Climate model fidelity and projections of climate change

J. Shukla,[1,2] T. DelSole,[1,2] M. Fennessy,[3] J. Kinter,[1,2] and D. Paolino[3]

Received 21 December 2005; revised 6 February 2006; accepted 23 February 2006; published 1 April 2006.

[1] Relative entropy, which is a measure of the difference between two probability distributions, has been calculated for the simulations of the climate of the 20th century from 13 climate models and the observed surface air temperature during the past 100 years. This quantity is used as a measure of model fidelity: a small value of relative entropy indicates that a given model's distribution is close to the observed. It is found that there is an inverse relationship between relative entropy and the sensitivity of the model to doubling of the concentration of CO_2. The models that have lower values of relative entropy, hence have higher fidelity in simulating the present climate, produce higher values of global warming for a doubling of CO_2. This suggests that the projected global warming due to increasing CO_2 is likely to be closer to the highest projected estimates among the current generation of climate models. **Citation:** Shukla, J., T. DelSole, M. Fennessy, J. Kinter, and D. Paolino (2006), Climate model fidelity and projections of climate change, *Geophys. Res. Lett.*, *33*, L07702, doi:10.1029/2005GL025579.

[2] The climate modeling community of the world has made major advances over the past 30 years in building models and estimating the nature of changes in the Earth's climate due to increases in CO_2. It has been found, however, that different models give different estimates of the increase in global temperature for a given increase in the concentration of CO_2 [*Houghton et al.*, 2001]. The primary reason for differences in the estimates of the influence of CO_2 has been attributed to differences in the parameterization of cloudiness and cloud-radiation interactions [*Houghton et al.*, 2001]. In the absence of any other information, estimates from equally plausible climate models are usually given equal weight. Forty years ago, when weather prediction models were used to estimate the predictability of weather, they produced very different estimates of the predictability [*Charney et al.*, 1966]; however, the weather prediction models of today have converged to a consistent value for the upper limit of weather predictability [*Simmons and Hollingsworth*, 2002]. Likewise, when the atmospheric general circulation models of 10 years ago were used for estimating the dynamical seasonal predictability, different models gave different estimates of the influence of a given global sea surface temperature anomaly (see the special 2000 issue of *Quarterly Journal of the Royal Meteorological Society, 126*, 1991–2350, on dynamical seasonal prediction). It is only recently that most atmospheric general circulation models produce comparable estimates of the atmospheric response to a given sea surface temperature anomaly. The simulations of the annual cycle in the tropical Pacific and El Niño and the Southern Oscillation (ENSO) were highly variable among different models about 10 years ago, and only recently have most of the coupled ocean-atmosphere models shown some similarity in their prediction of large ENSO events. These results suggest that improvements in the models' ability to simulate the observed variability leads consistently to improvements in the skill of forecasts, especially for daily weather and seasonal time scales.

[3] In this paper, we consider a common extension of these results and investigate whether or not the projection of the future climate due to the increase in CO_2 concentration made by a given climate model depends on that model's ability to simulate the annual cycle and the interannual variability of the present climate. In order to examine such a relationship, we need a metric to define models' fidelity in simulating the present climate. Typical measures such as root mean square error or anomaly pattern correlation are inappropriate for this purpose because they measure the "closeness" of two states (unless one is comparing mean states). Instead, we need to measure the closeness of two *distributions*, in particular, the observed and simulated distributions. A "perfect" model under this measure is one that correctly simulates the annual cycle and the statistics of intraseasonal variability, not necessarily the precise field at a given time. For this purpose, we have chosen a quantity known as *relative entropy* [*Kullback*, 1959; *Cover and Thomas*, 1991; *Kleeman*, 2002; *DelSole*, 2004; *Tippett et al.* 2004] (also see Appendix A). Some basic properties of relative entropy include: it is nonnegative, it vanishes if and only if the two distributions are identical, and it is invariant with respect to invertible non-linear transformations, and hence is independent of the basis set in which the variables are represented.

[4] We have computed the relative entropy between simulations of the climate of the 20th century and the observations of seasonal mean surface air temperature for the 100-year period 1899–1998. In this study, a season corresponds to one of a sequence of non-overlapping three-month periods starting from January–February–March. The annual cycle is defined at each grid point as the 100-year average value for each season minus the 100-year annual mean value; hence, the annual cycle is specified by four numbers at each grid point. The seasonal anomaly is defined at each grid point as the seasonal mean value minus the appropriate annual cycle value. To calculate relative entropy, all variables were assumed to be joint normally distributed; this assumption is reasonable in view of the Central Limit Theorem and the fact that the seasonal mean

[1]Climate Dynamics Program, School of Computational Sciences, George Mason University, Fairfax, Virginia, USA.
[2]Also at Center for Ocean-Land-Atmosphere Studies, Calverton, Maryland, USA.
[3]Center for Ocean-Land-Atmosphere Studies, Calverton, Maryland, USA.

0094-8276/06/2005GL025579$05.00

Table 1. Relative Entropy, Model Sensitivity, Cyclostationary Relative Entropy and Mean Square Difference for 13 IPCC AR4 Models[a]

Model	Relative Entropy	Sensitivity, °K (Land Only)	Cyclostationary Relative Entropy	Mean Square Difference
1	5.62	5.21	5.49	5.68
2	5.73	4.41	1.69	3.54
3	6.99	3.34	4.89	7.40
4	8.67	3.34	2.13	3.50
5	9.47	3.55	4.24	5.62
6	9.70	4.20	4.63	4.67
7	10.85	2.78	4.32	4.36
8	10.98	4.10	7.77	9.79
9	13.30	4.05	4.61	8.52
10	15.97	2.76	5.81	7.24
11	17.40	2.78	7.72	10.41
12	17.96	2.89	6.45	10.03
13	19.57	2.51	9.26	9.03

[a]Relative entropy, cyclostationary relative entropy and mean square difference are as defined in the text. Model sensitivity is the area-weighted global average change in surface temperature between the IPCC A1B scenario (years 71–100) and the present climate (20C3M, years 1971–2000). The rows of the table are ordered in ascending values of relative entropy, and identification of the models is omitted.

temperature is approximately a 90-day average of a quantity whose variability is well simulated by a first order autoregressive model with decorrelation time of about 6 days [*Madden*, 1979]. Even if the distributions are not Gaussian, relative entropy provides a useful measure of the difference in means and covariances that is invariant with respect to linear transformations, and hence does not depend on the coordinate system in which the data is represented. In addition, we assume that the structures in which we are most interested are the leading principal components of the observed system. The principal components in this study were computed in two different ways. In the first way, the principal components (PCs) that maximize the area-weighted variance of the seasonal anomalies were computed. The resulting PCs are of length 400—4 seasons per year for 100 years. The relative entropy computed from these PCs will be called simply "relative entropy." In the second way, four consecutive seasonal anomalies of each year were concatenated to form an augmented state vector that is four times larger than the original state vector, then the principal components that maximize the area-weighted variance of the augmented state vectors were computed. The resulting PCs are of length 100. The relative entropy computed from these PCs will be called "cyclostationary relative entropy." The essential difference between the two ways of computing relative entropy is that the former calculation has PCs of length 400 but neglects the seasonal variation in variance, whereas the latter calculation has PCs of length 100 but accounts for the seasonal variation in variance.

[5] We have also computed the sensitivity to changing greenhouse gases for 13 coupled climate models by taking the difference between two sets of numerical experiments. The first set is for scenario A1B, in which the CO_2 concentration is increased from the current value by 1% per year until it has doubled at a level of 720 ppm, and held constant thereafter. The second set is for the climate of the 20th century (20C3M), in which the models are forced based on the observed time series of CO_2 for the past 100 years. These two sets of numerical experiments are a subset of those prepared for the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4), whose data are available from the IPCC Data Archive (the data are defined on URL: http://www-pcmdi.llnl.gov/ipcc/standard_output.html). The 13 models (listed alphabetically: CNRM-CM3, GFDL-CM2.1, GISS-AOM, GISS-EH, GISS-ER, IPSL-CM4, MIROC3.2(hires), MIROC3.2(medres), MPI-ECHAM5, MRI-CGCM2.3.2, NCAR-CCSM3, NCAR-PCM, UKMO-HadCM3; see URL, http://www-pcmdi.llnl.gov/ipcc/model_documentation/ipcc_model_documentation.php) were chosen for this study based on data availability at the time of the analysis.

[6] We have chosen to compare the models and observations in a state space defined by the leading 15 principal components of the observed seasonal mean surface temperature anomaly field derived from HADCRUT2 (the 5°×5° gridded surface air temperature analysis of *Jones and Moberg* [2003]; see URL: http://www.cru.uea.ac.uk/cru/data/temperature/). This truncation for the PCs captures at least 65% of the variance of the seasonal means, and is thus a natural basis set for measuring how well each model reproduces the observations. Importantly, we include only those grid points in HADCRUT2 for which there are at least two months per season for the full 1899–1998 period.

[7] Table 1 shows the values of relative entropy and model sensitivity for each of the 13 models analyzed. The rows of Table 1 are ordered in ascending values of relative entropy, and identification of the models is omitted. The sensitivity is defined as the area-weighted global average of the change in surface air temperature between the 720 ppm stabilization experiment (A1B; years 71–100 average) and the 20C3M integration (years 1971–2000). The scenario A1B includes a 1% per year increase in CO_2 concentration, until year 70, after which the concentration is held constant at a value of 720 ppm. All models had not continued the integration for an additional 100 years after doubling, so years 71–100 were used in this study. In general, the models which had continued for an additional 100 years show that the global average temperature change for years 171–200 is about 1° K higher than that for years 71–100. The sensitivity, as well as the PCs, were calculated for land grid points only. The values of sensitivity are plotted against relative entropy in Figure 1. Estimates of the uncertainty in the surface temperature change (based on the average standard deviation among ensemble members for those models for which multiple realizations are available), for each model are shown as error bars.

[8] The 99% confidence interval of relative entropy for the sample size appropriate for this study and PC truncation of 15 is less than 1.1 under the null hypothesis that the two samples are drawn from identical distributions. A scan of Table 1 shows that all estimated values of relative entropy are well in excess of this value, indicating that all simulated distributions differ significantly from observations. Furthermore, for every model examined, the value of relative entropy is dominated by the term measuring the difference in the annual cycle, rather than the differences in covariances of seasonal anomalies. This result suggests

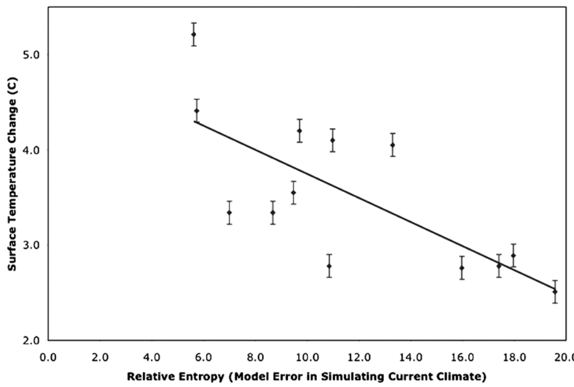

Figure 1. Model sensitivity (surface air temperature change over land) versus model relative entropy for 13 IPCC AR4 models. Estimates of the uncertainty in the surface temperature change are shown as vertical error bars. The line is a least-squares fit to the values.

that we should obtain similar results by measuring fidelity simply by the area-averaged mean square difference between the observed and simulated annual cycles. This measure has been tabulated in Table 1 and can be seen to give results consistent with relative entropy.

[9] We find that there is an apparent relationship between the models' ability to simulate the annual cycle and interannual variability of the present climate and their projection for the climate change due to doubling of CO_2. It is found that models that have lower values of relative entropy, and therefore can be considered to simulate the mean annual cycle and the interannual variability of the present climate more accurately, are more sensitive and produce higher estimates of global warming. This negative correlation between relative entropy and model sensitivity is found to be nearly invariant with respect to the number of PCs used to compute the relative entropy within the range of 5 to 15 PCs. This negative relation can be quantified in various ways. The correlation coefficient between sensitivity and the fidelity measures based on relative entropy, cyclostationary relative entropy, and mean square difference are -0.74, -0.38, and -0.35 respectively, which are statistically significant at the 0.25%, 20%, and 24% levels, respectively; the slope of the line plus its standard error are -0.13 ± 0.03, -0.14 ± 0.11, and -0.11 ± 0.09, all of which differ from zero.

[10] Figure 2 shows the zonal mean (land points only) sensitivity for two models: one with a low value of relative entropy (Model 1; thick curve) and one with the highest relative entropy (Model 10; thin curve). Once again, the sensitivity is consistently higher for the model with a lower value of relative entropy. Although the relationship between relative entropy and model sensitivity is not monotonic at all latitudes for all models (not shown), there is a clear separation between the sensitivity of the high and low relative entropy models, especially in the tropical and Northern Hemisphere extratropical regions.

[11] If we conjecture that models that better simulate the present climate should be considered more credible in projecting the future climate change, then this relationship suggests that the actual changes in global warming will be closer to the highest projected estimates among the current generation of models used in IPCC AR4.

Appendix A

[12] The relative entropy between two distributions, $p_1(x)$ and $p_2(x)$, is defined as

$$R(p_1, p_2) = \int_{R^M} p_1 \log\left(\frac{p_1}{p_2}\right) dx \tag{A1}$$

where the integral is a multiple integral over the range of the M-dimensional vector x. The distribution of seasonal mean surface temperature anomalies is assumed to be Gaussian. Two different assumptions on the time series are considered. The first, based on the "standard PCs" as defined in the main text assumes that the second order moments are stationary while the first order moments are periodically stationary. Under these assumptions, the relative entropy can be written as

$$R(p_1, p_2) = \frac{1}{2} \log\left(\frac{|\Sigma_2|}{|\Sigma_1|}\right) + \frac{1}{2} Tr\left\{\Sigma_1\left(\Sigma_2^{-1} - \Sigma_1^{-1}\right)\right\} + \sum_{k=1}^{4} \frac{1}{2}\left(\mu_1^k - \mu_2^k\right)^T \Sigma_1^{-1}\left(\mu_1^k - \mu_2^k\right) \tag{A2}$$

where μ_j^k is the mean of $p_j(x)$ in the k^{th} season, representing the annual cycle, Σ_j is the covariance matrix of $p_j(x)$, assumed independent of season and based on seasonal anomalies. The second, discussed in the main text, assumes that the augmented state vector is stationary, for which the relative entropy also is given by (A2), except that there is no summation in the last term, and all covariance matrices and means refer to the augmented state vector.

[13] The above expression is independent of nonsingular linear transformations of x. As discussed by *Kleeman* [2002] and *DelSole* [2004], the distribution of observed temperature is appropriately identified with p_1, and the distribution of model simulated temperature with p_2. It is perhaps worth mentioning that relative entropy is a measure

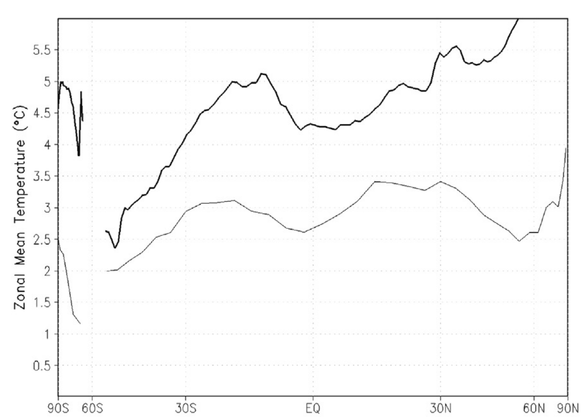

Figure 2. Zonal mean (land points only) sensitivity (surface air temperature change) for two selected models: (Model 1; relative entropy = 5.62; thick curve) and (Model 10; relative entropy = 15.97; thin curve).

of the difference in variability without regard to how that variability is correlated in time, and so is not a measure of how well the "trend" or low frequency variability is simulated. The first two terms essentially measure how well the variability of seasonal anomalies in the simulation approximates the variability in observations. The last term measures how well the annual cycle is simulated by the models.

[14] The principal components of observations were computed by first projecting all model fields onto the $5^\circ \times 5^\circ$ observational grid and masking out regions of missing data, then multiplying the value at each grid point by the square root of the cosine of latitude so that the square of the value is weighted by area. The masking procedure eliminates all but 301 grid points. The resulting defined points cover most of North America, Europe, and India, and certain coastal regions of East Asia, South America, and Australia.

[15] The sampling distribution of relative entropy, under the null hypothesis that the two M-dimensional multivariate normal distributions p_1 and p_2 are identical, was determined through Monte Carlo methods. The 99% confidence interval for relative entropy computed in this way was verified against the correct asymptotic results discussed extensively by *Kullback* [1959]. The 1% significance level for 100 years of data and 15 principal components was found to be 1.1.

[16] **Acknowledgments.** We thank Jennifer Adams and Brian Doty of COLA for their invaluable help. This research was supported by the NSF (ATM-0332910), NOAA (NA04OAR4310034) and NASA (NNG04GG46G). We thank Phil Jones and colleagues of CRU for the gridded observational data. We acknowledge the international modeling groups for providing their data for analysis, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) for collecting and archiving the model data, the JSC/CLIVAR Working Group on Coupled Modeling (WGCM) and their Coupled Model Intercomparison Project (CMIP) and Climate Simulation Panel for organizing the model data analysis activity, and the IPCC WG1 TSU for technical support. The IPCC Data Archive at Lawrence Livermore National Laboratory is supported by the Office of Science, U.S. Department of Energy.

References

Charney, J. G., R. Fleagle, V. Lally, H. Riehl, and D. Wark (1966), The feasibility of a global observation and analysis experiment, *Bull. Am. Meteorol. Soc.*, *47*, 200–220.

Cover, T. M., and J. A. Thomas (1991), *Elements of Information Theory*, 576 pp., John Wiley, Hoboken, N. J.

DelSole, T. (2004), Predictability and information theory. Part I: Measures of predictability, *J. Atmos. Sci.*, *61*, 2425–2440.

Houghton, J. T., Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, and D. Xiaosu (Eds.) (2001), *Climate Change 2001: The Scientific Basis—Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change (IPCC)*, 944 pp., Cambridge Univ. Press, New York.

Jones, P. D., 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.

Kleeman, R. (2002), Measuring dynamical prediction utility using relative entropy, *J. Atmos. Sci.*, *59*, 2057–2072.

Kullback, S. (1959), *Information Theory and Statistics*, 399 pp., John Wiley, Hoboken, N. J.

Madden, R. (1979), A simple approximation for the variance of meteorological time averages, *J. Appl. Meteorol.*, *18*, 703–706.

Simmons, A. J., and A. Hollingsworth (2002), Some aspects of the improvement in skill of numerical weather prediction, *Q. J. R. Meteorol. Soc.*, *128*, 647–677.

Tippett, M. K., R. Kleeman, and Y. Tang (2004), Measuring the potential utility of seasonal climate predictions, *Geophys. Res. Lett.*, *31*, L22201, doi:10.1029/2004GL021575.

T. DelSole, J. Kinter, and J. Shukla, Climate Dynamics Program, School of Computational Sciences, George Mason University, MS5C3, 4400 University Drive, Fairfax, VA 22030, USA. (shukla@cola.iges.org)

M. Fennessy and D. Paolino, Center for Ocean-Land-Atmosphere Studies, 4041 Powder Mill Road, Suite 302, Calverton, MD 20705, USA.