# Correction to "Sensitivity of distributions of climate system properties to the surface temperature data set"*
# &
# Sensitivity of distributions of climate system properties to the surface temperature data set**

Alex G. Libardoni and Chris E. Forest


MIT JOINT PROGRAM ON THE SCIENCE AND POLICY of GLOBAL CHANGE

*Reprinted from
*Geophysical Research Records*, 40(10): 2309–2311


**Reprinted from
*Geophysical Research Records*, 38(22): 1–6


Reprint 2013-9

The MIT Joint Program on the Science and Policy of Global Change combines cutting-edge scientific research with independent policy analysis to provide a solid foundation for the public and private decisions needed to mitigate and adapt to unavoidable global environmental changes. Being data-driven, the Program uses extensive Earth system and economic data and models to produce quantitative analysis and predictions of the risks of climate change and the challenges of limiting human influence on the environment—essential knowledge for the international dialogue toward a global response to climate change.

To this end, the Program brings together an interdisciplinary group from two established MIT research centers: the Center for Global Change Science (CGCS) and the Center for Energy and Environmental Policy Research (CEEPR). These two centers—along with collaborators from the Marine Biology Laboratory (MBL) at Woods Hole and short- and long-term visitors—provide the united vision needed to solve global challenges.

At the heart of much of the Program's work lies MIT's Integrated Global System Model. Through this integrated model, the Program seeks to: discover new interactions among natural and human climate system components; objectively assess uncertainty in economic and climate projections; critically and quantitatively analyze environmental management and policy proposals; understand complex connections among the many forces that will shape our future; and improve methods to model, monitor and verify greenhouse gas emissions and climatic impacts.

This reprint is one of a series intended to communicate research results and improve public understanding of global environment and energy challenges, thereby contributing to informed debate about climate change and the economic and social implications of policy alternatives.

Ronald G. Prinn and John M. Reilly,
Program Co-Directors

For more information, contact the Program office:

MIT Joint Program on the Science and Policy of Global Change
**Postal Address:**
Massachusetts Institute of Technology
77 Massachusetts Avenue, E19-411
Cambridge, MA 02139 (USA)
**Location:**
Building E19, Room 411
400 Main Street, Cambridge
**Access:**
Tel: (617) 253-7492
Fax: (617) 253-9845
Email: ***globalchange@mit.edu***
Website: ***http://globalchange.mit.edu/***

# Correction to "Sensitivity of distributions of climate system properties to the surface temperature data set"

Alex G. Libardoni[1] and Chris E. Forest[2]

Received 6 March 2013; revised 9 April 2013; accepted 17 April 2013; published 30 May 2013.

**Citation:** Libardoni, A. G., and C. E. Forest (2013), Correction to "Sensitivity of distributions of climate system properties to the surface temperature data set," *Geophys. Res. Lett.*, *40*, 2309–2311, doi:10.1002/grl.50480.

[1] In the paper "Sensitivity of distributions of climate system properties to the surface temperature data set" by A. G. Libardoni and C. E. Forest (*Geophys. Res. Lett.*, *38*, L22705, doi:10.1029/2011GL049431), two errors were made. First, there was an offset of 1 month when comparing model data to observational data. Second, the likelihood function that was used to relate model goodness-of-fit statistics to a probability distribution, while derived from basic understanding of probability distributions, is not acceptable as a likelihood function to statisticians.

[2] In this correction, we fix the mismatch that existed between model and observational annual averages. When calculating annual averages from model output, seasonal means were averaged, resulting in a given year being the average from December through November. When calculating annual averages from observational data, monthly means were averaged, resulting in a given year being the average from January through December. To correct for this 1 month mismatch, all annual mean temperatures derived from observations are calculated as December to November means, subject to the threshold criterion described in *Libardoni and Forest* [2011]. Because decadal mean temperatures are used for the surface temperature diagnostic, the 1 month shift in the averaging window has minimal impact on the resulting observational time series. Across all five decades and four zonal bands, the temperature differences due to the 1 month shift are at most 0.05°C. The revised time series are not shown.

[3] We also present revised results that implement a likelihood function proposed in *Lewis* [2013] that is more statistically sound as applied to the Bayesian methodology used in *Libardoni and Forest* [2011]. It will be shown below that the updated likelihood function alters the posterior distributions. The changes to the likelihood function involve changing the shape of the distributions used in statistical tests, changing the test statistic, and taking into account the necessary volumetric correction when making a change of variable. In total, the following changes were made to account for the likelihood estimate from *Lewis* [2013]:

[4] 1. We estimate the likelihood from goodness-of-fit statistics using the probability density function (PDF) of an $F$ distribution, as opposed to 1 minus the cumulative distribution function (CDF) of an $F$ distribution for the surface and upper-air diagnostics.

[5] 2. We use a $t$ distribution for $\Delta r$, rather than an $F$ distribution for $\Delta r^2$, for the ocean diagnostic.

[6] 3. We change the degrees of freedom in the statistical distributions to $\kappa$, the number of EOFs retained in estimates of the noise-covariance matrices, and $\nu$, the number of control run segments available to make these estimates, respectively. This results in a change from 3 and 24 to 16 and 49 degrees of freedom, respectively for the surface diagnostic, from 3 and 14 to 14 and 39 degrees of freedom respectively for the upper-air diagnostic, and from 3 and 24 degrees of freedom in an $F$ distribution to an effective degrees of freedom of 4.1 in a $t$ distribution for the ocean diagnostic.

[7] 4. We change the test statistic in the $F$ distribution for the surface and upper-air diagnostics from $\frac{\Delta r^2}{3}$ to $\frac{\Delta r^2}{\kappa}$, where $\kappa$ are 16 and 14 for the surface and upper-air diagnostics, respectively.

[8] 5. We multiply the likelihood from the $F$ distribution by $(\Delta r^2)^{-\left(\frac{\kappa}{2}-1\right)}$ to account for the transformation from the data space ($\Delta r^2$ values) to the model parameter space.

[9] When calculating a likelihood function, a probability density function should be used to calculate the likelihood value for a given value of the test statistic. In previous work, the corresponding cumulative density function was used and the likelihood was estimated as the probability of obtaining values greater than the test statistic as typically done for hypothesis testing. Incorporating the five changes presented above corrects the error from previous work and, as described in *Lewis* [2013], results in a likelihood function that ensures that a probability density function is now used. The changes, however, do not implement a switch from $\Delta r^2$ to $r^2$ in the test statistic as proposed in *Lewis* [2013]. Making this change would represent a change in the noise model that is not necessarily justified. Similar to assumptions used in linear regression, shifting from $r^2$ to $\Delta r^2$ implies that the noise model for the residual variability is appropriate in the vicinity of $\Delta r^2 = 0$, rather than assuming it is appropriate for all values of $r^2$.

[10] The differences between the likelihood functions discussed above, as implemented by each study for the surface diagnostic, indicate a stronger rejection of higher $\Delta r^2$ values using the *Lewis* [2013] likelihood function (Figure 1). Similar results hold true in the likelihood functions for the upper-air and deep ocean temperature diagnostics.

[1]Department of Meteorology, Pennsylvania State University, University Park, Pennsylvania, USA.

[2]Department of Meteorology, Earth and Environmental Systems Institute, Pennsylvania State University, University Park, Pennsylvania, USA.

Corresponding author: C. E. Forest, Pennsylvania State University507 Walker Building, University Park, PA 16802, USA. (ceforest@psu.edu)


0094-8276/13/10.1002/grl.50480

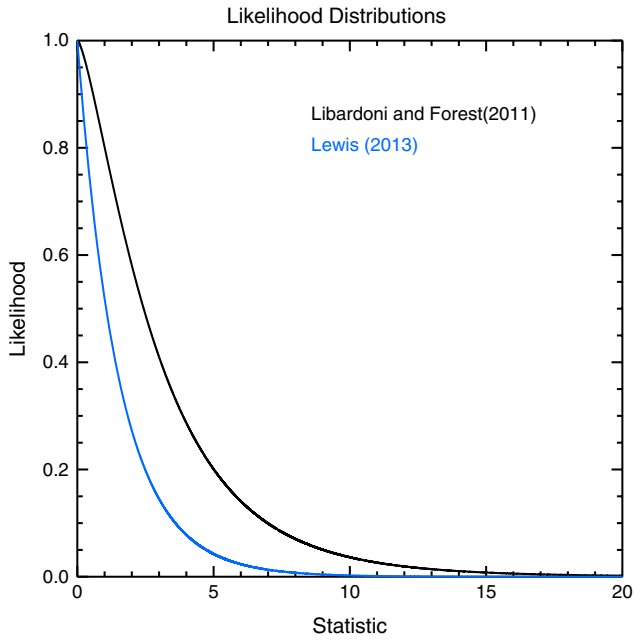


**Figure 1.** Likelihood functions for the surface diagnostic statistic using the (1 minus CDF) method from *Libardoni and Forest* [2011] and the method from *Lewis* [2013]. The test statistic used in *Libardoni and Forest* [2011] is $\frac{\Delta r^2}{3}$, while *Lewis* [2013] uses $\frac{\Delta r^2}{\kappa}$, where $\kappa = 16$.

[11] We present a revised Figure 3 from *Libardoni and Forest* [2011] to show the impact of using the likelihood from *Lewis* [2013] (Figure 2). Marginal distributions for each parameter are presented using both the original method and the corrected method. Updated parameter distributions have not been included for the results from *Forest et al.* [2008] because the previous results were provided only for comparison purposes in *Libardoni and Forest* [2011] and not re-derived using the new methods in *Libardoni and Forest* [2011]. The corrected method leads to shifts in climate sensitivity posteriors due to the likelihood function change, however, the general shifts in the resulting cumulative distribution functions are small compared with the ranges due to other factors such as the observational data source [*Libardoni and Forest*, 2011], confirming the original results (Figure 3). In general, the distribution modes are more pronounced under the new likelihood function, a compensating narrowing of the distribution is present, and the lower bounds of the distributions show small increases.

[12] When testing the sensitivity of the distributions to individual changes in the likelihood function, it was found that changing from 1 minus the CDF of the *F* distribution to the PDFs of *t* and *F* distributions for the diagnostics led to a narrowing of the distributions. Increasing the degrees of freedom in the *t* and *F* distributions led to a broadening of the distributions and a shift towards lower climate sensitivity and aerosol forcing values. The net impact of the changes results in the observed narrowing of the marginal distributions when using the *Lewis* [2013] method. After all

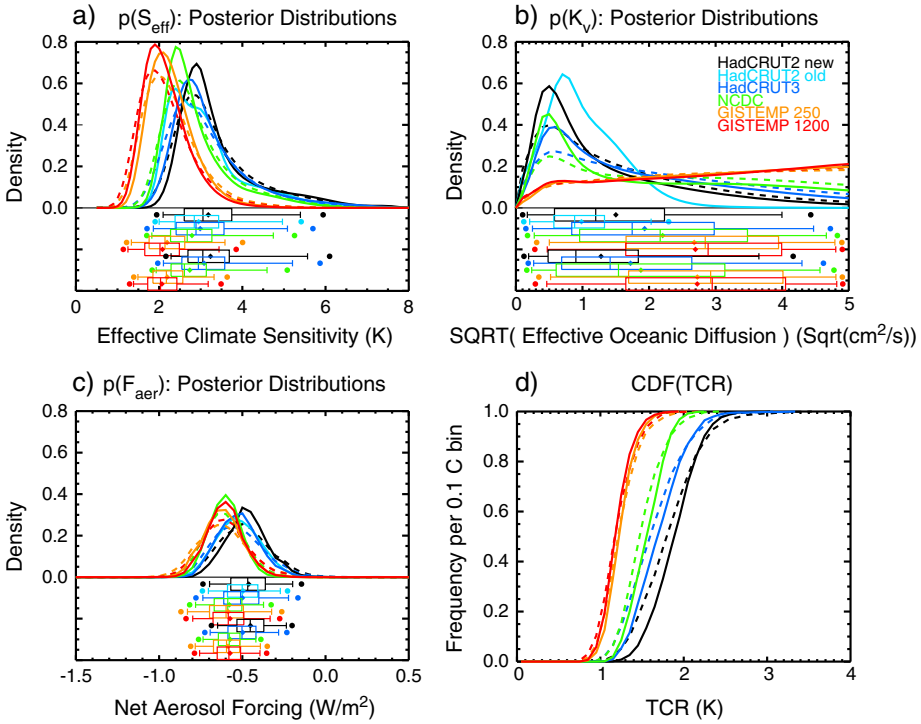


**Figure 2.** (a–d) Corrected Figure 3 from *Libardoni and Forest* [2011] which includes marginal probability distribution functions for each parameter and cumulative distribution functions for transient climate response using the different data sets. When two distributions are given for a particular data set, solid lines use the likelihood function from *Lewis* [2013] and dashed lines use the likelihood function from *Libardoni and Forest* [2011]. An update for HadCRUT2 old [*Forest et al.*, 2008] is not provided as discussed in the text.

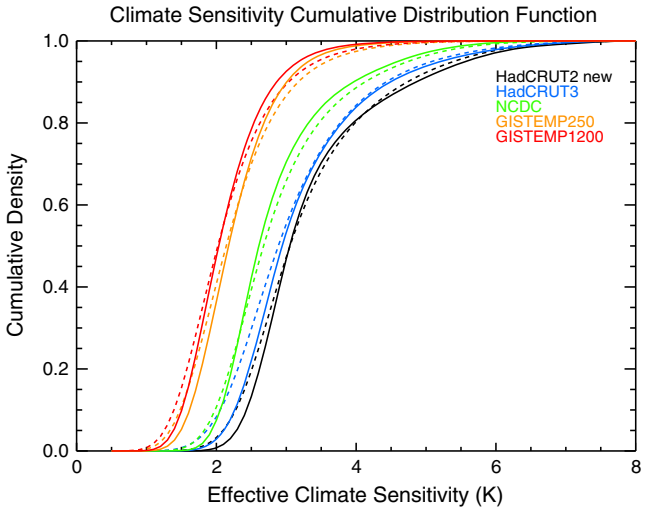


**Figure 3.** Corrected Figure 9 of the Auxiliary Material from *Libardoni and Forest* [2011] which shows cumulative distribution functions for climate sensitivity using the different data sets. When two distributions are given for a particular data set, solid lines use the likelihood function from *Lewis* [2013] and dashed lines use the likelihood function from *Libardoni and Forest* [2011]. An update for HadCRUT2 old [*Forest et al.*, 2008] is not provided as discussed in the text.

contributions from the individual changes are incorporated, the previously mentioned net narrowing of the distributions is observed (Figure 2).

[13] **Acknowledgments.** We acknowledge Nicholas Lewis for bringing the temporal mismatch between observation and model annual averages and the statistical errors to our attention.

## References

Forest, C. E., P. H. Stone, and A. P. Sokolov (2008), Constraining climate model parameters from observed 20th century changes, *Tellus*, *60A*, 911–920.

Lewis, N. (2013), An objective Bayesian, improved approach for applying optimal fingerprint techniques to climate sensitivity, *J. Clim.*, doi:10.1175/JCLI-D-12-00473.1, in press.

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.

# Sensitivity of distributions of climate system properties to the surface temperature dataset

Alex G. Libardoni[1] and Chris E. Forest[2]

Received 1 September 2011; revised 6 October 2011; accepted 11 October 2011; published 19 November 2011.

[1] Surface temperature, upper-air temperature, and ocean heat content data are used to constrain the distributions of the parameters that define three climate system properties in the MIT Integrated Global Systems Model: effective climate sensitivity, the rate of ocean heat uptake into the deep ocean, and net anthropogenic aerosol forcing. Five different surface temperature data records are used to show that the resulting parameter distribution functions are sensitive to the dataset used to estimate the likelihood of model output given the observed climate records. Estimates of effective climate sensitivity mode and mean differ by as much as 1 K between the datasets, with an overall range of 1.2 to 5.3 K. Ocean effective diffusivity distributions are poorly constrained by any dataset. The overall range of net aerosol forcing values, −0.19 to −0.83 $Wm^{-2}$, is small compared to other uncertainties in climate forcings. Transient climate response (TCR) estimates derived from these distributions range between 0.87 and 2.41 K and the shapes of individual TCR distributions depend on the surface dataset. Understanding the differences in parameter distributions and climate system properties derived from them is critical for understanding the full range of uncertainty involved in climate model calibration and prediction results. **Citation:** 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.

## 1. Introduction

[2] Developing climate models that produce reliable projections of future climate change is a critical research goal. To this end, models must be properly calibrated to have values of climate system properties that yield behavior similar to the true climate system. Uncertainties in the physical processes and feedbacks that define the climate system properties and resulting behavior introduce additional challenges into the calibration problem [*Randall et al.*, 2007]. Due to the multiple uncertainties, joint probability distributions are derived for the parameters used in the model rather than estimating individual values.

[3] Multiple studies have derived probability distribution functions for uncertain model parameters [*Andronova and Schlesinger*, 2001; *Forest et al.*, 2002, 2008; *Knutti et al.*, 2003; *Tomassini et al.*, 2007; *Sansó and Forest*, 2009; *Urban and Keller*, 2010]. While the specific approaches differ, the same general methodology is used for deriving the distribution functions. Using the efficiency of simple climate models and Earth Systems Models of Intermediate Complexity, hundreds of model runs are used to calibrate model parameters and resulting properties by comparing model output to observational data. Equilibrium climate sensitivity has been extensively studied. A synthesis of current work provided in the IPCC AR4 [*Hegerl et al.*, 2007] estimates climate sensitivity to be between 2.0 and 4.5 °C with a greater than 66-percent probability. However, several studies have shown that climate sensitivities can lie outside of the IPCC upper bound [*Andronova and Schlesinger*, 2001; *Forest et al.*, 2002, 2008; *Knutti et al.*, 2003; *Tomassini et al.*, 2007; *Tanaka et al.*, 2009; *Urban and Keller*, 2010] and a summary of sensitivity estimates are given by *Knutti and Hegerl* [2008]. Given that climate sensitivity is a measure of future warming, these uncertainties have a profound impact on policy decisions [*McInerney and Keller*, 2008].

[4] Transient climate response (TCR) offers an alternative metric of future climate behavior and is a function of both climate sensitivity and the rate of ocean heat uptake [*Sokolov et al.*, 2003; *Andrews and Allen*, 2008]. The IPCC AR4 estimates TCR to lie between 1 and 3 K [*Hegerl et al.*, 2007]. These bounds encompass TCR distributions derived in other studies and TCRs from individual AOGCMs [*Stott and Forest*, 2007; *Forest et al.*, 2008; *Knutti and Tomassini*, 2008]. Defined as the change in global-mean surface temperature at the time of $CO_2$ doubling in response to increasing $CO_2$ concentrations at 1% per year, TCR allows for all climate processes to be active and contribute fully to climate change.

[5] This study explores the impact that the surface temperature dataset used to compare model output to observed values has on the parameter constraints. To date, few studies have investigated how the surface temperature dataset used to compare model output with observational data impacts the parameter and TCR distributions. In total, five surface temperature data records representing three well-known climate centers are used in this study. Estimates of TCR are also investigated from the parameter distributions derived from each dataset. The resulting distributions show that model calibration is sensitive to the specific surface temperature dataset. Section 2 describes the datasets used and Section 3 describes the methods by which constraints are placed on parameter values. Section 4 presents the results with a discussion and summary in Section 5.

## 2. Data

[6] We use surface temperature data from five climate data records. The first two data records are HadCRUT2 [*Jones and Moberg*, 2003] and HadCRUT3 [*Brohan et al.*,

[1]Department of Meteorology, Pennsylvania State University, University Park, Pennsylvania, USA.
[2]Department of Meteorology, Earth and Environmental Systems Institute, Pennsylvania State University, University Park, Pennsylvania, USA.


0094-8276/11/2011GL049431

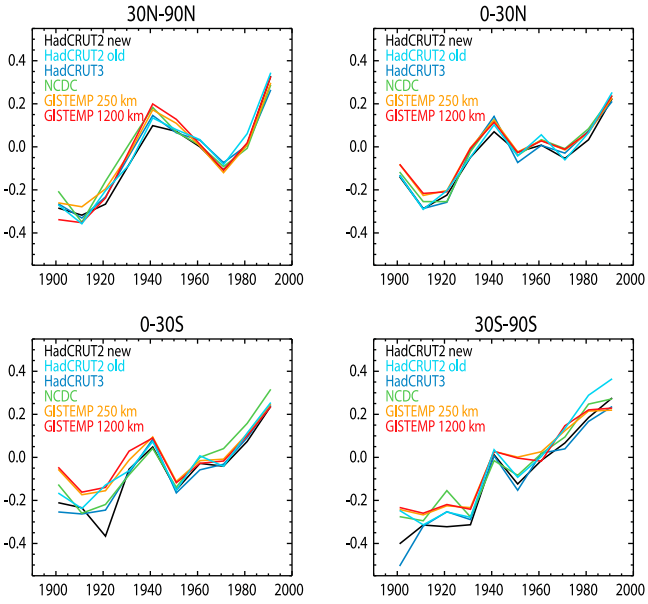


**Figure 1.** Decadal mean temperature time series derived from HadCRUT2 (black), HadCRUT3 (blue), NCDC (green), GISTEMP 250 (orange), and GISTEMP 1200 (red) surface data. Data used by *Forest et al.* [2008] is also shown (cyan). Temperatures are reported as anomalies with respect the 1906-1995 base period.

2006]. The third is the NCDC merged land-ocean dataset [*Smith et al.*, 2008]. The remaining two records are GISTEMP 250 and GISTEMP 1200 [*Hansen et al.*, 2010] from NASA, with the distinctions reflecting the 250 km and 1200 km radii of influence used in the interpolation algorithm. All data are reported as monthly surface temperature anomalies with respect to a given base period on a 5° × 5° grid. The data records differ from one another and potential reasons for these differences are now discussed briefly.

[7] One difference between the records is the land surface data used in the analyses. All records obtain a majority of their land surface data from the Global Historical Climatology Network (GHCN) [*Peterson and Vose*, 1997], but each utilizes the available data differently. For example, the Hadley Centre requires stations to have sufficient data between 1961 and 1990, their climate normal period, to be used in the analysis [*Jones and Moberg*, 2003; *Brohan et al.*, 2006]. Alternatively, NASA requires that stations have a period of overlap of at least 20 years with stations inside of a 1200 km radius to be used in the analysis [*Hansen et al.*, 2010]. A second difference between the data records is that each uses a different sea surface temperature (SST) dataset. Because oceans cover 70% of the Earth's surface, these choices lead to differences between the temperature data records [*Smith et al.*, 2008]. In a test of the sensitivity to ocean data choice, *Hansen et al.* [2010] showed that the global mean temperature calculated from GISTEMP data is affected by the choice of SST data. A last difference between the data records is the method used for filling regions with missing data and how the 5° × 5° grid box anomalies are calculated. Specific details of infilling and grid box averaging methods for each data record can be found in corresponding references. At this stage, we have five surface temperature data records and choose to treat them each as equally plausible. We present the results derived from each of them and do not attempt to merge the results into a single posterior distribution.

## 3. Methods

[8] Following the work of *Forest et al.* [2008], this study estimates the joint probability distribution of climate model parameters for effective climate sensitivity ($S_{eff}$), effective ocean diffusivity of heat anomalies ($K_v$), and net anthropogenic aerosol forcing ($F_{aer}$). Using the climate model component of the MIT Integrated Global Systems Model [*Sokolov and Stone*, 1998; *Sokolov et al.*, 2005], the model simulates historical temperature responses given choices of the three model parameters. In this study, the parameter space is sampled by varying $S_{eff}$ between 0 and 8 K, $K_v$ between 0 and 25 $cm^2s^{-1}$, and $F_{aer}$ between −1.5 and 0.5 $Wm^{-2}$. The value of $F_{aer}$ sets the amplitude of the net anthropogenic aerosol forcing in the 1980s in a spatially prescribed forcing pattern and is scaled by historical emissions to represent all unmodeled forcings in simulations [*Forest et al.*, 2001]. Each model run is forced by historical records of greenhouse gas concentrations, sulfate aerosol loadings, tropospheric and stratospheric ozone concentrations, solar irradiance changes, and stratospheric aerosols from volcanic eruptions [*Forest et al.*, 2008]. Model performance under a given set of parameter values is evaluated through comparison of model output to historical data using surface temperature, upper-air temperature, and ocean heat content diagnostics as described by *Forest et al.* [2008].

## 4. Results

[9] Time series of temperature for each zonal band used in the surface temperature diagnostic have been derived using the averaging techniques described in the auxiliary material and are shown in Figure 1, along with the time series used by *Forest et al.* [2008] from *Allen et al.* [2000].[1] Although also derived from HadCRUT2 data, it is important to note that the time series from *Allen et al.* [2000] is different than the time series derived in this study and is not used to allow for identical treatment of each dataset. In general, the patterns in each zonal band are similar, with the sign of the temperature change consistent across a majority of the decades for each dataset. However, the magnitudes of the changes differ. In particular, agreement is weakest among trends in the Southern Hemisphere, particularly for the 30–90 °S region. This is not surprising given that SST datasets differ between the records and a greater fraction of the Earth's surface is ocean in the Southern Hemisphere. Differences are also related to the factors discussed in Section 2.

[10] Another significant difference is the linear temperature trend observed from beginning to end in the period used in the surface diagnostic, 1946–1995. In all zonal bands, GISTEMP datasets show either similar or weaker warming trends than the other datasets. This is most evident in the 30–90 °S zonal band, where in the first decade of GISTEMP

[1] Auxiliary materials are available in the HTML. doi:10.1029/2011GL049431.

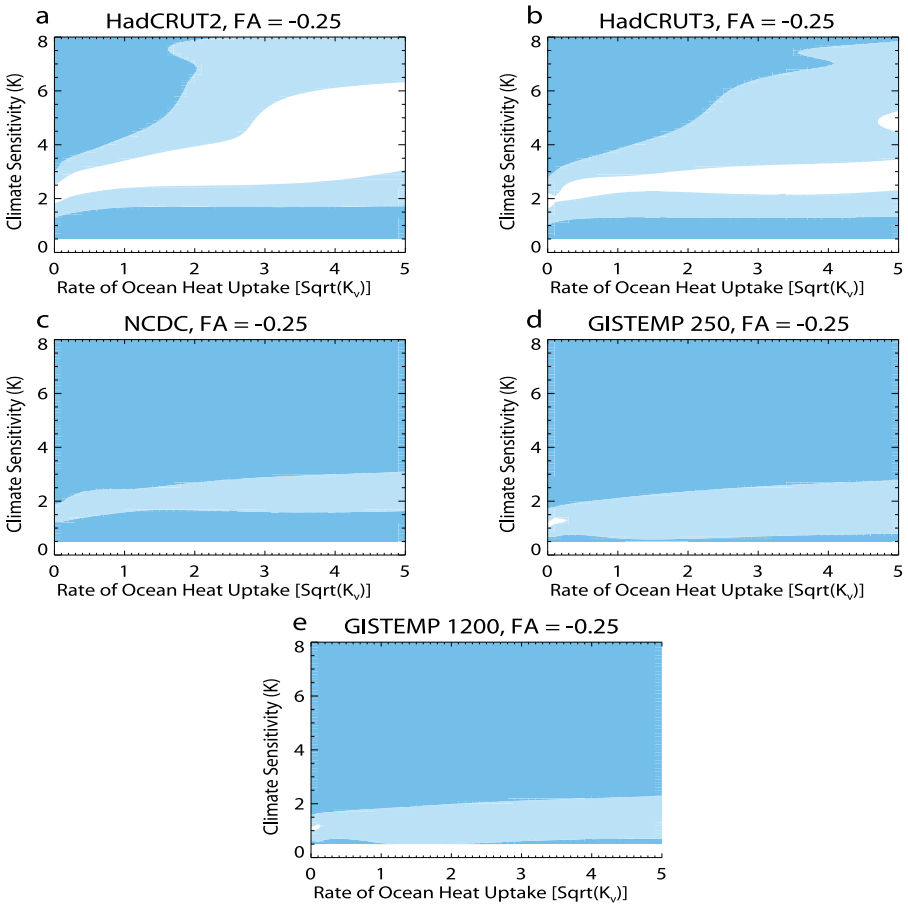


**Figure 2.** Goodness-of-fit statistics for surface diagnostics using (a) HadCRUT2, (b) HadCRUT3, (c) NCDC, (d) GISTEMP 250, and (e) GISTEMP 1200 surface data at an aerosol forcing of −0.25 $Wm^{-2}$. 90- (white) and 99-percent (light blue) confidence regions are shown. Areas outside of the 99-percent confidence region (dark blue) are rejected at the 1-percent significance level for being inconsistent with the observed data. All likelihoods are based off of an F-test.

data are the warmest, yet in the last decade GISTEMP data are the coldest relative to other records. Similar, yet weaker, patterns hold in the remaining zonal bands. In general, the NCDC time series yields the next weakest warming trends, followed by the HadCRUT2 and HadCRUT3 datasets. However, the extent of the differences is much less pronounced than with the GISTEMP datasets and the rank order of the trends is not consistent across all zonal bands.

[11] Using the derived time series, new surface temperature diagnostics are calculated for each model run and used to estimate of a goodness-of-fit statistic ($r^2$) which evaluates how well model output matches the observed temperature trends. For a fixed $F_{aer}$ and $K_v$, the model is varied over ten values of $S_{eff}$. For fixed values of $F_{aer}$ and $K_v$, the $r^2$ values are smoothed by fitting a sixth-order polynomial to the ten data points obtained when varying the values of $S_{eff}$. After the values have been smoothed, $r^2$ values are interpolated onto a finer grid scale (see auxiliary material) using least-squares quadratic interpolation of the smoothed data.

[12] Using an F-test, each $r^2$ value is then converted to the likelihood that a given model produces output which matches the observations [*Forest et al.*, 2002]. Further details of this process are provided in the auxiliary material. The resulting $r^2$ values at a net aerosol forcing of −0.25 $Wm^{-2}$ indicate that the regions of the parameter space which are rejected by the surface diagnostic are dependent on which surface dataset is used (Figure 2). At this particular $F_{aer}$ level, model results are consistent with HadCRUT data over a much larger region of the parameter space. In particular, regions that are not inconsistent with the surface data at the 10-percent level are only present when HadCRUT data are used. Significant differences between the surface diagnostics are also evident at the −0.50 and −0.75 $Wm^{-2}$ $F_{aer}$ levels and tend to show larger acceptance regions for the HadCRUT2 and HadCRUT3 datasets and strong rejection of high $S_{eff}$ values for the GISTEMP datasets (see auxiliary material). Upper-air and ocean diagnostics were not changed from *Forest et al.* [2008].

[13] Upon interpolation between $F_{aer}$ levels and repeated application of Bayes' Theorem for each diagnostic, the joint probability distribution for the parameter space is derived. The expert prior on $S_{eff}$ used by *Forest et al.* [2008] has been applied along with uniform priors on $K_v$ and $F_{aer}$. The resulting marginal distributions for each parameter are presented in Figures 3a–3c for each surface dataset, along with the distributions derived by *Forest et al.* [2008]. From these distributions, it is clear that the dataset used for the surface diagnostic impacts the parameter distributions. For $S_{eff}$, the distributions derived from the GISTEMP datasets yield the lowest values with 5–95% confidence intervals of 1.3 to

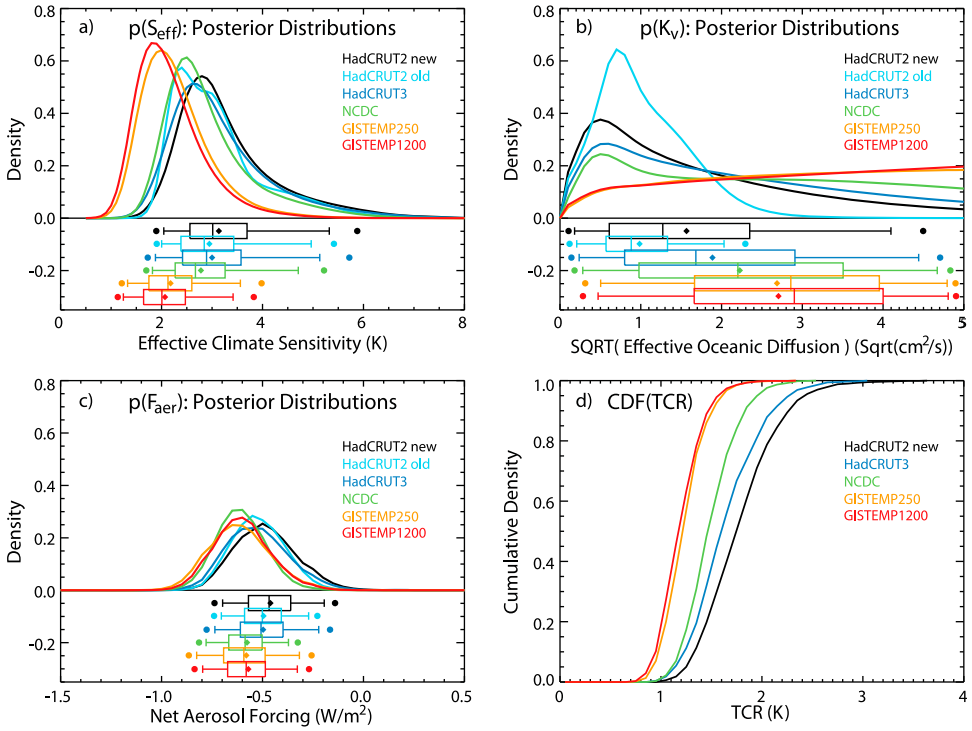


**Figure 3.** Marginal probability distribution functions and TCR cumulative distribution functions derived from HadCRUT2 (black), HadCRUT3 (blue), NCDC (green), GISTEMP 250 (orange), and GISTEMP 1200 (red) surface data. (a) $S_{eff}$, (b) $K_v$, and (c) $F_{aer}$ marginal distributions are compared to those presented by *Forest et al.* [2008] (cyan). Whisker plots indicate boundaries for the 2.5–97.5 (dots), 5–95 (vertical bar ends), 25–75 (box ends), and 50 (vertical bar in box) percentiles. Distribution means are represented by diamonds and modes are the peaks in the distribution. (d) TCR CDFs are derived from 1000 member Latin Hypercube samples from the joint parameter distributions.

3.6 K (GISTEMP 250) and 1.2 to 3.4 K (GISTEMP 1200). This can be traced to the failure of the surface diagnostic in constraining the lower bound of the distribution and can be attributed to the weaker warming trends previously discussed (i.e., model runs with higher $S_{eff}$ values yield warming which is too strong to be consistent with the GISTEMP datasets). The fifth percentiles fall outside of the lower bound of 2.0 K given in the IPCC AR4. The remaining datasets yield similar, yet still noticeably different results. Of these, HadCRUT datasets yield wider 5–95% confidence intervals of 2.0 to 5.3 K (HadCRUT2) and 1.9 to 5.1 K (HadCRUT3) than the NCDC distributions 1.8 to 4.7 K bounds. Each upper bound is greater than the upper bound of 4.5 K in the IPCC AR4.

[14] Based on the wide confidence intervals regardless of which surface dataset is used, $K_v$ is poorly constrained by the observations. With the exclusion of the GISTEMP datasets, the mode in the distribution is found for low values of ocean heat uptake. This results from the high $S_{eff}$ and high $K_v$ regions being rejected for positive values of net aerosol forcing. GISTEMP datasets demonstrate a long right tail for high values of $K_v$ and show no pronounced mode. A major difference between the distributions derived in this study and those from *Forest et al.* [2008] is that an estimate of natural variability has been included in the ocean heat content diagnostic. This is analogous to the treatment of the surface diagnostic and accounts for observational errors and natural sources of variability by combining the variability of both sources into a single estimate of the total variability. This estimation results in a decrease in the significance of the ocean heat content signal and leads to weaker constraints on $K_v$. As a result, broad distributions for $K_v$ are derived when the natural variability estimate is included.

[15] Weaker $F_{aer}$ values are estimated when using HadCRUT data, with 5–95% intervals of −0.19 to −0.70 $Wm^{-2}$ (HadCRUT2) and −0.22 to −0.74 $Wm^{-2}$ (HadCRUT3). The remaining datasets yield approximately 0.1 $Wm^{-2}$ stronger aerosol forcing, with 5–95% confidence intervals of −0.37 to −0.78 $Wm^{-2}$ (NCDC), −0.32 to −0.83 $Wm^{-2}$ (GISTEMP 250), and −0.33 to −0.80 $Wm^{-2}$ (GISTEMP 1200). The slightly weaker $F_{aer}$ values from the HadCRUT datasets can be attributed to the larger acceptance regions at the −0.25 $Wm^{-2}$ aerosol level seen in Figure 2. However, the overall range of $F_{aer}$ based off of the 5–95% confidence intervals, −0.19 to −0.83 $Wm^{-2}$, is smaller than errors in other model forcing terms.

[16] TCR distributions are derived from the parameter distributions. From each joint distribution, a 1000 member Latin Hypercube sample [*McKay et al.*, 1979] is estimated, whereby $S_{eff}$-$K_v$ pairs are drawn. Using a functional fit calibrated by prior runs of the model, the resulting TCR has been calculated for each pair [*Sokolov et al.*, 2003] and cumulative density functions are estimated (Figure 3d).

[17] We note that the lower bound on TCR values for the GISTEMP results are less than the lower bound of 1 K from the IPCC AR4. Ranges of 0.87 to 1.32 K (GISTEMP 250)

and 0.91 to 1.35 K (GISTEMP 1200) mark the 5–95% confidence intervals. These lower TCRs can be attributed to the low values of $S_{eff}$ and long right tails in the $K_v$ distributions derived using the GISTEMP datasets. With a lower $S_{eff}$, the equilibrium temperature change will be less for a given forcing. Furthermore, more efficient mixing of heat into the deep ocean acts to reduce surface temperatures as well. The HadCRUT2, HadCRUT3, and NCDC 5–95% intervals are bounded by 1.24 to 2.31 K, 1.13 to 2.41 K, and 1.10 to 1.96 K, respectively. All distributions fall within the range of TCR values given in the IPCC AR4. Given that the $K_v$ distributions are similar for these datasets, similar values are drawn in the Latin Hypercube sample and it follows that the $S_{eff}$ distributions should dominate the TCR distributions (Figure 3d). Given that the $F_{aer}$ distributions are nearly identical across all datasets, these results show that TCR follows shifts in the distributions of $S_{eff}$ and $K_v$ rather than those for $F_{aer}$.

[18] As a surrogate for future warming, TCR distributions measure the model response global mean temperature change for idealized forcing scenarios. Similar to climate sensitivity estimates, TCR has a profound impact on policy decisions regarding climate change adaptation and mitigation strategies. If the information from TCR distributions are properly included in policy decisions, effective strategies can be potentially improved.

## 5. Conclusions

[19] The results presented here show that climate model parameter constraints are sensitive to the surface dataset used to compare with model output. In general, the ranges of the effective climate sensitivity parameter distributions are comparable, but are shifted relative to each other depending on which surface dataset is used. The biggest shift in effective climate sensitivity distributions is observed when the GISTEMP datasets are used. Using the 95-percent confidence intervals and considering all datasets, climate sensitivity is found to be between 1.2 and 5.3 K. Regardless of the surface data used, effective ocean diffusivity is poorly constrained by the data. Anthropogenic aerosol forcing is found to be between −0.19 and −0.83 $Wm^{-2}$ when considering all datasets.

[20] TCR estimates are also sensitive to the choice of surface data. When all surface datasets are considered, transient warming is found to lie between 0.87 and 2.31 K. However, this range masks the differences that exist between the individual distributions. The TCR distribution derived from GISTEMP data is narrower and yields only minimal warming. In contrast, distributions derived from Hadley Centre datasets are wider and yield stronger warming. Given that both the parameter and TCR distributions differ when using different datasets, additional uncertainty is present in model calibration and climate projection studies. Future studies using these datasets must account for these differences to avoid overconfidence in predictions through mistreatment of the uncertainty.

[21] **Acknowledgments.** This project was supported in part by NOAA award NA09OAR4310176, DOE award DE-SC0004956, NSF award SES-0825915, and the MIT Joint Program on the Science and Policy of Global Change. We also thank the National Climatic Data Center, the Hadley Centre for Climate Prediction and Research, and the NASA Goddard Institute for Space Studies for producing publicly available surface data products. The authors thank B. Sanderson and K. Tanaka for helpful and constructive reviews.

[22] The Editor thanks the two anonymous reviewers for their assistance in evaluating this paper.

## References

Allen, M. R., P. A. Stott, R. Schnur, T. Delworth, and J. F. B. Mitchell (2000), Quantifying the uncertainty in forecasts of anthropogenic climate change, *Nature*, *407*, 617–620.

Andrews, D. G., and M. R. Allen (2008), Diagnosis of climate models in terms of transient climate response and feedback response time, *Atmos. Sci. Lett.*, *9*, 7–12.

Andronova, N. G., and M. E. Schlesinger (2001), Objective estimation of the probability density function for climate sensitivity, *J. Geophys. Res.*, *106*(D19), 22,605–22,612.

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

Forest, C. E., M. R. Allen, A. P. Sokolov, and P. H. Stone (2001), Constraining climate model properties using optimal fingerprint detection methods, *Clim. Dyn.*, *18*, 277–295.

Forest, C. E., 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.

Forest, C. E., P. H. Stone, and A. P. Sokolov (2008), Constraining climate model parameters from observed 20th century changes, *Tellus, Ser. A*, *60*(5), 911–920.

Hansen, J., R. Ruedy, M. Sato, and K. Lo (2010), Global surface temperature change, *Rev. Geophys.*, *48*, RG4004, doi:10.1029/2010RG000345.

Hegerl, G. C., et al. (2007), Understanding and attributing climate change, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon et al., pp. 663–746, Cambridge Univ. Press, Cambridge, U. K.

Jones, P., and A. Moberg (2003), Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001, *J. Clim.*, *16*, 206–223.

Knutti, R., and G. C. Hegerl (2008), The equilibrium sensitivity of the Earth's temperature to radiation changes, *Nat. Geosci.*, *1*, 735–743.

Knutti, R., and L. Tomassini (2008), Constraints on the transient climate response from observed global temperature and ocean heat uptake, *Geophys. Res. Lett.*, *35*, L09701, doi:10.1029/2007GL032904.

Knutti, R., T. F. Stocker, F. Joos, and G. Plattner (2003), Probabilistic climate change projections using neural networks, *Clim. Dyn.*, *21*, 257–272.

McInerney, D., and K. Keller (2008), Economically optimal risk reduction strategies in the face of uncertain climate thresholds, *Clim. Change*, *91*, 29–41.

McKay, M. D., R. J. Beckman, and W. J. Conover (1979), A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics*, *21*, 239–245.

Peterson, T. C., and R. S. Vose (1997), An overview of the Global Historical Climatology Network temperature database, *Bull. Am. Meteorol. Soc.*, *78*(12), 2837–2849.

Randall, D., et al. (2007), Climate models and their evaluation, in *Climate Change 2007, The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Integovernmental Panel on Climate Change*, edited by S. Solomon et al., pp. 589–662, Cambridge Univ. Press, Cambridge, U. K.

Sansó, B., and C. Forest (2009), Statistical calibration of climate system properties, *Appl. Stat.*, *58*, 485–503.

Smith, T. M., R. W. Reynolds, T. C. Peterson, and J. Lawrimore (2008), Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880–2006), *J. Clim.*, *21*, 2283–2296.

Sokolov, A., et al. (2005), The MIT integrated global system model (IGSM) version 2: Model description and baseline evaluation, *MIT JP Rep. 124*, Mass Inst. of Technol., Cambridge. [Available at http://web.mit.edu/globalchange/www/MITJPSPGC_Rpt124.pdf.]

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

Sokolov, A. P., C. E. Forest, and P. H. Stone (2003), Comparing oceanic heat uptake in AOGCM transient climate change experiments, *J. Clim.*, *16*, 1573–1582.

Stott, P. A., and C. E. Forest (2007), Ensemble climate predictions using climate models and observational constraints, *Philos. Trans. R. Soc. A*, *365*, 2029–2052.

Tanaka, K., T. Raddatz, B. C. O'Neill, and C. H. Reick (2009), Insufficient forcing uncertainty underestimates the risk of high climate sensitivity, *Geophys. Res. Lett.*, *36*, L16709, doi:10.1029/2009GL039642.

Tomassini, L., P. Reichert, R. Knutti, T. F. Stocker, and M. E. Borsuk (2007), Robust Bayesian uncertainty analysis of climate system properties using Markov Chain Monte Carlo estimates, *J. Clim.*, *20*, 1239–1254.

Urban, N. M., and K. Keller (2010), Probabilistic hindcasts and projections of the coupled climate, carbon cycle and Atlantic meridional overturning circulation system: A Bayesian fusion of century-scale observations with a simple model, *Tellus, Ser. A*, *62*(5), 737–750.

C. E. Forest, Department of Meteorology, Pennsylvania State University, 507 Walker Bldg., University Park, PA 16802, USA. (ceforest@psu.edu)

A. G. Libardoni, Department of Meteorology, Earth and Environmental Systems Institute, Pennsylvania State University, 408 Walker Bldg., University Park, PA 16802, USA. (agl127@psu.edu)

MIT JOINT PROGRAM ON THE SCIENCE AND POLICY OF GLOBAL CHANGE

**REPRINT SERIES Recent Issues**

Joint Program Reprints are available free of charge (limited quantities). To order: please use contact information on inside of front cover.

**2012-29** Proposed Vehicle Fuel Economy Standards in the United States for 2017 to 2025: Impacts on the Economy, Energy, and Greenhouse Gas Emissions, Karplus, V.J. and S. Paltsev, *Transportation Research Record: Journal of the Transportation Research Board*, 2289: 132–139 (2012)

**2012-30** The Role of Stocks and Shocks Concepts in the Debate Over Price versus Quantity, Parsons, J.E. and L. Taschini, *Environmental and Resource Economics*, doi: 10.1007/s10640-012-9614-y (2012)

**2012-31** Green growth and the efficient use of natural resources, Reilly, J.M., *Energy Economics*, 34(S1): S–S93 (2012)

**2012-32** The role of China in mitigating climate change, Paltsev, S., J. Morris, Y. Cai, V. Karplus and H.D. Jacoby, *Energy Economics*, 34(S3): S444–S450 (2012)

**2012-33** Valuing climate impacts in integrated assessment models: the MIT IGSM, John Reilly, Sergey Paltsev, Ken Strzepek, Noelle E. Selin, Yongxia Cai, Kyung-Min Nam, Erwan Monier, Stephanie Dutkiewicz, Jeffery Scott, Mort Webster and Andrei Sokolov, *Climatic Change*, online first, doi:10.1007/s10584-012-0635-x (2012)

**2012-34** From "Green Growth" to sound policies: An overview, Schmalensee, Richard, *Energy Economics*, 34: S–S6 (2012)

**2012-35** Shale gas production: potential versus actual greenhouse gas emissions, O'Sullivan, F. and S. Paltsev, *Environmental Research Letters*, 7(4): 044030 (2012)

**2013-1** The Impact of Climate Policy on US Aviation, Winchester, Niven, Christoph Wollersheim, Regina Clewlow, Nicholas C. Jost, Sergey Paltsev, John M. Reilly and Ian A. Waitz, *Journal of Transport Economics and Policy*, 47(1): 1–15 (2013)

**2013-2** Impact of anthropogenic absorbing aerosols on clouds and precipitation: A review of recent progresses, Wang, Chien, *Atmoshperic Research*, 122: 237–249 (2013)

**2013-3** Applying engineering and fleet detail to represent passenger vehicle transport in a computable general equilibrium model, Karplus, Valerie, Sergey Paltsev, Mustafa Babiker and John M. Reilly, *Economic Modelling,* 30: 295–305 (2013)

**2013-4** Should a vehicle fuel economy standard be combined with an economy-wide greenhouse gas emissions constraint? Implications for energy and climate policy in the United States, Karplus, Valerie, Sergey Paltsev, Mustafa Babiker and John M. Reilly, *Energy Economics*, 36: 322–333 (2013)

**2013-5** Climate impacts of a large-scale biofuels expansion, Hallgren, W., C.A. Schlosser, E. Monier, D. Kicklighter, A. Sokolov and J. Melillo, *Geophysical Research Letters*, 40(8): 1624–1630 (2013)

**2013-6** Non-nuclear, low-carbon, or both? The case of Taiwan, Chen, Y.-H.H., *Energy Economics*, 39: 53–65 (2013)

**2013-7** The Cost of Adapting to Climate Change in Ethiopia: Sector-Wise and Macro-Economic Estimates, Robinson, S., K. Strzepek and Raffaello Cervigni, *IFPRI ESSP WP 53* (2013)

**2013-8** Historical and Idealized climate model experiments: an intercomparison of Earth system models of intermediate complexity, Eby, M., A.J. Weaver, K. Alexander, K. Zickfield, A. Abe-Ouchi, A.A. Cimatoribus, E. Crespin, S.S. Drijfhout, N.R. Edwards, A.V. Eliseev, G. Feulner, T. Fichefet, C.E. Forest, H. Goosse, P.B. Holden, F. Joos, M. Kawamiya, D. Kicklighter, H. Kiernert, M. Matsumoto, I.I. Mokov, E. Monier, S.M. Olsen, J.O.P. Pedersen, M. Perrette, G. Phillpon-Berthier, A. Ridgwell, A Schlosser, T. Schneider von Deimling, G. Shaffer, R.S. Smith, R. Spahni, A.P. Sokolov, M. Steinacher, K. Tachiiri, K. Tokos, M. Yoshimori, N Zeng and F. Zhao, *Clim. Past*, 9:1111–1140 (2013)

**2013-9** Correction to "Sensitivity of distributions of climate system properties to the surface temperature data set", and Sensitivity of distributions of climate system properties to the surface temperature data set, Libardoni, A.G. and C.E. Forest, *Geophysical Research Letters*, 40(10): 2309–2311 (2013), and 38(22): 1–6 (2011)

**For a complete list of titles see:**
**http://globalchange.mit.edu/research/publications/reprints**

**MIT Joint Program on**
**The Science and Policy of Global Change**
Massachusetts Institute of Technology
77 Massachusetts Avenue, E19-411
Cambridge, MA 02139
USA