Discriminating between climate observations in terms of their ability to improve an ensemble of climate predictions

Yi Huang[1], Stephen Leroy, and Richard M. Goody[2]

School of Engineering and Applied Science, Harvard University, Cambridge, MA 02138

Contributed by Richard M. Goody, May 10, 2011 (sent for review March 14, 2011)

In view of the cost and complexity of climate-observing systems, it is a matter of concern to know which measurements, by satellite or in situ, can best improve the accuracy and precision of long-term ensembles of climate projections. We follow a statistical procedure to evaluate the relative capabilities of a wide variety of observable data types for improving the accuracy and precision of an ensemble of Intergovernmental Panel on Climate Change (IPCC) models. Thirty-two data types are evaluated for their potential for improving a 50-y surface air temperature trend prediction with data from earlier periods, with an emphasis on 20 y. Data types can be ordered in terms of their ability to increase the precision of a forecast. Results show that important conclusions can follow from this ordering. The small size of the IPCC model ensemble (20 members) creates uncertainties in these conclusions, which need to be substantiated with the larger ensembles expected in the future. But the larger issue of whether the methodology can provide useful answers is demonstrated.

climate monitoring | climate model | remote sensing | climate change

Satellite observations are an increasingly important source of climate observations. They can give continuous global coverage with the same instruments, and it is now possible to design simple instruments that can be referred confidently to international standards, assuring the value of the observations in perpetuity. An example of this approach is National Aeronautics and Space Administration's Climate Absolute Radiance and Refractivity Observatory (CLARREO) mission, which has recently been suspended indefinitely due to national fiscal constraints. CLARREO, as last configured, contained a thermal infrared interferometer, a global positioning system (GPS) radio occultation instrument, and a shortwave spectrometer to measure reflected solar radiation. It becomes a matter of concern to know whether these are the best measurements for improving long-term climate projections, and if not, which measurements are.

In this paper, we ask how far a given observable data type is capable of improving an ensemble of climate trend predictions, directly. This is a standard problem in the statistical literature, which we have adapted to the requirements of this paper. The widely used ensemble of climate models employed by the Intergovernmental Panel on Climate Change (1) Fourth Assessment Report (IPCC-AR4) provides an example of an ensemble, and we anticipate that ensembles of projections will be a common feature of future climate studies. An ensemble of projections can be characterized, at a minimum, by a mean and a standard deviation, which may be identified with the most likely projection and its precision, respectively. These important parameters can be modified by taking account of measured data not previously incorporated into the models. The data need not be the same as the predicted quantities. For example, the prediction may be of globally averaged surface temperature trends. The measured data could be trends of entirely different quantities measured at different times from the prediction. These data will generally result in improvements in the accuracy and precision of the ensemble. The greater the improvement, the more effective is the measurement, at least in this context, providing the kind of direct, quantitative assessment that has been lacking in the past. This context is, however, relative. A different long-term prediction will probably lead to a different assessment of observing systems. The aim of this investigation is not to improve our knowledge of the climate, a task that requires a different approach. The program that we adopt to obtain an assessment is unique for climate prediction but familiar to the more mature field of weather prediction. Observing-system simulation experiments for weather observations were first proposed by Smagorinsky (2) and were more recently reviewed by Arnold and Dey (3).

Because the ensemble that we use has only 20 members, the significance of the present assessment is not high. Nevertheless, the larger issue as to whether appropriate methodologies exist is demonstrated.

In *Method*, we outline the theoretical basis for this paper. The methods are well-established in the statistical literature, but are not familiar to the climate community. In *Calculations*, we apply these results to the IPCC ensemble of 50-y projections, using a *perfect model* approach to provide "data" at 10-y intervals before that time. We shall then use 20-y data to modify a 50-y prediction. We choose a 20-y period for data because although longer periods of data do more to suppress the influence of internal variability, they may be too long to be of value. The data will include 32 candidate climate variables, and specifically those that can be observed by the CLARREO instrumentation. In *Discussions and Conclusions*, we discuss these results and present some conclusions.

Method

Our aim is to improve an ensemble of climate trend predictions, using data. The solution to this problem (with small adjustments) can be found in both the literature of Bayesian inference (see ref. 4) and the literature of classical (frequentist) statistics (for example, see ref. 5, section 5-5). We have chosen to use the language of Bayesian inference because it is convenient and because it introduces the evidence function, which is needed for later discussion.

We formulate a statistical model wherein an ensemble of x, y pairs is produced by an ensemble of climate models $\mathscr{E}$, the x variable referring to quantities that can be observed and the y variable referring to quantities to be predicted. Without modification by collected data, we estimate a prior probability density function (PDF) in these x, y pairs as $P(x,y|\mathscr{E})$. Available climate

Author contributions: R.M.G. designed research; Y.H. performed research; and Y.H., S.L., and R.M.G. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed at: 12 Oxford Street, Link 284, Cambridge, MA 02138. E-mail: yihuang@huarp.harvard.edu.

[2]To whom correspondence should be addressed at: 101 Cumloden Drive, Falmouth, MA 02540. E-mail: goody@huarp.harvard.edu.

model ensembles are limited in the number of members, so we must devise a prior $P(x,y|\mathscr{E})$ by assuming a continuum of models, thus making the functional form of the prior continuous. For the sake of simplicity, we assume the functional forms of all PDFs to be Gaussian, there being little evidence to suggest something other than Gaussian PDFs, given the small size of the ensemble (20 members). We apply Bayes's theorem (see ref. 6) using data d, and the likelihood function $P(d|x)$ dependent only on x by definition:

$$P(x,y|d,\mathscr{E}) = P(d|x)P(x,y|\mathscr{E})/P(d|\mathscr{E}). \quad [1]$$

The posterior $P(x,y|d,\mathscr{E})$ is the ensemble prediction x, y modified by data; the likelihood function $P(d|x)$ is the likelihood of the data given a predicted observation; the prior $P(x,y|\mathscr{E})$ is the PDF of the ensemble before data are considered; and the evidence function $P(d|\mathscr{E})$ is the evidence that the data are valid in light of the ensemble, which also serves as a normalization constant on the posterior distribution. In order to arrive at the posterior projection for y, we treat the x in the posterior as a nuisance parameter and integrate over it. Thus,

$$P(y|d,\mathscr{E}) = \int dx P(x,y|d,\mathscr{E}). \quad [2]$$

The solution to Eqs. **1** and **2** is a slight modification to the solution given by Gelman et al. (see ref. 4, Eqs. A.1 and A.2 on p. 579 and §2.6 on p. 46). If $\mathscr{N}$ is the normal distribution, the prior in x, y is

$$P(x,y|\mathscr{E}) \sim \mathscr{N}([\mu_x,\mu_y],\mathbf{\Sigma}), \quad [3]$$

where the uncertainty covariance matrix $\mathbf{\Sigma}$ has submatrices $\mathbf{\Sigma}_{xx}$, $\mathbf{\Sigma}_{yy}$, $\mathbf{\Sigma}_{xy}$, and $\mathbf{\Sigma}_{yx} = \mathbf{\Sigma}_{xy}^T$ corresponding to the x and y subspaces, and μ_x and μ_y are the means of the x and y ensembles.

Then the posterior in y is

$$P(y|d,\mathscr{E}) \sim \mathscr{N}(\mu_{y|d},\mathbf{\Sigma}_{y|d}), \quad [4]$$

where the updated most probable prediction $\mu_{y|d}$ and its uncertainty covariance $\mathbf{\Sigma}_{\mathbf{y|d}}$ are

$$\mu_{y|d} = \mu_y + \mathbf{\Sigma}_{yx}(\mathbf{\Sigma}_{xx} + \mathbf{\Sigma}_d)^{-1}(d - \mu_x), \quad [5]$$

$$\mathbf{\Sigma}_{y|d} = \mathbf{\Sigma}_{yy} - \mathbf{\Sigma}_{yx}(\mathbf{\Sigma}_{xx} + \mathbf{\Sigma}_d)^{-1}\mathbf{\Sigma}_{xy}. \quad [6]$$

$\mathbf{\Sigma_d}$ is the variance of the data d.

When x and y are reduced to one dimension each so that $\mathbf{\Sigma}_{yy} \to \sigma_y^2$, $\mathbf{\Sigma}_{xx} \to \sigma_x^2$, $\mathbf{\Sigma}_{xy} \to \rho\sigma_x\sigma_y$, and $\mathbf{\Sigma}_d \to \sigma_d^2$, the following equations result:

$$\mu_{y|d} = \mu_y + \frac{\rho\sigma_x\sigma_y}{\sigma_x^2 + \sigma_d^2}(d - \mu_x), \quad [7]$$

$$\sigma_{y|d}^2 = \sigma_y^2\left(1 - \frac{\rho^2\sigma_x^2}{\sigma_x^2 + \sigma_d^2}\right). \quad [8]$$

ρ is the correlation coefficient relating measurement to prediction.

For both the posterior prediction $\mu_{y|d}$ and posterior uncertainty $\sigma_{y|d}$, the correlation coefficient ρ plays a crucial role. The update of the posterior mean prediction from the prior mean prediction is linear in the correlation coefficient ρ, and the reduction of the prediction uncertainty is always greater for larger values of $|\rho|$.

The uncertainty described by $\mathbf{\Sigma}_{xx}$ includes both the uncertainty in climate modeling due to model errors, and uncertainty due to natural variability. For the purposes of this paper, these two uncertainties need not be separated. $\mathbf{\Sigma_d}$ describes the uncertainty in the measurement.

Calculations

We apply the results of *Method* to the models of the World Climate Research Programme's (WCRP's) Coupled Model Intercomparison Project phase 3 (CMIP3) multimodel dataset, the ensemble that is used in the projections of IPCC-AR4. We use the results of the forcing scenario A1B, for which carbon dioxide concentrations stabilize at about twice year 2,000 levels after 70 y. All calculations of x and y are trends produced by A1B forcing, and all are annual and global averages of trends except where otherwise stated. Trends are obtained by linear regression on a time series of data. The means of x and y produced by CMIP3, one run per model, are μ_x and μ_y, and their uncertainties across the ensemble are σ_x and σ_y. For d, the uncertainty, σ_d, is obtainable from the trend calculation. This procedure captures only natural variability and does not include possible measurement errors (7).

There are 21 members of this ensemble but one model (ncar_pcm1) is used to represent the data (a *perfect model* test), so 20 members remain. Calculations were made for surface air temperature trends over 50 y (y, the 50-y prediction) using data types (d) and data simulations (x) evaluated at 10-y intervals up to 50 y (but with most attention directed to 20 y). Therefore, the importance of a data type is based on the increase in quality of a 50-y prediction given information on that data type for the first 20 y. The greater the increase in quality, the more valuable the data type.

Calculations were made with 32 different data types, including surface air temperature, column integrated cloud water and ice, total cloud amount, precipitation, precipitable water, surface downwelling and upwelling long-wave and short-wave radiation, and atmospheric temperature, relative humidity, specific humidity, and geopotential height at 500, 200, and 50 *hPa* levels. Also included are upwelling monochromatic nadir radiances measured from space. These data are not in the CMIP3 archive but were calculated off-line with the radiative transfer code MODTRAN at 2 cm^{-1} resolution (8, 9) using archived data on physical variables. These were clear-sky calculations, which do not account for the presence of clouds because the CMIP3 archive does not contain the needed cloud information. Fig. 1 shows the correlation coefficient between trends of radiances in an outgoing radiance

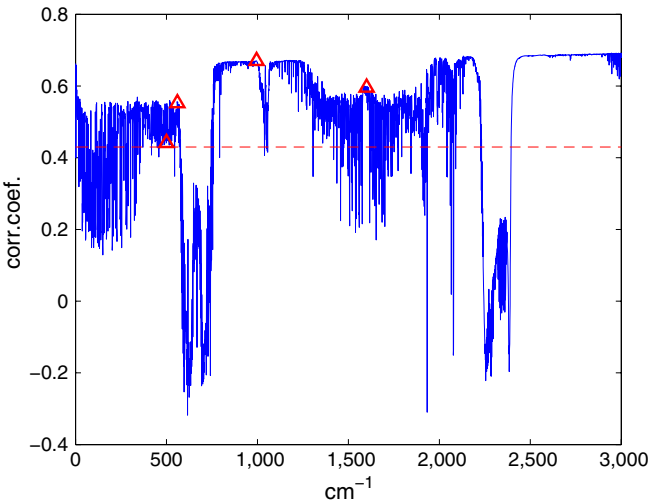

Fig. 1. Correlation coefficients between 20-y trends in spectral radiances and the 50-y trends in global surface temperature. Values higher than the dashed line are significant at 95% confidence level. The radiances used in this paper are indicated.

spectrum at 20 y and trends of global surface temperatures at 50 y. The four radiances chosen in this study are indicated in Fig. 1. They are 1,600 cm^{-1}, which has a strong absorption by water vapor; 995 cm^{-1}, which is an infrared window to the surface; 500 cm^{-1}, which sounds lower atmospheric water vapor; and 560 cm^{-1}, which is influenced slightly by carbon dioxide concentrations.

Of the 32 data types, both in situ and satellite based, 14 were selected for further analysis because their correlation coefficients relating observation to prediction were significantly different from zero at the 95% confidence level ($|\rho| > 0.43$). In addition, we selected three combinations of data types. Fig. 2 shows a time series of data at 10-y intervals for surface temperature trends and for one multiple dataset. These time series have the expected features that, as the correlation increases, the accuracy converges on the truth, and the precision tends to that of the data.

In Tables 1 and 2, results are shown for stand-alone data types (cf Eqs. **7** and 8). Table 1 is for satellite data types and Table 2 is for in situ data types. The reason for the separation of in situ from satellite data types is the very large practical differences in the methods required to obtain them, and specifically to investigate CLARREO instrumentation. All calculations in Tables 1 and 2 were for 20-y global average trends in the measured quantity x and for 50-y global average trends in surface air temperature. The diagnostic quantities are improvement in accuracy of the prediction ($\Delta\hat{\mu}$) and improvement in precision ($\Delta\hat{\sigma}$):

$$\Delta\hat{\mu} = (\mu_y - \mu_{y|d})/(\mu_y - y_t), \quad [9]$$

$$\Delta\hat{\sigma} = (\sigma_y - \sigma_{y|d})/\sigma_y, \quad [10]$$

with y_t the truth of the predicted quantity y, which can be known only in a perfect model test.

Discussion and Conclusions

Findings. There is no single way to evaluate the relative merit of data types when making a specific prediction, but comparing abilities to improve accuracy and precision of prediction are two obvious candidates. But $\Delta\hat{\mu}$ and $\Delta\hat{\sigma}$ do not follow the same order in Tables 1 and 2, and a choice must be made between them.

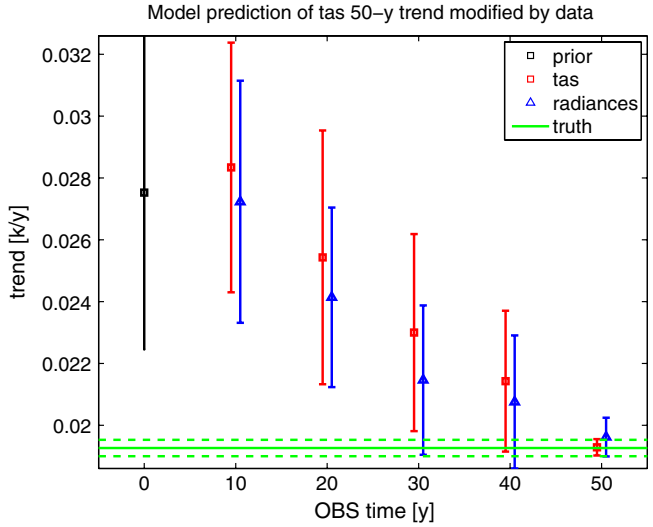

Fig. 2. Model projections of surface temperature trends at 50 y, modified by data at 10, 20, 30, 40, and 50 y. Two data types are shown: surface temperature trends (tas, in situ data) and all satellite radiance trends together (see Table 1). The green line represents d_y, the data for y at 50 y, the "truth." Not all data show such regular behaviors.

Table 1. Perfect model tests for data measurable from space

Data type	ρ	σ_ρ	$\Delta\hat{\sigma}$	$\Delta\hat{\mu}$
Total outgoing long-wave radiation	0.67	0.13	0.27	0.75
Radiance at 995 cm^{-1}	0.67	0.13	0.23	0.44
Dry pressure at 5.5 km	0.64	0.14	0.22	0.42
Radiance at 1,600 cm^{-1}	0.60	0.16	0.18	0.34
Radiance at 560 cm^{-1}	0.55	0.17	0.15	0.35
Reflected solar radiation	−0.47	0.19	0.13	0.32
Radiance at 500 cm^{-1}	0.44	0.20	0.09	0.30

A 50-y projection is modified by 20-y data. ρ is the correlation coefficient between the 20-y trend of the variable and the 50-y trend in surface temperature. σ_ρ is the standard deviation of ρ, $\Delta\hat{\mu} = (\mu_y - \mu_{y|d})/(\mu_y - y_t)$, $\Delta\hat{\sigma} = (\sigma_y - \sigma_{y|d})/\sigma_y$. Listed in order of decreasing $|\rho|$.

The dry pressure at 5.5 km is a surrogate for a GPS occultation measurement.

Priority according to ability to improve accuracy proves to be far more problematic than according to ability to improve precision. The evaluation of $\Delta\hat{\mu}$ depends strongly on the choice of the perfect model, and on the data type. It is also subject to the natural variability of d, which is particularly important if the 20-y forecast is accurate (when $[d - \mu_x]$ is small). On the other hand, $\Delta\hat{\sigma}$ depends only very weakly on the perfect model chosen and contains no terms that depend strongly on the natural variability of d. Consequently, we use ability to improve the precision of projection as the basis for setting priorities of data types. For most purposes, this criterion means that the correlation coefficient can be used as the ordering parameter.

Nevertheless, we include $\Delta\hat{\mu}$ in Tables 1, 2, and 3 to illustrate the magnitude of the improvement in accuracy that can result. For example, in Table 3, the first entry shows that satellite measurements may be capable of improving accuracy by 81%.

Taken at face value (but see, however, *Uncertainties*), the single data types listed in Tables 1 and 2 suggest some interesting conclusions. Comparing data types for satellite and in situ measurements, there are three satellite measurements [total outgoing long-wave radiation trend, radiance trend at 995 cm^{-1}, and dry pressure trend at 5.5 km, the occultation surrogate (10)] that appear to rank above or equal to the best in situ measurement (geopotential height trend at 500 hPa). This ranking could be taken to indicate that the CLARREO instrumentation is well chosen.

Table 3 suggests that, if the radiances could be chosen in an optimal manner, the thermal interferometer could, alone, provide a powerful constraint on ensemble projections. (We note that total outgoing thermal radiation can also be derived from interferometric measurements even though it is not measured directly.) The results are improved if radio occultation data are added.

Among in situ measurements, it appears that surface temperature trend may not necessarily be the best variable to use to control projections of surface temperature trends; there is a suggestion that the trend of geopotential height at 500 hPa may be equally useful.

Table 2. As for Table 1 but for data measurable in situ

Data type	ρ	σ_ρ	$\Delta\hat{\sigma}$	$\Delta\hat{\mu}$
Geopotential height at 500 hPa	0.64	0.14	0.22	0.44
Surface air temperature	0.61	0.15	0.21	0.27
Upwelling long-wave radiation at surface	0.61	0.15	0.21	−0.47
Precipitable water	0.61	0.15	0.18	0.42
Geopotential height at 200 hPa	0.56	0.17	0.17	0.39
Air temperature at 500 hPa	0.53	0.17	0.15	0.44
Specific humidity at 500 hPa	0.45	0.19	0.10	0.43

See Table 1 for definitions of symbols.

Table 3. Three multivariable calculations

Data	$\Delta\hat{\sigma}$	$\Delta\hat{\mu}$
All satellite quantities in Table 2	0.53	0.81
Spectral radiances only in Table 2	0.44	0.42
All in situ quantities in Table 1	0.35	0.57

A 50-y prediction is modified by 20-y data. See Table 1 for definition of symbols.

Uncertainties. There are uncertainties related to the ensembles. An ensemble may not span enough physical uncertainty to include the actual observations. If it does not, i.e., if d is outside the range of $P(x)$ (and there are members of the climate community who believe this to be true for the IPCC ensemble; see, for example, ref. 11), the application of Eqs. **1**–**8** in *Method* may decrease the accuracy of the solution while the precision appears to increase—a nonsensical result. Ultimately, once data are collected, ensembles can be rated according to the Bayesian evidence function $P(d|\mathscr{E})$ (see Eq. **1**), the probability that the data can exist given the ensemble. Small probabilities must be rejected, and the largest retained for prediction purposes.

The small size of the ensemble (20 members) introduces uncertainties into all calculated statistical quantities. The correlation coefficients for all variables that we consider are statistically significant at the >95% confidence level; but the absolute correlation coefficients $|\rho|$ for these observables do not necessarily differ significantly. In order to estimate significance of differences we have used Fisher's approximation to evaluate standard deviations for correlation coefficients (see ref. 12, §35), as shown in Tables 1 and 2. If we compare the largest and smallest correlation coefficients in Table 1, the difference is 0.23, and the combined standard deviations are 0.24, a result of small significance.

We have assumed Gaussian PDFs throughout in order to simplify the analysis. A Kolmogorov test validates our assumption, most likely only because the ensemble is as small as it is. If larger ensembles, suited to our purposes, become available in the future, non-Gaussian PDFs may be required. But we would not expect relative priorities of data types to depend critically on whether or not Gaussian statistics are assumed.

We have used clear-sky calculations of radiances because CMIP3 did not distribute the information necessary to simulate cloudy radiances off-line. The priorities of some data types may change in the reality of Earth's half-cloudy atmosphere. When the models give the data needed to calculate cloudy radiances, this limitation can be removed.

Finally, our approach does nothing to eliminate the uncertainties in forcing in the IPCC models. But because of the differential nature of our calculations this uncertainty should not have overriding importance.

Use of New Data. When new observed climate data become available, the question will arise as to how they may best be used. It is possible to make direct use of new data to increase both the precision and accuracy of an ensemble prediction through the approach of this paper. Table 3 suggests that the satellite instruments of CLARREO could lead to a 53% increase in the precision of a calculated ensemble and an 81% increase in its accuracy, for the specific procedure that we have used.

Other methods for modifying an ensemble climate prediction using data are subjects of active research. A method close to the one we have used is the "reliability ensemble average" (REA) method (see ref. 13), in which the posterior accuracy and precision are calculated by assigning a weight or "reliability" to each model projection. The reliabilities are estimated from two factors, the "bias" of a model or $(x - d)$, and the "convergence" of a model or $(y - \mu_y)$. Smith et al. (14) point out that the method has "several seemingly *ad hoc* features." Tebaldi et al. (15) and Smith et al. (14) claim to have removed these ad hoc features (with some changes in the definition of the reliability) using Bayesian inference, but they have done so without explicitly giving precedence to empirical data in the Bayesian likelihood function.

Future Work. The present procedure constitutes only a first attempt at the problem of observing-system selection, and improvements are possible. One area concerns spectral radiances. We have chosen four radiances to represent the spectrum. This choice might not be far from optimal for the restricted region of the spectrum used (see Fig. 1). But if the size of the ensemble could be increased, so that the two strong CO_2 bands become available, there is much more useful information. The use of spectral empirical orthogonal functions (EOFs) suggests itself as a way to handle large numbers of spectral radiances; then the variables become projections of the spectra on the first few EOFs. The variations between annually averaged spectra are a possible basis for calculating the EOFs. A preliminary attempt to conduct the EOF analysis was made for this paper, but given the limited spectral region available, the results were inconclusive.

Additional improvements for our technique include the following: The ensemble can be increased in size by using a perturbed-physics ensemble (see ref. 16). A larger ensemble would not only allow more data types to be used in the analysis, but it would decrease the uncertainties of difference between entries in Tables 1, 2, and 3 ($\sigma_\rho \propto \frac{1}{\sqrt{n-3}}$) and enable more confident choices to be made. Dealing with the cloudy/clear-sky problem has already been mentioned. Elimination of the Gaussian PDF assumption is possible, but will require numerical solutions. It may also be possible to use some observed time series of data in place of the perfect model assumption. The Earth Radiation Budget Experiment (ERBE) (17, 18) and Clouds and Earth's Radiation Energy System (CERES) (19) instruments have been measuring the total outgoing thermal radiation for many years. Nine years of atmospheric infrared sounder (AIRS) thermal spectra (20) will soon be available. New and independent in situ data will always be available.

Conclusions The differences between data types, as shown in Tables 1 and 2 are, at best, of marginal significance and could not be used for making practical choices between instruments. But the validity of the methodology is not affected by the size of the ensemble or any of the other factors discussed in *Future Work*. Consequently, this paper demonstrates the important proposition that objective methods can be used to constrain the accuracy and precision of ensembles of climate models using measured data.

The paper also demonstrates that the significance of data choices can be readily improved with the use of larger ensembles and that other improvements are also possible, opening up a major area for further research.

It is not the purpose of this paper to discuss the evidence in favor of the IPCC ensemble of predictions, or possible improvements to it to which new data might contribute. But our results demonstrate that the assessment or improvement of the IPCC ensemble is possible, with the caveat that the Bayesian evidence function $P(d|\mathscr{E})$ must be evaluated to assess whether the ensemble is capable of explaining the data.

ACKNOWLEDGMENTS. We thank Professor Art Dempster for helpful discussions of statistical issues and Professor Guido Visconti for a valuable review. We also thank the two PNAS reviewers for very constructive comments. We acknowledge the modeling groups, the Program for Climate Model Diagnosis and Intercomparison and the WCRP's Working Group on Coupled Modelling, for their roles in making available the WCRP CMIP3 multimodel dataset. Support of this dataset is provided by the Office of Science, US Department of Energy. This work was supported in part by National Science Foundation Grant ATM-0755099 and by National Aeronautics and Space Administration Grant NNX11AE74G.

1. Solomon S, et al., ed. (2007) *Climate Change 2007: The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ Press, Cambridge, UK) p 996.
2. Smagorinsky J, Minakoda Kl, Strickler RF (1970) The relative importance of variables in initial conditions for dynamical weather prediction. *Tellus* 22:141–157.
3. Arnold CP, Dey CH (1986) Observing-systems simulation experiments: Past, present and future. *Bull Am Meteoro Soc* 67:687–695.
4. Gelman A, Carlin J, Stern H, Rubin D (2003) *Bayesian Data Analysis* (Chapman & Hall, Boca Raton, FL), 2nd Ed, p 668.
5. Liebelt PB (1967) *An Introduction to Optimal Estimation* (Addison-Wesley, Reading, MA) p 273.
6. Sivia D (2007) *Data Analysis: A Bayesian Tutorial* (Oxford Univ Press, Oxford, UK) p 189.
7. Leroy S, Anderson J, Ohring G (2008) Climate signal detection times and constraints on climate benchmark accuracy requirements. *J Climate* 21:841–846.
8. Bernstein L, et al. (1996) Very narrow band model calculations of atmospheric fluxes and cooling rate. *J Atmos Sci* 53:2887–2904.
9. Berk A, et al. (1998) MODTRAN cloud and multiple scattering upgrades with application to AVIRIS. *Remote Sens Environment* 65:367–375.
10. Leroy S, Anderson J, Dykema J (2006) Testing climate models using GPS radio occultation: A sensitivity analysis. *J Geophys Res* 111:D17105, 10.1029/2005JD006145.
11. Stainforth DA, et al. (2005) Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature* 433:403–406.
12. Fisher R (1973) *Statistical Methods for Research Workers* (Hafner Publishing Co, New York), 14th Ed, p 353.
13. Giorgi F, Mearns L (2002) Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the "reliability ensemble averaging" (REA) method. *J Climate* 15:1141–1158.
14. Smith R, Tebaldi C, Nychka D, Mearns L (2009) Bayesian modeling of uncertainty in ensembles of climate models. *J Am Stat Assoc* 104:97–116.
15. Tebaldi C, Smith R, Nychka D, Mearns L (2005) Quantifying uncertainty in projections of regional climate change: A Bayesian approach to the analysis of multimodel ensembles. *J Climate* 18:1524–1540.
16. Allen M (1999) Do-it-yourself climate prediction. *Nature* 401:642.
17. Ramanathan V, et al. (1989) Cloud-radiative forcing and climate results from the Earth Radiation Budget Experiment. *Science* 243:57–63.
18. Wong T, et al. (2006) Reexamination of the observed decadal variability of the Earth radiation budget using altitude-corrected ERBE/ERBS nonscanner WFOV data. *J Climate* 19:4028–4040.
19. Wielicki B, et al. (1996) Clouds and the Earth's Radiant Energy System (CERES): An earth observing system experiment. *Bull Am Meteorol Soc* 77:853–868.
20. Aumann H, et al. (2003) AIRS/AMSU/HSB on the Aqua mission: Design, science objectives, data products, and processing systems. *IEEE Trans Geosci Remote Sensing* 41:253–264.