Constraints on climate change from a multi-thousand member ensemble of simulations

C. Piani, D. J. Frame, D. A. Stainforth, and M. R. Allen
Department of Physics, University of Oxford, Oxford, UK

Received 2 September 2005; revised 31 October 2005; accepted 3 November 2005; published 15 December 2005.

[1] The first multi thousand member "perturbed physics" ensemble simulation of present and future climate, completed by the distributed computing project climate*prediction*.net, is used to search for constraints on the response to increasing greenhouse gas levels among present day observable climate variables. The search is conducted with a systematic statistical methodology to identify correlations between observables and the quantities we wish to predict, namely the climate sensitivity and the climate feedback parameter. A sensitivity analysis is conducted to ensure that results are minimally dependent on the parameters of the methodology. Our best estimate of climate sensitivity is 3.3 K. When an internally consistent representation of the origins of model-data discrepancy is used to calculate the probability density function of climate sensitivity, the 5th and 95th percentiles are 2.2 K and 6.8 K respectively. These results are sensitive, particularly the upper bound, to the representation of the origins of model-data discrepancy. **Citation:** Piani, C., D. J. Frame, D. A. Stainforth, and M. R. Allen (2005), Constraints on climate change from a multi-thousand member ensemble of simulations, *Geophys. Res. Lett.*, *32*, L23825, doi:10.1029/2005GL024452.

1. Introduction

[2] Single forecasts of the climate response to increasing greenhouse gas levels, are far more useful to policy makers when they are accompanied by some measure of the associated uncertainty [*Schneider*, 2001; *Allen and Stainforth*, 2002]. Establishing the best guess for mean global temperature change at 3 K or 3.5 K is a moot point, while determining whether there is a 5% or 25% chance of a mean global temperature change of more than 5 K under a 550 ppm CO_2 stabilization scenario is all important [*Palmer and Raisanen*, 2002; *Gregory et al.*, 2002; *Stott and Kettleborough*, 2002; *Smith*, 2002; *Forrest et al.*, 2002; *Andronova and Schlesinger*, 2001]. Past studies have addressed this issue by assembling models from different research groups to constitute an 'ensemble of opportunity' of which the mean and a variance of climate sensitivity (CS) are used to produce a probability density function (PDF) [*Covey et al.*, 2003]. The results from such studies are, by construction, dependent on the choice of ensemble. A natural progression of this method leads to weight members of these ensembles by some measure of their similarity to observations. This approach is very likely to underestimate uncertainties since the individual models that constitute the ensemble are tuned to simulate present day climate well, hence observations are likely to be double-counted during the weighting process [*Allen et al.*, 2002; *Giorgi and Mearns*, 2002].

[3] Recently *Murphy et al.* [2004] presented results from a 58 member ensemble in which poorly constrained parameters are varied within expert-defined ranges. By allowing the parameters to vary the authors reduce the extent to which the model is tuned to observations. The results from the ensemble are weighted according to their associated likelihood, given a range of present day climate observations, to produce an estimated PDF for CS. Results were found to be sensitive to the sampling strategy used to perturb model parameters. Likewise, a frequency distribution of CS from the first 2000+ simulations from climate*prediction*.net was presented by *Stainforth et al.* [2005] and was also found to be highly sensitive to the parameter sampling strategy used, leading those authors to conclude they could not assign objective probabilities to the different outcomes observed. The only instance in which a likelihood weighted ensemble will produce a result that is independent of the ensemble itself, is if the ensemble samples 'model space' uniformly [*Frame et al.*, 2005]. This is, of course, conceptually impossible since there is no way of objectively establishing a metric for the space of all possible models [*Allen et al.*, 2002].

[4] Here we will use the latest results from the climate*prediction*.net (CPDN) project [*Allen*, 1999; *Stainforth et al.*, 2002; *Hansen et al.*, 2001] to apply a methodology for constraining predictions of future climate variables from grand ensembles of experiments. The CPDN project uses distributed computing resources [*Stainforth et al.*, 2002] to involve members of the public (with over 100000 participants to date) in an unprecedented global climate simulation project. Each participant conducts one or more experiments, each comprised of three separate simulations using a predetermined setting of parameter values and initial conditions. In this initial phase of the project, a slab version of the Met Office unified model (referred to as HadSM3) was used [*Pope et al.*, 2000]. The model has a horizontal resolution of $3.75^\circ \times 2.5^\circ$ and 19 atmospheric levels reaching 10mb. The ocean is simulated by a single layer. Each individual model experiment is comprised of 3 phases. In the first 15yr "calibration" phase the sea surface temperatures (SSTs) are held constant, and are the same for all experiments, while the resulting heat-flux convergence field is recorded and merged into a seasonal climatology. In the following 15yr "control" and "double CO_2" phases the SSTs are allowed to vary subject to atmospheric-ocean heat exchange with the additional input of heat-flux convergence from the seasonal climatology obtained from the calibration

0094-8276/05/2005GL024452$05.00

Table 1. Climate Variables Chosen for This Analysis[a]

Climate Variable	Source	Space Subset
1.5m Temperature	CRU	85°N to 85°S, land only, latitude-longitude grid
Mean sea level pressure	ERA40	85°N to 85°S, latitude-longitude grid
Precipitation	*Xie and Arkin* [1997]	85°N to 85°S over land, 30°N to 30°S over ocean, latitude-longitude grid
Surface sensible & latent heat fluxes	SOC	85°N to 40°S ocean only, latitude-longitude grid
Relative Humidity, Temperature, Zonal and Meridional winds	ERA40	85°N to 40°S, zonal mean height-latitude grid
Outgoing longwave & shortwave radiation	ERBE	85°N to 40°S, 1D zonal mean latitude grid

[a]Also shown are corresponding observational data set used for each climate variable [*Jones et al.*, 1999; *Gibson et al.*, 1997; *Xie and Arkin*, 1997; *New et al.*, 1999; *Harrison et al.*, 1990] and choice of grid and spatial averaging. The observational and reanalysis data used are multi year averages representing near present day climate.

phase. The heat-flux-correction minimizes the drift in SSTs. In very few cases the flux-correction does not eliminate the drift entirely, these models are purged during quality control which leaves roughly over 2000 viable experiments. In the 3rd phase, CO_2 concentrations are doubled. The global mean surface temperatures from the 2nd and 3rd phases are used to derive CS [*Andronova et al.*, 2005] and the feedback parameter (FP). This is done using the same extrapolation technique described by *Stainforth et al.* [2005].

2. Methodology

[5] A predictor of CS and the FP (defined here as 1/CS) is searched for among the variables of the control climate (phase 2) for which we have observational and reanalysis data sets. In this context the word ‘climate’ refers to the set of variables described in Table 1. These variables were chosen according to availability, from both observation datasets and model output, and to maximize consistency with prior perturbed physics ensemble experiments [*Murphy et al.*, 2004]. Here we focus on linear predictors to simplify the methodology and minimize the risk of over fitting, but a natural generalization would be to relax the linearity condition [*Knutti et al.*, 2003; R. T. Knutti, Constraining climate sensitivity from the seasonal cycle in surface temperature, submitted to *Geophysical Research Letters*, 2005]. Accurate uncertainty analysis requires us to allow for correlations between these observable quantities. Following the practice of climate change detection studies, we assume the correlation structure of model-data discrepancies can be modeled by internal climate variability [*Hergel and Allen*, 2002; *Morgan and Keith*, 1995]. Hence we begin by taking a 496 year control climate simulation of the HadCM3 coupled model. This is then divided into 62 non-overlapping 8 year periods over which the mean is taken. From each of these the mean over the entire 496 year period is subtracted leaving 62 independent climate anomalies which we take to represent internal climate variability. We continue by projecting both observations and members of the CPDN ensemble onto the EOFs of the 62 HadCM3 climate anomalies, with individual variables within these EOFs weighted by the inverse of the globally averaged standard deviation of HadCM3 climate variability. We then find linear combinations of these EOFs (“rotated” or REOFs) such that, when members of the CPDN ensemble are projected onto them, the resulting projections are uncorrelated with each other. We assume the quantities we wish to estimate (CS and FP) can be predicted by a simple linear model:

$$y_i = \sum_{j=1}^{K} x_{ij}\beta_j + noise \tag{1}$$

where y_i is CS or FP of the ith member of the CPDN ensemble, x_{ij} is the amplitude of the jth REOF in the ith ensemble member, expressed as anomalies about the ensemble mean, and K is the number of REOFs retained. The values of β_j that minimize the squared prediction error across the CPDN ensemble are given by the standard regression formula with uncorrelated independent variables:

$$\hat{\beta}_j = \sum_{i=1}^{N} \mu_j^{-2} x_{ji}^T y_i, \tag{2}$$

where μ_j^2 is the variance in the jth REOF across the CPDN ensemble. Our best estimate of the CS in the real world is given by:

$$y_o = \sum_{j=1}^{K} x_{oj}\hat{\beta}_j, \tag{3}$$

where the x_{oj} is the projection of the observations, also expressed as anomalies about the CPDN ensemble mean, onto the jth REOF; and the variance associated with it is given by:

$$\sigma(y_m)^2 = \frac{1}{M-1}\sum_{m=1}^{M}\left(\sum z_{mj}\hat{\beta}_j\right)^2, \tag{4}$$

where z_{mj} is a segment of HadCM3 control climate, independent of the one used to derive the REOFs, also expressed as an anomaly from the CPDN ensemble mean.

[6] We can examine the adequacy of our representation of model-data discrepancy by comparing the variance of the observations and reanalysis data in directions perpendicular to our linear predictor of either CS or FP (in the subspace spanned by the K REOFs) with the variance expected from HadCM3. If the representation of model-data discrepancy is adequate, this quantity should conform to the $F(K, \nu)$ distribution, with ν being the estimated degrees of freedom of our control climate. Results are shown in Figure 1. The shaded area in Figure 1 is the region of values that pass the

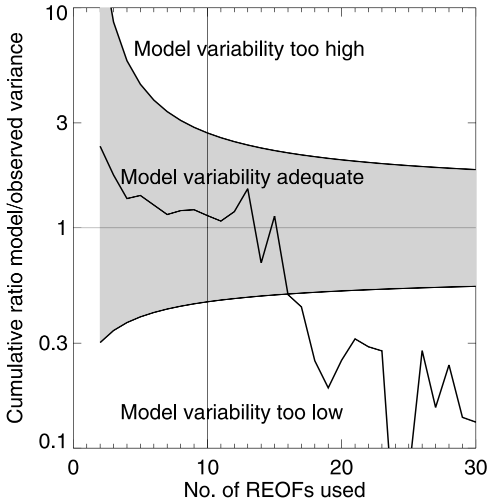

Figure 1. Cumulative ratio of model/observations variance, in the direction perpendicular to β, as a function of no. of REOFs used in the prediction equation (2). See color version of this figure in the HTML.

F-test at the 90% level. From Figure 1, it is clear that we are justified in truncating our REOF series at 10. Slightly higher truncations would also be acceptable, but give identical results. Expressing HadCM3 model segments about the HadCM3 control climate results in an unrealistically low representation of model-data discrepancy, meaning even with the best-fit combination of parameters, the observation-model discrepancy cannot (unsurprisingly) be attributed entirely to internal climate variability. Expressing coupled model variability about the CPDN (slab model) ensemble mean provides a crude measure of the magnitude of possible "systematic" model errors.

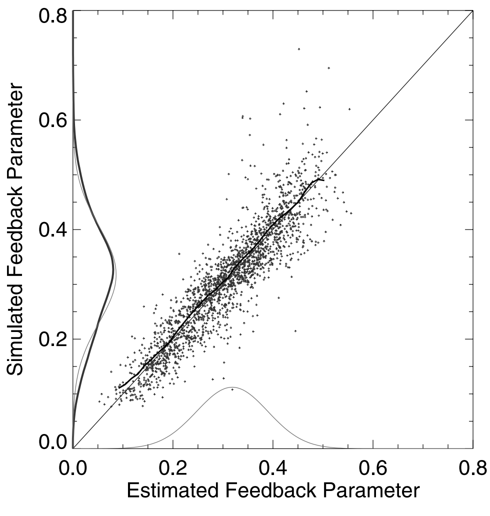

Figure 2. Scatter plot of simulated feedback parameter from the Climate*Prediction*.net ensemble (each dot represents one model) versus feedback parameter estimated using equation (1) (REOF truncation number is 10). The thin line on the x-axis is the estimated PDF of the feedback parameter, obtained as a Gaussian with mean given by equation (3) and variance given by equation (4). The thin line on the y-axis is the PDF obtained by simply adding the unexplained variance onto the x-axis PDF. The thick line on the y-axis is the PDF obtained using the scatter plot itself as a transfer function. See color version of this figure in the HTML.

3. Results

[7] Equation (3) gives us our best estimate of CS and FP (3.3 K and 0.32 K^{-1} respectively). Note that the best estimate for FP is not the exact inverse of the best estimate of CS because a different optimal linear predictor is derived from equation (2). The variance from equation (4) and the best estimate of FP determine the Gaussian PDF on the x-axis of Figure 2. Here the scatter plot represents values of FP (one dot for each CPDN model) obtained from the double CO_2 experiments (y-axis) versus estimates of the same from equation (1) (x-axis). With a truncation number of 10, only 0.26 of the variance in FP is unexplained by the linear model. When this variance is added uniformly to all points, the PDF on the y-axis is obtained (thin line). If instead one uses the median of the scatter plot and the distribution about the median as a transfer function for the horizontal distribution one obtains a PDF that is only slightly different (thick line on the y-axis).

[8] Figure 3 is similar to Figure 2 but predicting CS instead of FP. We immediately see that, even though we have specifically looked for an optimal linear predictor, the relationship between the values of the predictor and the simulated climate sensitivity is non-linear, as one would expect if all observable aspects of present-day climate scale with the strength of atmospheric feedbacks and not the climate sensitivity. The fraction of unexplained variance in CS is 0.31. The PDF on the y-axis in Figure 3 (solid thick

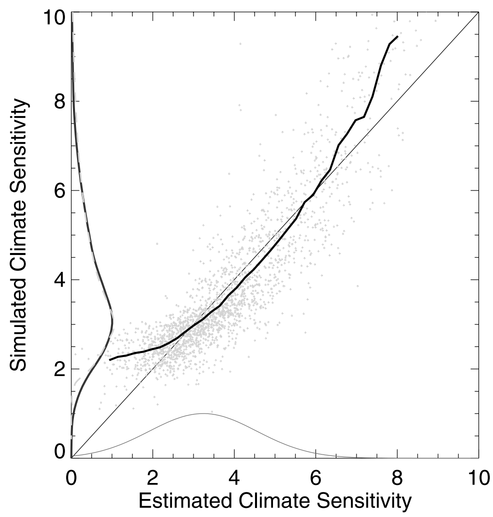

Figure 3. The scatter plot, median and thin PDF are the same as Figure 2 but for climate sensitivity. The solid thick PDF is obtained from the thick PDF in Figure 2 by plotting it onto the inverse axis. The thick dashed line is obtained by using the median and the scatter about the median as a transfer function for the x-axis PDF. The shape of the median graphically explains why observationally constrained forecasts of CS are likely to be skewed to higher values. See color version of this figure in the HTML.

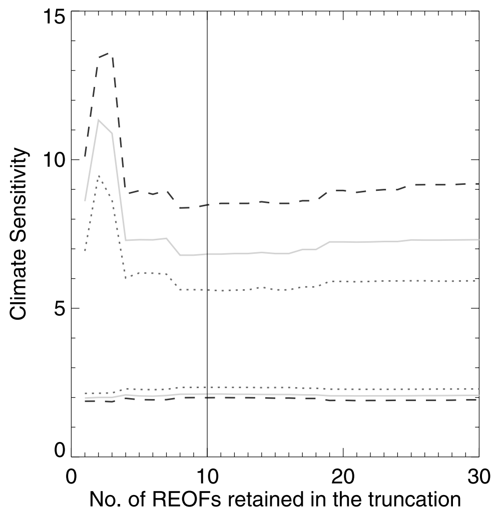

Figure 4. Dashed line: values of the 2.5% and 97.5% cutoff values of climate sensitivity as a function of the REOF truncation number. Solid line: same but for the 5% and 95% cutoff values. Dotted line: same but for the 10% and 90% cutoff values. See color version of this figure in the HTML.

line) is not obtained as in Figure 2 because there are insufficient data points at the lower end where the PDF on the x-axis is still significantly large. Instead the thick solid PDF in Figure 3 is obtained simply by plotting the thick PDF in Figure 2 against sensitivity rather than feedback parameter. This distribution can be interpreted as a probability density function for CS assuming all values of sensitivity are considered equally likely at the outset, just as the thick solid distribution in Figure 2 can be interpreted as a PDF of FP assuming a similar "uniform prior" in that quantity. Unlike the distributions shown in previous analyses of perturbed-physics ensembles, this distribution should depend on the validity of the transfer function between observable and predicted climate variables, which could be tested, for example with an independent climate model, and not on subjective decisions made about how to sample possible models or parameter settings. In the regions where the scatter plot can be used to project the x-axis PDF onto the y-axis in Figure 3, the resulting PDF (dashed thick line) matches the solid thick one.

[9] Comparing Figures 2 and 3 it is clear that observable quantities scale with FP and not with CS. (That is why the thick PDF on Figure 2 matches the thin y-axis Gaussian PDF almost perfectly while the thick solid PDF in Figure 3 is 'inverse Gaussian' and is hence significantly skewed towards larger values.) As long as observational error is Gaussian, the observationally-constrained PDF of CS must be 'inverse Gaussian' and upper bounds remain rather sensitive to the adopted representation of model-data discrepancy.

[10] When truncating the REOF series at 10, the 5th and 95th percentiles of the estimated distribution for CS are 2.2 K and 6.8 K respectively. Although we can calculate more extreme cutoff values, they are of questionable significance, because we only have a limited segment of control variability with which to represent expected model-data discrepancies. The qualitative observation that likelihood falls off considerably more slowly towards high sensitivities than towards low values arises, however, from the physically reasonable result that observable quantities scale with feedback strength, not sensitivity, and so is likely to be robust. To asses the dependence of our results on the number of REOFs retained, we plot the cutoff values of climate sensitivity as a function of the truncation number (Figure 4). This shows that adding of removing a few REOFs to our analysis has little effect.

[11] Predictor-based constraints on future climate, are expected to be quite insensitive to the ensemble sampling strategy when compared to constraints based on likelihood weighting techniques. This is because predictor-based methods identify intrinsic links in the climate system between observable variables and quantifiable future changes. To the extent that the HadSM3 is a correct representation of the real climate system (+ noise), and as confirmed by preliminary studies of sensitivity to choices of parameter perturbations, these results are independent of the particular parameter perturbation sampling strategy used in the construction of the perturbed physics ensemble.

[12] **Acknowledgments.** We thank the research staff of the climate-*prediction*.net group for their support. Special thanks to David Sexton of the Hadley Centre for Climate Prediction and Research for providing part of the observation data-sets and comments. This work was supported by the Natural Environmental Research Council's COAPEC, e-Science and fellowship programmes.

References

Allen, M. R. (1999), Do-it-yourself climate prediction, *Nature*, *401*, 627.
Allen, M. R., and D. A. Stainforth (2002), Towards objective probabilistic climate forecasting, *Nature*, *419*, 228.
Allen, M. R., J. A. Kettleborough, and D. Stainforth (2002), Model error in weather and climate forecasting, paper presented at 2002 ECMWF Predictability Seminar, Eur. Cent. for Medium Range Weather Forecasting, Reading, U. K.
Andronova, N. G., and M. E. Schlesinger (2001), Objective estimation of the probability density function for climate sensitivity, *J. Geophys. Res.*, *106*, 22,605–22,612.
Andronova, N., et al. (2005), The concept of climate sensitivity: History and development, in *Human Induced Climate Change: An Interdisciplinary Assessment*, edited by M. E. Schlesinger et al., Cambridge Univ. Press, New York, in press.
Covey, C., et al. (2003), An overview of results from the coupled model intercomparison project, *Global Planet. Change*, *37*(1–2), 103–133.
Forrest, C., et al. (2002), Quantifying uncertainties in climate system properties with the use of recent climate observations, *Science*, *295*, 113–117.
Frame, D. J., B. B. B. Booth, J. A. Kettleborough, D. A. Stainforth, J. M. Gregory, M. Collins, and M. R. Allen (2005), Constraining climate forecasts: The role of prior assumptions, *Geophys. Res. Lett.*, *32*, L09702, doi:10.1029/2004GL022241.
Gibson, J. K., et al. (1997), ERA description, *ECMWF Re-Anal. Proj. Rep. Ser. 1*, Eur. Cent. for Medium Range Weather Forecasting, Reading, U. K.
Giorgi, F., and L. O. Mearns (2002), Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the reliability ensemble averaging (REA) method, *J. Clim.*, *15*, 1141–1158.
Gregory, J. M., et al. (2002), An observationally-based estimate of climate sensitivity, *J. Clim.*, *15*, 3117–3121.
Hansen, J. A., et al. (2001), Casino-21: Climate simulation of the 21st century, *World Res. Rev.*, *13*, 187–189.
Harrison, E. F., P. Minnis, B. R. Barkstrom, V. Ramanathan, R. D. Cess, and G. G. Gibson (1990), Seasonal variation of cloud radiative forcing derived from the Earth radiation budget experiment, *J. Geophys. Res.*, *95*, 18,687–18,703.
Hergel, G. C., and M. R. Allen (2002), Origins of model-data discrepancy in optimal fingerprinting, *J. Clim.*, *15*, 1348–1356.
Jones, P. D., M. New, D. E. Parker, S. Martin, and I. G. Rigor (1999), Surface air temperature and its changes over the past 150 years, *Rev. Geophys.*, *37*, 173–200.

Knutti, R. T., et al. (2003), Probabilistic climate change projections using neural networks, *Clim. Dyn.*, *21*, doi:10.1007/s00382-003-0345-1.

Morgan, M. G., and D. W. Keith (1995), Climate change—Subjective judgements by climate experts, *Environ. Sci. Technol.*, *29*, 468–476.

Murphy, J. M., D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. I. Webb, M. Collins, M. R. Allen, and D. J. Stainforth (2004), Quantifying uncertainties in climate change using a large ensemble of general circulation model predictions, *Nature*, *430*, 768–772.

New, M., M. Hulme, and P. Jones (1999), Representing twentieth-century space-time climate variability. part I: Development of a 1961–90 mean monthly terrestrial climatology, *J. Clim.*, *12*, 829–856.

Palmer, T. N., and J. Raisanen (2002), Quantifying the risk of extreme seasonal precipitation events in a changing climate, *Nature*, *415*, 512–517.

Pope, V. D., et al. (2000), The impact of new physical parameterizations in the Hadley Centre climate model, HadAM3, *Clim. Dyn.*, *16*, 123–146.

Schneider, S. H. (2001), What is dangerous climate change?, *Nature*, *411*, 17–19.

Smith, L. (2002), What might we learn from climate forecasts?, *Proc. Natl. Acad. Sci. U.S.A.*, *99*, 2487–2492.

Stott, P. A., and J. A. Kettleborough (2002), Origins and estimates of uncertainty in predictions of twenty-first century temperature rise, *Nature*, *416*, 723–726.

Stainforth, D., et al. (2002), Distributed computing for public interest climate modelling research, *Comput. Sci. Eng.*, *4*, 82–89.

Stainforth, D., et al. (2005), Uncertainty in predictions of climate response to rising levels or greenhouse gases, *Nature*, *433*, 403–406.

Xie, P., and P. A. Arkin (1997), Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates and numerical model outputs, *Bull. Am. Meteorol. Soc.*, *78*, 2539–2558.

M. R. Allen, D. J. Frame, C. Piani, and D. A. Stainforth, Department of Physics, University of Oxford, Parks Road, Oxford OX1 3PU, UK. (cpiani@atm.ox.ac.uk)