method revised diagnostics results, $\kappa_{\mathrm{sfc}} = 16$. Its shape varies little using $\kappa_{\mathrm{sfc}} = 14$ and/or the F06 diagnostics. A PDF-weighted mean over F_{aer} values has been shown. When instead conditioned on F_{aer}, the prior retains its broad shape over the range where the F_{aer} likelihood is significant but scales up by a factor of several times as F_{aer} becomes less negative. The sharp decline in the prior with $\sqrt{K_v}$ reflects reducing sensitivity of modeled temperatures to parameter changes as ocean heat uptake increases, so that volumes in parameter space correspond to progressively smaller volumes in whitened difference space.

Model-prediction variability remaining after interpolation accounts for departures from smoothness and monotonicity in the noninformative prior. The upturn in the low S_{eq}, high $\sqrt{K_v}$ corner is an artifact arising because at low S_{eq}, where temperature changes are small, model variability results in some of the simulated surface diagnostic temperatures changing in the wrong direction from $\sqrt{K_v} = 5$ to 8, between which there are no modeled values.

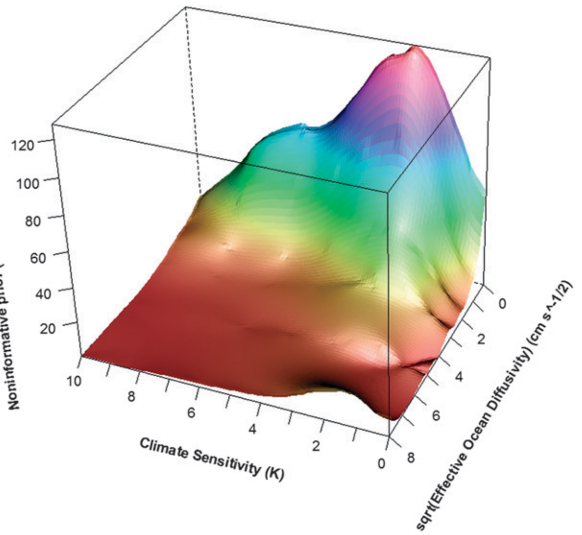

FIG. 4. Scaling factor from whitened difference space to parameter space employed by the objective Bayesian method, equivalent to a noninformative joint prior for the three climate system parameters. The plot shows variation in S_{eq}–$\sqrt{K_v}$ space, integrating over F_{aer} weighted by its marginal PDF with $\kappa_{\mathrm{sfc}} = 16$, using the revised diagnostics (longer surface and deep-ocean diagnostics, no upper-air diagnostic).

Figure 5 shows marginal joint credible regions in S_{eq}–$\sqrt{K_v}$ space, using our method for both the F06 and revised diagnostics. The tongue of probability heading toward high S_{eq} levels, using the F06 diagnostics, is only reduced to low levels (producing better-constrained PDFs, particularly for S_{eq}) by the falling value of the noninformative prior. The other diagnostic likelihoods cannot by themselves sufficiently reduce the substantial surface diagnostic likelihood that exists at very high S_{eq}, even at fairly low $\sqrt{K_v}$. The revised surface diagnostic discriminates more, so the S_{eq} PDF is reasonably constrained even when employing the F06 uniform priors method.

7. Discussion

The Forest papers develop a powerful means of estimating climate sensitivity jointly with uncertain ocean diffusivity and aerosol forcing. We develop a revised, objective Bayesian, statistical inference approach that improves their methods, principally by use of a noninformative prior but also by the avoidance of Bayesian updating (which is incompatible therewith) and of dependence on parameter surface flatness, and by incorporating a geometric volume adjustment.

Using our objective Bayesian method, the F06 approach of comparing observed with model-simulated spatiotemporal surface temperature patterns fairly well

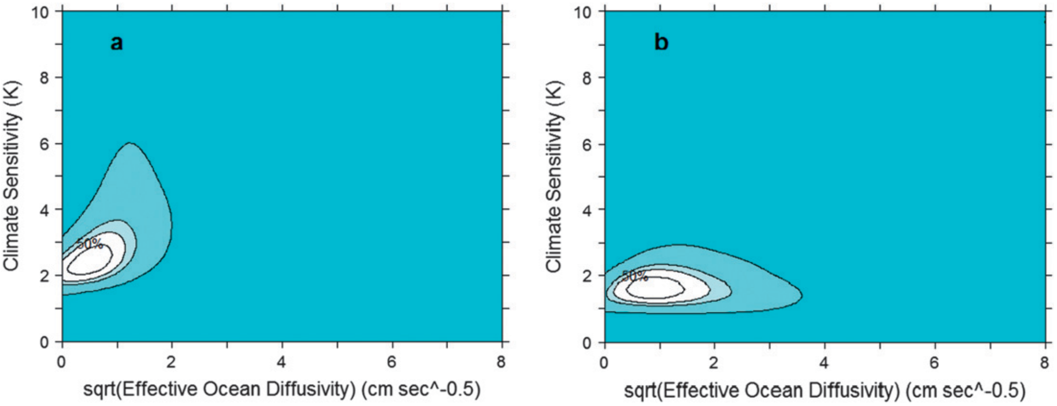

FIG. 5. Joint credible regions in S_{eq}–$\sqrt{K_v}$ space using the objective Bayesian method at $\kappa_{\mathrm{sfc}} = 16$ and all relevant diagnostics (with $\kappa_{\mathrm{ua}} = 12$ when using the upper-air diagnostic). (a) Original F06 diagnostics and (b) revised diagnostics (longer surface and deep-ocean diagnostics, no upper-air diagnostic). The contours enclose regions of highest joint marginal posterior density with various levels of total probability, the innermost contour being at 50%. Shading shows rejection regions at significance levels of 20% (lightest), 10%, and 1% (darkest).

constrains parameter estimates using the same diagnostics as in F06, whereas S_{eq} is badly constrained using the F06 method and uniform priors. We find the upper-air diagnostic—which provides relatively weak inference—problematic, with natural variability in many of its variables likely highly correlated with that in surface diagnostic variables and sensitivity to the weightings and truncation used, with the weightings–truncation combination used in F06 seemingly producing unsatisfactory inference.

We resolve these issues by employing only surface and deep-ocean diagnostics, revising these to use longer diagnostic periods, taking advantage of previously unused post-1995 model simulation data and correctly matching model simulation and observational data periods (mismatched by 9 months in the F06 surface diagnostic). Using the revised diagnostics, estimates of S_{eq} are lower and more tightly constrained, with a 1.1–2.9-K range obtained using the F06 method and 1.2–2.2 K using the new method. Switching from the original HadCRUT observational surface temperature dataset to the updated HadCRUT4 may have contributed significantly to this reduction: Ring et al. (2012) reported that doing so caused a 0.5-K reduction in their S_{eq} estimate.

Comparison, at the best-fit parameter combinations, of the model-simulated and observed rises in global mean temperature between the first 20 and last 20 simulation years provides a key reality check for the validity of inference arising from the alternative diagnostics, testing all stages of the optimal fingerprint method. Taking the best-fit point at $\kappa_{\mathrm{sfc}} = 16$ under the new method and using the revised diagnostics, the two temperature rises are virtually identical. Using the original F06 diagnostics, the model-simulated rise is one-third higher than observed. These test results, while not proving that parameter inference is correct using the revised diagnostics or incorrect using the original diagnostics, provide substantial support for preferring use of the revised diagnostics. However, the sensitivity of inference about S_{eq} to the data used, its processing and analysis, indicates that results using the revised diagnostics should not be regarded as definitive.

Our method of converting directly from a PDF in whitened variable space to one in parameter space yields a conversion factor equating to a noninformative joint prior for the parameters. Its shape is far removed from the uniform priors mainly used in F06, substantially affecting results. Although, coincidentally, the central section of the shape of the F06 expert prior for S_{eq} is broadly similar to a cross section of the noninformative joint prior at fixed K_v and F_{aer}, the expert prior declines much more rapidly at low and high S_{eq}. Moreover, the noninformative joint prior is far from uniform in K_v and F_{aer} and is not a separable function of the three parameters. The shape of the noninformative prior reflects how informative the data are about the parameters as their values vary, rather than the preexisting information as to parameter values. Any such prior information is unlikely to be independent of information provided by the data, invalidating Bayesian inference. We recommend that a computed noninformative joint parameter prior, not separate uniform (or expert) priors, be used in future Bayesian climate parameter studies.

Using the same diagnostics and method as F06, we obtain tighter bounds (5%–95% points per marginal posterior PDFs) at $\kappa_{\mathrm{sfc}} = 16$ than reported with uniform priors in F06. This is particularly so for S_{eq}, the central estimate of which is, moreover, reduced by 0.4 K. This principally reflects use of a preferable upper-air truncation parameter and a corrected implementation of the F06 method. Our 90% range of 2.0–3.6 K for S_{eq}, obtained using the new method with the F06 diagnostics, improves on the IPCC's 2–4.5-K "likely" range given in Hegerl et al. (2007). Our 90% range of 1.2–2.2 K for S_{eq}, obtained using the preferred revised diagnostics and the new method, appears low in relation to that range, partly because uncertainty in nonaerosol forcings and surface temperature measurements is ignored. Incorporation of such uncertainties is estimated to increase the S_{eq} range to 1.0–3.0 K, with the median unchanged (see supplemental material for derivation and additional discussion).

Our 1.6-K mode for S_{eq} obtained with the objective Bayesian method and the preferred revised diagnostics is identical to that from the main results in two recent studies providing observationally constrained estimates of S_{eq}: Aldrin et al. (2012) and (using the same HadCRUT4 dataset) Ring et al. (2012). Our 1.1–2.9-K 90% range for S_{eq} obtained using the F06 method (uniform priors) and revised diagnostics compares with the 1.2–3.5 K obtained using a uniform prior for S_{eq} in Aldrin et al. (2012).

Acknowledgments. We thank Charles Curry, Chris Forest, and Bruno Sansó for making available respectively CSF05, GRL06_reproduce, and SFZ08 archives, Tim Johns for providing HadCM2/3 AOGCM control run data, and Hu McCullough, Andrew Montford, Steven Mosher, and four reviewers for helpful comments.

APPENDIX A

The F06 Method and $mF_{m,\nu}$ Distribution

Each F06 goodness-of-fit statistic is based on a vector $\tilde{\mathbf{u}} = \tilde{\mathbf{T}}_o - \mathbf{T}_m(\boldsymbol{\theta}_m)$ of the p differences between observations and model predictions. Through premultiplication by a "whitening" matrix $\hat{\mathbf{C}}_N^{-1/2*} = \mathbf{\Lambda}_\kappa^{-1}\mathbf{V}_\kappa^{\mathrm{T}}$ ($\mathbf{V}_\kappa\mathbf{\Lambda}_\kappa^2\mathbf{V}_\kappa^{\mathrm{T}}$ being the truncated eigen-decomposition of $\hat{\mathbf{C}}_N$), $\tilde{\mathbf{u}}$ is transformed into a set of error variables (whitened differences) $\tilde{\mathbf{s}} = \hat{\mathbf{C}}_N^{-1/2*}\tilde{\mathbf{u}}$ that would, if inter alia the estimated noise covariance matrix $\hat{\mathbf{C}}_N$ and the model predictions were accurate, have independent $N(0, 1)$ distributions when the model parameter settings equaled the climate system parameters' hypothetical true values.

Since only the $\kappa < p$ largest covariance matrix EOFs are retained, there are only κ nonzero whitened differences. The whitened differences' sum of squares, r^2, is computed for each parameter combination setting. AT99 states, in the context of climate change detection and attribution, that where m scaling factors for the ratios of observed to model-predicted pattern amplitudes are estimated before model minus observation differences are determined, $\tilde{\mathbf{u}}^{\mathrm{T}}\hat{\mathbf{C}}_N^{-1}\tilde{\mathbf{u}} \sim (\kappa - m)F_{\kappa-m,\nu}$ ($\hat{\mathbf{C}}_N$ from here on referring to the version retaining κ EOFs). When, as in F06, the aim is instead to estimate climate system parameters, the scaling factors are all set to unity rather than being estimated, as discussed in F01; so $m = 0$. Accordingly, assuming the model is accurate and uses the true parameter values (resulting in the underlying whitened differences being zero)

$$\tilde{\mathbf{u}}^{\mathrm{T}}\hat{\mathbf{C}}_N^{-1}\tilde{\mathbf{u}} \sim \kappa F_{\kappa,\nu}. \tag{A1}$$

The realized differences at each candidate model parameter settings are whitened and the whitened sum of squares r^2 computed. That sum represents the squared length of $\tilde{\mathbf{s}}$, a vector in a κ-dimensional whitened difference space, with origin where all those differences are zero:

$$r^2 = \|\tilde{\mathbf{s}}\|^2 = \tilde{\mathbf{s}}^{\mathrm{T}}\tilde{\mathbf{s}} = (\hat{\mathbf{C}}_N^{-1/2*}\tilde{\mathbf{u}})^{\mathrm{T}}(\hat{\mathbf{C}}_N^{-1/2*}\tilde{\boldsymbol{u}}) = \tilde{\mathbf{u}}^{\mathrm{T}}\hat{\mathbf{C}}_N^{-1}\tilde{\mathbf{u}}. \tag{A2}$$

It is argued in F01 [see Eq. (6) therein] that

$$E(r^2) = E(\breve{\mathbf{u}}^{\mathrm{T}}\hat{\mathbf{C}}_N^{-1}\breve{\mathbf{u}}) = \hat{\mathbf{u}}^{\mathrm{T}}\hat{\mathbf{C}}_N^{-1}\hat{\mathbf{u}} + mF_{m,\nu}, \tag{A3}$$

where $\breve{\mathbf{u}}$ represents the actual differences at the true model parameter settings and $\hat{\mathbf{u}}$ the estimated differences, being the actual differences at the best-fit model parameters (where r^2 is minimized), and m now represents the number of unknown model parameters. Here $m = 3$.

Defining r_{min}^2 as the minimum r^2 as model parameters are varied and Δr^2 as the excess of r^2 at an arbitrary location in parameter space over r_{min}^2, it is asserted [F01, their Eq. (7)] that

$$\Delta r^2 \sim mF_{m,\nu}. \tag{A4}$$

Given known observational values, all realizable whitened difference vectors $\tilde{\mathbf{s}}$ extend from the origin to a point on an $m = 3$ dimensional hypersurface (the parameter surface) embedded in the κ-dimensional space, the location of that surface depending on the observational values. In the κ-dimensional space, r_{min}^2 is the squared length of $\hat{\boldsymbol{s}}$, the vector $\tilde{\boldsymbol{s}}$ ending at the point on the surface nearest to the origin, representing the best-fit estimated differences $\hat{\mathbf{u}}^{\mathrm{T}}\hat{\mathbf{C}}_N^{-1}\hat{\mathbf{u}}$ in (A3):

$$r_{\mathrm{min}}^2 = \|\hat{\mathbf{s}}\|^2 = \hat{\mathbf{s}}^{\mathrm{T}}\hat{\mathbf{s}}. \tag{A5}$$

At this point, the parameter surface is orthogonal to $\hat{\boldsymbol{s}}$ (otherwise the minimum would lie elsewhere). An m-dimensional tangent hyperplane coincident with the parameter surface where it meets $\hat{\mathbf{s}}$, and therefore also orthogonal to $\hat{\mathbf{s}}$, can be constructed. If the parameter surface is flat, and so everywhere is coincident with the tangent hyperplane, Δr^2 will, as assumed in F06, have an $mF_{m,\nu}$ distribution per (A4). However, the parameter surface is more likely to be convex, resulting in an $mF_{m,\nu}$ distribution producing tighter bounds on the parameters than are justified (see supplemental material for additional discussion).

APPENDIX B

New Method

We start by defining whitened versions of the modeled, observed, and underlying surface diagnostic temperatures: $\mathbf{q}_{\rm sfc} = \hat{\mathbf{C}}_{N_{\rm sfc}}^{-1/2*}\mathbf{T}_{m_{\rm sfc}}(\boldsymbol{\theta}_m)$, $\tilde{\mathbf{w}}_{\rm sfc} = \hat{\mathbf{C}}_{N_{\rm sfc}}^{-1/2*}\tilde{\mathbf{T}}_{o_{\rm sfc}}$, and $\mathbf{w}_{\rm sfc}(\boldsymbol{\theta}_t) = \hat{\mathbf{C}}_{N_{\rm sfc}}^{-1/2*}\mathbf{T}_{o_{\rm sfc}}(\boldsymbol{\theta}_t)$. The underlying temperatures $\mathbf{T}_{o_{\rm sfc}}(\boldsymbol{\theta}_t)$ are what the observed temperatures would have been in the absence of climate noise and on the assumption of model accuracy are functions of the true parameter vector $\boldsymbol{\theta}_t$.

Since the whitening operation is linear the whitening may be carried out before calculating differences or after, so

$$\begin{aligned}\tilde{\mathbf{s}}_{\rm sfc} &= \hat{\mathbf{C}}_{N_{\rm sfc}}^{-1/2*}\tilde{\mathbf{u}} = \hat{\mathbf{C}}_{N_{\rm sfc}}^{-1/2*}[\tilde{\mathbf{T}}_{o_{\rm sfc}} - \mathbf{T}_{m_{\rm sfc}}(\boldsymbol{\theta}_m)]\\ &= \hat{\mathbf{C}}_{N_{\rm sfc}}^{-1/2*}\tilde{\mathbf{T}}_{o_{\rm sfc}} - \hat{\mathbf{C}}_{N_{\rm sfc}}^{-1/2*}\mathbf{T}_{m_{\rm sfc}}(\boldsymbol{\theta}_m) = \tilde{\mathbf{w}}_{\rm sfc} - \mathbf{q}_{\rm sfc}.\end{aligned} \quad \text{(B1)}$$

By supposition, variability in the whitened observed temperatures has a $N(0, \mathbf{I}_{\kappa_{\rm sfc}})$ distribution:

$$p[\tilde{\mathbf{w}}_{\rm sfc} \,|\, \mathbf{w}_{\rm sfc}(\boldsymbol{\theta}_t)] \propto \exp\left\{-\sum_{i=1}^{\kappa_{\rm sfc}} [\tilde{w}_{{\rm sfc}_i} - w_{{\rm sfc}_i}(\boldsymbol{\theta}_t)]^2/2\right\}. \quad \text{(B2)}$$

Now consider the κ_c vectors of whitened observed $\tilde{\mathbf{w}}_c = (\tilde{w}_{{\rm sfc}_1}, \ldots, \tilde{w}_{{\rm sfc}_{\kappa_{\rm sfc}}}, \tilde{w}_{{\rm ua}_1}, \ldots, \tilde{w}_{{\rm ua}_{\kappa_{\rm ua}}}, \tilde{w}_{\rm do})$, underlying $\mathbf{w}_c = (w_{{\rm sfc}_1}, \ldots, w_{{\rm sfc}_{\kappa_{\rm sfc}}}, w_{{\rm ua}_1}, \ldots, w_{{\rm ua}_{\kappa_{\rm ua}}}, w_{\rm do})$, and modeled $\mathbf{q}_c = (q_{{\rm sfc}_1}, \ldots, q_{{\rm sfc}_{\kappa_{\rm sfc}}}, q_{{\rm ua}_1}, \ldots, q_{{\rm ua}_{\kappa_{\rm ua}}}, q_{\rm do})$ temperatures for all three diagnostics combined, where $\kappa_c = \kappa_{\rm sfc} + \kappa_{\rm ua} + 1$, and define $\tilde{\mathbf{s}}_c = \tilde{\mathbf{w}}_c - \mathbf{q}_c$. On the assumption, implicit in F06's use of Bayesian updating, that the whitened differences for the three diagnostics are independent, we can obtain equations corresponding to (B2) for the other diagnostics and then multiply them to obtain a PDF for $\tilde{\mathbf{w}}_c$:

$$\begin{aligned}p[\tilde{\mathbf{w}}_c \,|\, \mathbf{w}_c(\boldsymbol{\theta}_t)] &\propto \exp\left\{-\sum_{i=1}^{\kappa_{\rm sfc}} [\tilde{w}_{{\rm sfc}_i} - w_{{\rm sfc}_i}(\boldsymbol{\theta}_t)]^2/2\right\}\\ &\times \exp\left\{-\sum_{i=1}^{\kappa_{\rm ua}} [\tilde{w}_{{\rm ua}_i} - w_{{\rm ua}_i}(\boldsymbol{\theta}_t)]^2/2\right\}\\ &\times \exp\{-[\tilde{w}_{\rm do} - w_{\rm do}(\boldsymbol{\theta}_t)]^2/2\}\end{aligned} \quad \text{(B3)}$$

or

$$p[\tilde{\mathbf{w}}_c \,|\, \mathbf{w}_c(\boldsymbol{\theta}_t)] \propto \exp\{-[\tilde{\mathbf{w}}_c - \mathbf{w}_c(\boldsymbol{\theta}_t)]^{\rm T}[\tilde{\mathbf{w}}_c - \mathbf{w}_c(\boldsymbol{\theta}_t)]/2\}. \quad \text{(B4)}$$

The density for $\tilde{\mathbf{w}}_c$ depends only on the pivot variable $(\tilde{\mathbf{w}}_c - \mathbf{w}_c)$, so $\mathbf{w}_c$ is a location parameter vector. The noninformative joint prior for estimating location parameters in the presence of independent normal errors of known variance is uniform (Datta and Sweeting 2005): $p(\mathbf{w}_c) = \text{constant}$. Applying Bayes's theorem at this point, we derive a posterior joint PDF for the underlying whitened temperatures as

$$p_{\mathbf{w}_c}[\mathbf{w}_c(\boldsymbol{\theta}_t) \,|\, \tilde{\mathbf{w}}_c] = p_{\tilde{\mathbf{w}}_c}[\tilde{\mathbf{w}}_c \,|\, \mathbf{w}_c(\boldsymbol{\theta}_t)]p_{\mathbf{w}_c}(\mathbf{w}_c)/p_{\tilde{\mathbf{w}}_c}(\tilde{\mathbf{w}}_c). \quad \text{(B5)}$$

Omitting constant values of the uniform prior and the denominator we obtain

$$p_{\mathbf{w}_c}[\mathbf{w}_c(\boldsymbol{\theta}_t) \,|\, \tilde{\mathbf{w}}_c] \propto p_{\tilde{\mathbf{w}}_c}[\tilde{\mathbf{w}}_c \,|\, \mathbf{w}_c(\boldsymbol{\theta}_t)], \quad \text{(B6)}$$

which, substituting from (B4), becomes

$$p_{\mathbf{w}_c}[\mathbf{w}_c(\boldsymbol{\theta}_t) \,|\, \tilde{\mathbf{w}}_c] \propto \exp\{-[\tilde{\mathbf{w}}_c - \mathbf{w}_c(\boldsymbol{\theta}_t)]^{\rm T}[\tilde{\mathbf{w}}_c - \mathbf{w}_c(\boldsymbol{\theta}_t)]/2\}. \quad \text{(B7)}$$

This Bayesian posterior is identical to the error distribution PDF when estimating the underlying whitened temperatures using a frequentist approach.

Given the assumption of model prediction accuracy $\mathbf{q}_c(\boldsymbol{\theta}_m)\,|_{\boldsymbol{\theta}_m=\boldsymbol{\theta}_t} = \mathbf{w}_c(\boldsymbol{\theta}_t)$, we can, since (B1) extended to the combined diagnostics then implies $\tilde{\mathbf{s}}_c(\boldsymbol{\theta}_m = \boldsymbol{\theta}_t) = \tilde{\mathbf{w}}_c - \mathbf{w}_c$, simplify (B7) to

$$p_{\mathbf{w}_c}[\mathbf{w}_c(\boldsymbol{\theta}_t) \,|\, \tilde{\mathbf{w}}_c] \propto \exp\{-[\tilde{\mathbf{s}}_c(\boldsymbol{\theta}_m = \boldsymbol{\theta}_t)]^{\rm T}[\tilde{\mathbf{s}}_c(\boldsymbol{\theta}_m = \boldsymbol{\theta}_t)]/2\}. \quad \text{(B8)}$$

The resulting joint posterior PDF for $\mathbf{w}_c$ can now be converted into a joint PDF for the parameters using a standard generalization [Mardia et al. 1979, their Eq. (2.5.16)] of the usual formula for converting PDFs upon a change of variables with unchanged dimensionality. If $\mathbf{x} = \boldsymbol{\phi}(\boldsymbol{\theta}), \boldsymbol{\theta} \in R^m, \mathbf{x} \in R^\kappa$ is a parameterization of a m-dimensional hypersurface in $R^\kappa (m \le \kappa)$ and $g(\boldsymbol{\theta})$ is a PDF on R^m, then $\mathbf{x}$ has a PDF on the hypersurface given by

$$f(\mathbf{x}) = g[\boldsymbol{\phi}^{-1}(\mathbf{x})]|\mathbf{D}^{\rm T}\mathbf{D}|^{-1/2}, \quad \text{(B9)}$$

where

$$\mathbf{D} = \mathbf{D}(\mathbf{x}) = \left[\frac{\partial \phi_i(\boldsymbol{\theta})}{\partial \theta_j}\right]\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\phi}^{-1}(\mathbf{x})} \qquad \text{(B10)}$$

is the $\kappa \times m$ Jacobian matrix evaluated at $\boldsymbol{\theta} = \boldsymbol{\phi}^{-1}(\mathbf{x})$. Here, we take $g(\boldsymbol{\theta}) = p(\boldsymbol{\theta}_t)$, the joint PDF for the true parameter vector, and $\mathbf{x} = \mathbf{w}_c(\boldsymbol{\theta}_t)$ so that $\boldsymbol{\phi}^{-1}(\mathbf{x}) = \boldsymbol{\theta}_t$ and we make the PDFs conditional on $\tilde{\mathbf{w}}_c$ (which does not affect $\mathbf{D}$), with the result that (B9) becomes

$$p[\mathbf{w}_c(\boldsymbol{\theta}_t) \mid \tilde{\mathbf{w}}_c] = p(\boldsymbol{\theta}_t|\tilde{\mathbf{w}}_c)|\mathbf{D}^{\mathrm{T}}\mathbf{D}|^{-1/2}, \qquad \text{(B11)}$$

where

$$\mathbf{D} = \mathbf{D}[\mathbf{w}_c(\boldsymbol{\theta}_t)] = \left[\frac{\partial w_{c_i}(\boldsymbol{\theta}_t)}{\partial \theta_{t_j}}\right]\Bigg|_{\boldsymbol{\theta}_t}. \qquad \text{(B12)}$$

Rearranging (B11),

$$p(\boldsymbol{\theta}_t \mid \tilde{\mathbf{w}}_c) = p[\mathbf{w}_c(\boldsymbol{\theta}_t) \mid \tilde{\mathbf{w}}_c]|\mathbf{D}^{\mathrm{T}}\mathbf{D}|^{1/2}. \qquad \text{(B13)}$$

By the presumed model prediction accuracy and (B1)

$$\mathbf{w}_c(\boldsymbol{\theta}_t) = \mathbf{q}_c(\boldsymbol{\theta}_m|\boldsymbol{\theta}_m = \boldsymbol{\theta}_t) = \tilde{\mathbf{w}}_c(\tilde{\mathbf{T}}_{o_c}) - \tilde{\mathbf{s}}_c(\tilde{\mathbf{T}}_{o_c}, \boldsymbol{\theta}_m|\boldsymbol{\theta}_m = \boldsymbol{\theta}_t), \qquad \text{(B14)}$$

and noting that $\tilde{\mathbf{w}}_c$ is not a function of $\boldsymbol{\theta}$ we can therefore change (B12) to

$$\mathbf{D} = -\left[\frac{\partial \tilde{s}_{c_i}(\boldsymbol{\theta}_m)}{\partial \theta_{m_j}}\right]\Bigg|_{\boldsymbol{\theta}_m=\boldsymbol{\theta}_t} \qquad \text{(B15)}$$

Substituting now from Eq. (B6) in Eq. (B13), noting that conditionality on $\mathbf{w}_c(\boldsymbol{\theta}_t)$ is the same as conditionality on $\boldsymbol{\theta}_t$, gives

$$p(\boldsymbol{\theta}_t|\tilde{\mathbf{w}}_c) \propto p(\tilde{\mathbf{w}}_c|\boldsymbol{\theta}_t)|\mathbf{D}^{\mathrm{T}}\mathbf{D}|^{1/2}, \qquad \text{(B16)}$$

where $|\mathbf{D}^{\mathrm{T}}\mathbf{D}|^{1/2}$ represents a conversion factor $\pi(\boldsymbol{\theta}_t)$ from probability density in whitened observation space to that in parameter space (i.e., on the parameter surface).

REFERENCES

Aldrin, M., M. Holden, P. Guttorp, R. B. Skeie, G. Myhre, and T. K. Berntsen, 2012: Bayesian estimation of climate sensitivity based on a simple climate model fitted to observations of hemispheric temperatures and global ocean heat content. *Environmetrics,* **23,** 253–271.

Allen, M. R., and S. F. B. Tett, 1999: Checking internal consistency in optimal fingerprinting. *Climate Dyn.,* **15,** 419–434.

Andronova, N. G., and M. E Schlesinger, 2001: Objective estimation of the probability density function for climate sensitivity. *J. Geophys. Res.,* **106** (D19), 22 605–22 611.

Bernardo, J. M., and A. F. M. Smith, 1994: *Bayesian Theory.* Wiley, 608 pp.

Box, G. E. P., and G. C. Tiao, 1973: *Bayesian Inference in Statistical Analysis.* Addison-Wesley, 588 pp.

Curry, C. T., 2007: Inference for climate system properties. M.S. Science, Dept. of Applied Mathematics & Statistics, University of California, Santa Cruz, 40 pp. [Available online at http://www.webcitation.org/68VqaIbVK.]

——, B. Sansó, and C. E. Forest, 2005: Inference for climate system properties. Applied Mathematics and Statistics, University of California, Santa Cruz, Tech. Rep. Ams2005-13, 7 pp. [Available online at http://www.soe.ucsc.edu/research/technical-reports/ams2005-13.]

Datta, G. S., and T. J. Sweeting, 2005: Probability matching priors. *Handbook of Statistics,* Vol. 25, D. K. Dey and C. R. Rao, Eds., Elsevier, 91–114.

Delworth, T. L., R. L. Stouffer, K. W. Dixon, M. J. Spelman, T. R. Knutson, A. J. Broccoli, P. J. Kusher, and R. T. Wetherald, 2002: Review of simulations of climate variability and change with the GFDL R30 coupled climate model. *Climate Dyn.,* **19,** 555–574.

Drignei, D., C. Forest, and D. Nychka, 2008: Parameter estimation for computationally intensive nonlinear regression with an application to climate modeling. *Ann. Appl. Stat.,* **2,** 1217–1230, doi:10.1214/08-AOAS210.

Forest, C. E., M. R. Allen, P. H. Stone, and A. P. Sokolov, 2000: Constraining uncertainties in climate models using climate change detection methods. *Geophys. Res. Lett.,* **27,** 569–572.

——, ——, A. P. Sokolov, and P. H. Stone, 2001: Constraining climate model properties using optimal fingerprint detection methods. *Climate Dyn.,* **18,** 277–295.

——, P. H. Stone, A. P. Sokolov, M. R. Allen, and M. D. Webster, 2002: Quantifying uncertainties in climate system properties with the use of recent climate observations. *Science,* **295,** 113–117.

——, ——, and ——, 2006: Estimated PDFs of climate system properties including natural and anthropogenic forcings. *Geophys. Res. Lett.,* **33,** L01705, doi:10.1029/2005GL023977.

——, ——, and ——, 2008: Constraining climate model parameters from observed 20th century changes. *Tellus,* **60A,** 911–920.

Forster, P. M. de F., and J. M. Gregory, 2006: The climate sensitivity and its components diagnosed from Earth radiation budget data. *J. Climate,* **19,** 39–52.

Frame, D. J., B. B. B. Booth, J. A. Kettleborough, D. A. Stainforth, J. M. Gregory, M. Collins, and M. R. Allen, 2005: Constraining climate forecasts: The role of prior assumptions. *Geophys. Res. Lett.,* **32,** L09702, doi:10.1029/2004GL022241.

Gordon, C., C. Cooper, C. A. Senior, H. Banks, J. M. Gregory, T. C. Johns, J. F. B. Mitchell, and R. A. Wood, 2000: The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dyn.,* **16,** 147–168.

Gregory, J., R. J. Stouffer, S. C. B. Raper, P. A. Stott, and N. A. Rayner, 2002: An observationally based estimate of the climate sensitivity. *J. Climate,* **15,** 3117–3121.

Hegerl, G. C., T. C. Crowley, W. T. Hyde, and D. J. Frame, 2006: Climate sensitivity constrained by temperature reconstructions over the past seven centuries. *Nature,* **440,** 1029–1032, doi:10.1038/nature04679.

——, 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.

Jeffreys, H., 1946: An invariant form for the prior probability in estimation problems. *Proc. Roy. Soc. London,* **186A,** 453–461.

Jewson, S., D. Rowlands, and M. Allen, cited 2009: A new method for making objective probabilistic climate forecasts from numerical climate models based on Jeffreys' Prior. [Available online at http://arxiv.org/pdf/0908.4207.pdf.]

Johns, T. C., R. E. Carnell, J. F. Crossley, and J. M. Gregory, 1997: The second Hadley Centre coupled ocean-atmosphere GCM: Model description, spinup and validation. *Climate Dyn.,* **13,** 103–134.

Jones, P. D., M. New, D. E. Parker, S. Martin, and I. G. Rigor, 1999: Surface air temperature and its changes over the past 150 years. *Rev. Geophys.,* **37,** 173–199.

Kass, R. E., 1989: The geometry of asymptotic inference. *Stat. Sci.,* **4,** 188–219.

——, and L. Wasserman, 1996: The selection of prior distributions by formal rules. *J. Amer. Stat. Assoc.,* **91,** 1343–1370.

Knutti, R., T. F. Stocker, F. Joos, and G.-K. Plattner, 2002: Constraints on radiative forcing and future climate change from observations and climate model ensembles. *Nature,* **416,** 719–723.

Levitus, S., J. Antonov, and T. Boyer, 2005: Warming of the world ocean, 1955–2003. *Geophys. Res. Lett.,* **32,** L02604, doi:10.1029/2004GL021592.

Libardoni, A. G., and C. E. Forest, 2011: Sensitivity of distributions of climate system properties to the surface temperature dataset. *Geophys. Res. Lett.,* **38,** L22705, doi:10.1029/2011GL049431.

Mardia, K. V., J. T. Kent, and J. M. Bibby, 1979: *Multivariate Analysis.* Academic Press, 518 pp.

Morice, C. P., J. J. Kennedy, N. A. Rayner, and P. D. Jones, 2012: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 dataset. *J. Geophys. Res.,* **117,** D08101, doi:10.1029/2011JD017187.

Mosegaard, K., and A. Tarantola, 2002: Probabilistic approach to inverse problems. *International Handbook of Earthquake and Engineering Seismology,* W. H. K. Lee et al., Eds., International Geophysics Series, Vol. 81A, Elsevier, 237–265.

Parker, D. E., M. Gordon, D. P. N. Cullum, D. M. H. Sexton, C. K. Folland, and N. Rayner, 1997: A new global gridded radiosonde temperature data base and recent temperature trends. *Geophys. Res. Lett.,* **24,** 1499–1502.

Pueyo, S., 2012: Solution to the paradox of climate sensitivity. *Climatic Change,* **113,** 163–179, doi:10.1007/s10584-011-0328-x.

Ring, M. J., D. Lindner, E. F. Cross, and M. E. Schlesinger, 2012: Causes of the global warming observed since the 19th century. *Atmos. Climate Sci.,* **2,** 401–415.

Roe, G. H., and M. B. Baker, 2007: Why is climate sensitivity so unpredictable? *Science,* **318,** 629–632.

Sansó, B., and C. Forest, 2009: Statistical calibration of climate system properties. *J. Roy. Stat. Soc.,* **58C,** 485–503.

——, C. E. Forest, and D. Zantedeschi, 2008: Inferring climate system properties using a computer model (with discussion). *Bayesian Anal.,* **3,** 1–62.

Sokolov, A. P., and P. H. Stone, 1998: A flexible climate model for use in integrated assessments. *Climate Dyn.,* **14,** 291–303.

——, C. E. Forest, and P. H. Stone, 2003: Comparing oceanic heat uptake in AOGCM transient climate change experiments. *J. Climate,* **16,** 1573–1582.