Uncertainty in predictions of the climate response to rising levels of greenhouse gases

D. A. Stainforth[1], T. Aina[1], C. Christensen[2], M. Collins[3], N. Faull[1], D. J. Frame[1], J. A. Kettleborough[4], S. Knight[1], A. Martin[2], J. M. Murphy[3], C. Piani[1], D. Sexton[3], L. A. Smith[5], R. A. Spicer[6], A. J. Thorpe[7] & M. R. Allen[1]

[1]*Department of Physics, University of Oxford, Parks Road, Oxford OX1 3PU, UK*
[2]*Computing Laboratory, University of Oxford, Parks Road, Oxford OX1 3QD, UK*
[3]*Hadley Centre for Climate Prediction and Research, Met Office, Exeter EX1 3PB, UK*
[4]*Rutherford Appleton Laboratory, Chilton, Oxfordshire, OX11 0QX, UK*
[5]*London School of Economics, London WC2A 2AE, UK*
[6]*Department of Earth Sciences, The Open University, Milton Keynes MK7 6AA, UK*
[7]*Department of Meteorology, University of Reading, Reading RG6 6BB, UK*

The range of possibilities for future climate evolution[1–3] needs to be taken into account when planning climate change mitigation and adaptation strategies. This requires ensembles of multi-decadal simulations to assess both chaotic climate variability and model response uncertainty[4–9]. Statistical estimates of model response uncertainty, based on observations of recent climate change[10–13], admit climate sensitivities—defined as the equilibrium response of global mean temperature to doubling levels of atmospheric carbon dioxide—substantially greater than 5 K. But such strong responses are not used in ranges for future climate change[14] because they have not been seen in general circulation models. Here we present results from the 'climate*prediction*.net' experiment, the first multi-thousand-member grand ensemble of simulations using a general circulation model and thereby explicitly resolving regional details[15–21]. We find model versions as realistic as other state-of-the-art climate models but with climate sensitivities ranging from less than 2 K to more than 11 K. Models with such extreme sensitivities are critical for the study of the full range of possible responses of the climate system to rising greenhouse gas levels, and for assessing the risks associated with specific targets for stabilizing these levels.

As a first step towards a probabilistic climate prediction system we have carried out a grand ensemble (an ensemble of ensembles) exploring uncertainty in a state-of-the-art model. Uncertainty in model response is investigated using a perturbed physics ensemble[4] in which model parameters are set to alternative values considered plausible by experts in the relevant parameterization schemes[9]. Two or three values are taken for each parameter (see Methods); simulations may have several parameters perturbed from their standard model values simultaneously. For each combination of parameter values (referred to here as a 'model version') an initial-condition ensemble[22] is used, creating an ensemble of ensembles. Each individual member of this grand ensemble (referred to here as a 'simulation') explores the response to changing boundary conditions[22] by including a period with doubled CO_2 concentrations.

The general circulation model (GCM) is a version of the Met Office Unified Model consisting of the atmospheric model HadAM3[23], at standard resolution[9] but with increased numerical stability, coupled to a mixed-layer ocean. This allows us to explore the effects of a wide range of uncertainties in the way the atmosphere is represented, while avoiding a long spin-up for each model version. Each simulation involves three 15-year phases: (1) calibration, to deduce the ocean heat-flux convergence field used in the subsequent phases; (2) control, used to quantify the relevance of the particular model version and heat-flux convergence field; and (3) doubled CO_2, to explore the response to changing boundary conditions.

Individual simulations are carried out using idle processing capacity on personal computers volunteered by members of the general public[19]. This distributed-computing method[16,18,19] leads to a continually expanding data set of results, requiring us to use a specified subset of data available at a specific point in time. The analysis presented here uses 2,578 simulations (>100,000 simulated years), chosen to explore combinations of perturbations in six parameters.

The 2,578 simulations contain 2,017 unique simulations (duplicates are used to verify the experimental design—see Methods). Figure 1a shows the grand ensemble frequency distribution of global mean, annual mean, near-surface temperature (T_g) in these 2,017 simulations, as it develops through each phase. Some model versions show substantial drifts in the control phase owing to the use of a simplified ocean (see Supplementary Information). We remove unstable simulations (see Methods) and average over initial-condition ensembles of identical model versions to reduce sampling uncertainty. The frequency distribution of initial-condition-ensemble-mean time series of T_g for the resulting 414 model versions (for which the initial-condition ensembles involve 1,148 independent stable simulations) is shown in Fig. 1b. Six of these model versions show a significant cooling tendency in the doubled-CO_2 phase. This cooling is also due to known limitations with the use of a simplified ocean (see Supplementary Information) so these simulations are excluded from the remaining analysis of sensitivity.

The frequency distribution of the simulated climate sensitivities (see Methods) for the remaining model versions is shown in Fig. 2a and ranges from 1.9 to 11.5 K. Two key features are that relatively few model versions have sensitivities less than 2 K, and the long tail of the distribution extending to very high values; 4.2% are >8 K. Most sensitivities cluster round 3.4 K, the value for the unperturbed model, suggesting that many of the parameter combinations

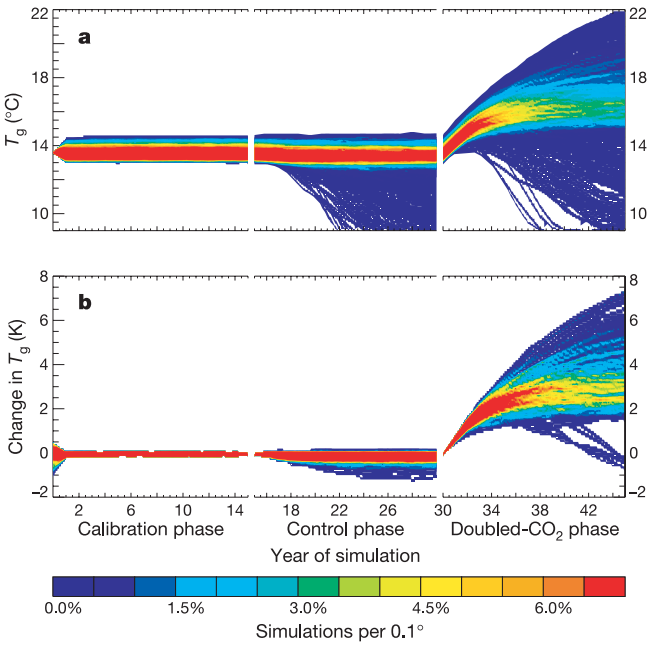

Figure 1 Frequency distributions of T_g (colours indicate density of trajectories per 0.1 K interval) through the three phases of the simulation. **a**, Frequency distribution of the 2,017 distinct independent simulations. **b**, Frequency distribution of the 414 model versions. In **b**, T_g is shown relative to the value at the end of the calibration phase and where initial-condition ensemble members exist, their mean has been taken for each time point.

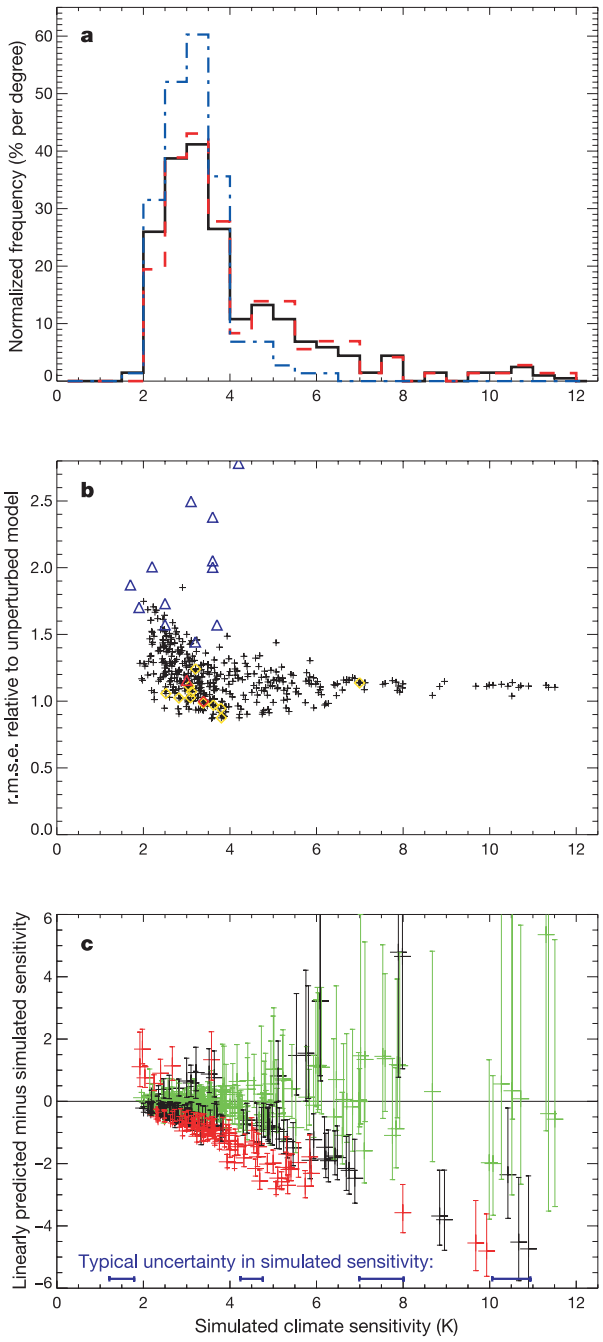

Figure 2 The response to parameter perturbations. **a**, The frequency distribution of simulated climate sensitivity using all model versions (black), all model versions except those with perturbations to the cloud-to-rain conversion threshold (red), and all model versions except those with perturbations to the entrainment coefficient (blue). **b**, Variations in the relative r.m.s.e. of model versions. The unperturbed model is shown by the red diamond. Model versions with only a single parameter perturbed are highlighted by yellow diamonds. The triangles show the CMIP II models for which data are available; HadCM3 (having the same atmosphere as the unperturbed model but with a dynamic ocean) is shown in red and the others in blue. **c**, Linear prediction of climate sensitivity based on summing the change in λ for the relevant single-parameter-perturbation model versions, to estimate λ when multiple perturbations are combined. Error bars show the resulting uncertainty ($\pm$ one sigma) caused by the combination of a number of $\Delta\lambda$ values where each λ has an uncertainty deduced from the initial-condition ensembles having only a single parameter perturbed. Linear predictions within one sigma of the simulated value are shown in green, between one and two sigma in black, and above two sigma in red. Mean uncertainties in the simulated value (two-sigma range, inferred from the initial-condition ensembles) are shown at the bottom for four regions of sensitivity (0–3, 3–6, 6–9, 9–12).

explored have relatively little effect on this global variable. There are a number of possible reasons for this clustering: the relevant processes may in fact have only a limited impact on sensitivity, the parameter ranges used may be too small to influence substantially the response in this model, and/or multiple perturbations may have mutually compensating effects when averaged on global scales. Of course, many significant regional impacts are invisible in a global average.

The range of sensitivities across different versions of the same model is more than twice that found in the GCMs used in the IPCC Third Assessment Report[14]. The possibility of such high sensitivities has been reported by studies using observations to constrain this quantity[9,11,24,25], but this is the first time that GCMs have generated such behaviour. The shape of the distribution is determined by the parameters selected for perturbation and the perturbed values chosen, which were relatively arbitrary. Model developers provided plausible high and low values for each model parameter; however, we cannot interpret these as absolute upper and lower bounds because experts are known to underestimate uncertainty even in straightforward elicitation exercises where the import of the question is clear[26]. In our case even the physical interpretation of many of these parameters is ambiguous[27]. We can illustrate the importance of the parameter choices by subsampling the model versions. If all perturbations to one parameter (the cloud-to-rain conversion threshold) are omitted, the red histogram in Fig. 2a is obtained, with a slightly increased fraction (4.9%) of model versions >8 K. If perturbations to another parameter (the entrainment coefficient) are omitted, the blue histogram in Fig. 2a is obtained, with no model versions >8 K. (See Supplementary Information for further sensitivity analyses.)

Can either high-end or low-end sensitivities be rejected on the basis of the model-version control climates? Fig. 2b suggests not; it illustrates the relative ability of model versions to simulate observations using a global root-mean-squared error (r.m.s.e.) normalized by the errors in the unperturbed model (see Methods). For all model versions this relative r.m.s.e. is within (or below) the range of values for other state-of-the-art models, such as those used in the second Coupled Model Inter Comparison (CMIP II) project[28] (triangles). The five variables used for this comparison are each standard variables in model evaluation and inter-comparison exercises[29] (see Methods). This lack of an observational constraint, combined with the sensitivity of the results to the way in which parameters are perturbed, means that we cannot provide an objective probability density function for simulated climate sensitivity. Nevertheless, our results demonstrate the wide range of behaviour possible within a GCM and show that high sensitivities cannot yet be neglected as they were in the headline uncertainty ranges of the IPCC Third Assessment Report (for example, the 1.4–5.8 K range for 1990 to 2100 warming).[14] Further, they tell us about the sensitivities of our models, allowing better-informed decisions on resource allocation both for observational studies and for model development.

Can we coherently predict the model's response to multiple parameter perturbations from a small number of simulations each of which perturbs only a single parameter[9]? The question is important because it bears on the applicability of linear optimization methods in the design and analysis of smaller ensembles. Figure 2c shows that assuming that changes in the climate feedback parameter[14] λ combine linearly provides some insight, but fails in two important respects. First, combining uncertainties gives large fractional uncertainties for small predicted λ and hence large uncertainties for high sensitivities. This effect becomes more pronounced the greater the number of parameters perturbed. Second, this method systematically underestimates the simulated sensitivity, as shown in Fig. 2c, and consequently artificially reduces the implied likelihood of a high response. Furthermore, more than 20% of the linear predictions are more than two standard errors from the

simulated sensitivities. Thus, comprehensive multiple-perturbed-parameter ensembles appear to be necessary for robust probabilistic analyses.

Figure 3 shows the initial-condition ensemble-mean of the temperature and precipitation changes for years 8–15 after doubling CO_2 concentrations, for three model versions: (1) the unperturbed model; (2) a version with low sensitivity; and (3) a version with high sensitivity (see Supplementary Information for details of the control climates in these model versions). All three models show the familiar increased warming at high latitudes and the overall surface-temperature pattern scales with sensitivity. Even in the low-sensitivity model version the warming in certain regions is substantial, exceeding 3 K in Amazonia and 4 K in much of North America. The precipitation field shows a greater variety of response. For instance, this particular low-sensitivity model version shows a region of substantially reduced precipitation east of the Mediterranean; something not evident in either the standard or high-sensitivity model versions shown. It is critical to note that model versions with similar sensitivities often also show differences in such regional details[9]. The use of a GCM-based grand ensemble allows the significance of such details to be ascertained.

Thanks to the participation and enthusiasm of tens of thousands of individuals world-wide we have been able to discover GCM versions with comparatively realistic control climates and with sensitivities covering a much wider range than has ever been seen before. These results are a critical step towards a better under-

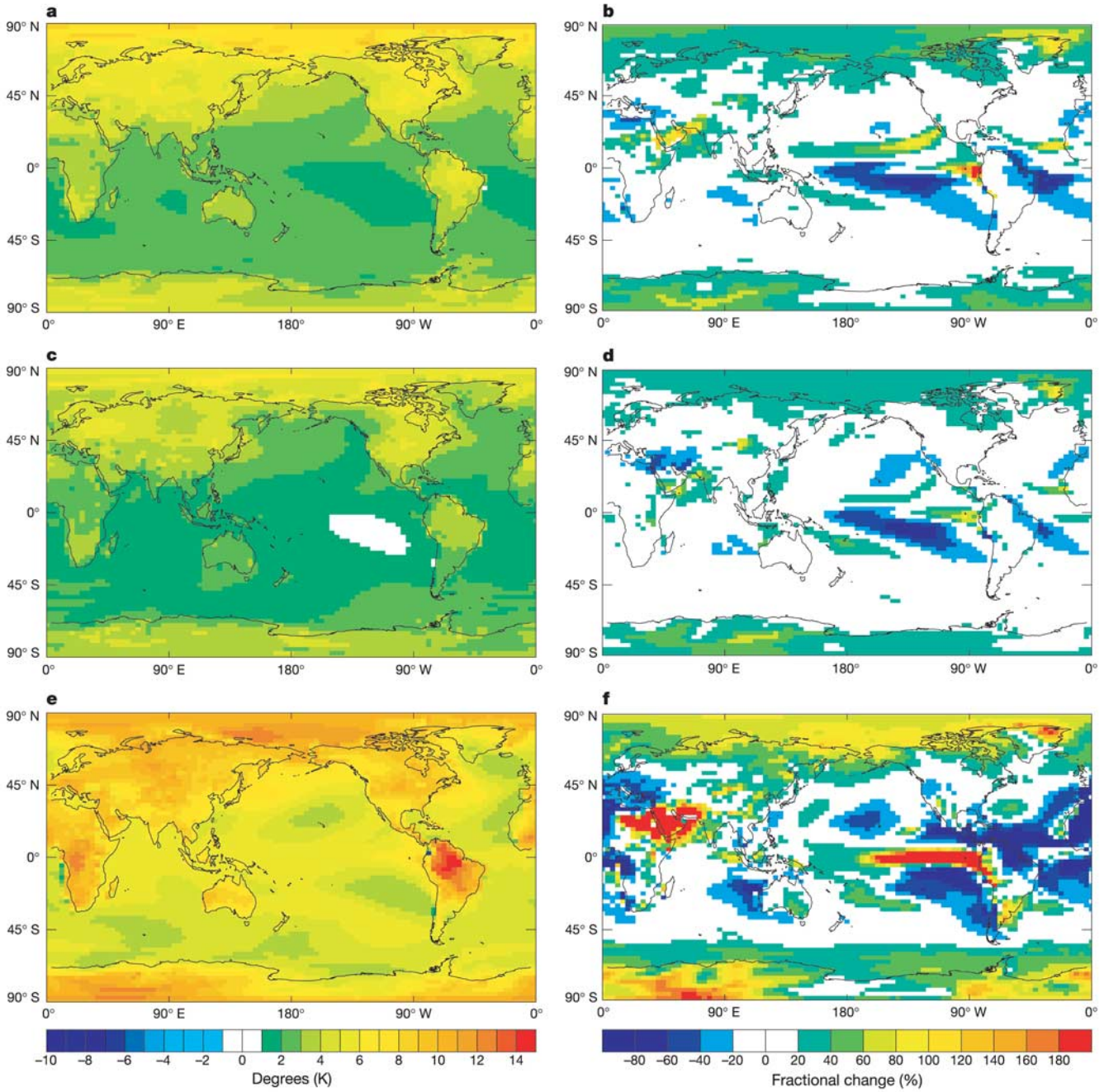

Figure 3 The temperature (left panels) and precipitation (right panels) anomaly fields in response to doubling the CO_2 concentrations. **a**, **b**, The unperturbed model (simulated climate sensitivity, 3.4 K). **c**, **d**, A model version with low simulated climate sensitivity (2.5 K). **e**, **f**, A model version with high simulated climate sensitivity (10.5 K). These fields are the means of years eight to fifteen after the change of forcing is applied, averaged over initial-condition ensemble members; they are not the equilibrium response.

standing of the potential responses to increasing levels of greenhouse gases, regional and seasonal impacts, our models and internal variability. Future experiments will include a grand ensemble of transient simulations of the years 1950–2100 using a model with a fully dynamic ocean. □

Methods

Model simulations

Participants in the climate*prediction*.net experiment download an executable version of a full GCM. They are allocated a particular set of parameter perturbations and initial conditions enabling them to run one simulation: that is, one member of the grand ensemble. Their personal computer then carries out 45 years of simulation and returns results to the project's servers. Over 90,000 participants from more than 140 countries have registered to date. The model, based on HadSM3[30], is a climate resolution version of the Met Office Unified Model with the usual horizontal grid of 3.75° longitude × 2.5° latitude and 19 layers in the vertical. The ocean consists of a single thermodynamic layer with ocean heat transport prescribed using a heat-flux convergence field that varies with position and season but has no inter-annual variability. For each simulation the heat-flux convergence field is calculated in the calibration phase where sea surface temperatures (SSTs) are fixed; in subsequent phases the SSTs vary according to changes in the atmosphere–ocean heat flux. The initial-condition ensemble members have different starting conditions for the calibration and therefore allow for uncertainty in the heat-flux convergence fields used in the control and doubled-CO_2 phases.

Data quality

Most model simulations are unique members of the grand ensemble, each being a combination of perturbed model parameters and perturbed initial conditions. To evaluate the reliability of the experimental design a certain number of identical simulations are distributed; most give identical results. Where they do not, they are usually very similar, suggesting that a few computational bits were lost at some point and consequently they are essentially different members of the initial-condition ensemble. In these cases the mean of the simulations is taken.

There are a small number of simulations (1.6%) which show obvious flaws in the data: for example, sudden jumps of data values from of the order of 10^2 to of the order of 10^8. These probably result from loss of information, for instance during a PC shut-down at a critical point in processing or a result of machine 'overclocking'. These are removed from this analysis. Finally, runs that show a drift in T_g greater than 0.02 K yr^{-1} in the last eight years of the control are judged to be unstable and are also removed from this analysis.

Perturbations

Perturbations are made to six parameters, chosen to affect the representation of clouds and precipitation: the threshold of relative humidity for cloud formation, the cloud-to-rain conversion threshold, the cloud-to-rain conversion rate, the ice fall speed, the cloud fraction at saturation and the convection entrainment rate coefficient. This is a subset of those explored by ref. 9. In each model version each parameter takes one of three values (the same values as those used by ref. 9); for cloud fraction at saturation only the standard and intermediate values are used. As climate*prediction*.net continues, the experiment is exploring 21 parameters covering a wider range of processes and values.

Climate sensitivity calculations

The simulated climate sensitivity is taken as the difference between the predicted equilibrium T_g in the doubled-CO_2 and control phases. The latter is simply the mean of the last eight years of that phase. The former is deduced by fitting the change in T_g, relative to the start of the phase, to the exponential expression: $\Delta T_g(t) = \Delta T_{g\,(2\times CO2)}(1 - \exp(-t/\tau))$, giving us a value of $T_{g\,(2\times CO2)}$ that allows for uncertainty in the response timescale, τ. Even for high simulated climate sensitivities the uncertainty in this procedure is small (see Fig. 2c) and alternative methods give similar results. Because it is based on the first 15 years' response, the λ associated with this simulated climate sensitivity reflects the decadal timescale feedbacks in the system. Longer, centennial-timescale processes could affect the ultimate value of the equilibrium sensitivity and are best studied using models with dynamic oceans and cryospheres.

Relative root-mean-square error

Models are compared with gridded observations of annual mean temperature, sea level pressure, precipitation and atmosphere–ocean sensible and latent heat flux. The total error in variable j is defined simply as:

$$\varepsilon_{js}^2 = (\Sigma_i w_i (m_{is} - o_i)^2 \quad (1)$$

where m_{is} is the simulated value in grid-box i averaged over the last 8 yr of the control phase of simulation s, o_i is the observed value[9] and w_i is an area weighting. Mean squared errors relative to the standard model are computed as:

$$\varepsilon_s^2 = (\Sigma_j \varepsilon_{js}^2 / \varepsilon_{ju}^2)/N \quad (2)$$

where N is the number of variables and ε_{ju}^2 is the mean ε_{js}^2 for the unperturbed model, and averaged across initial-condition ensembles. Normalizing errors in individual variables by the corresponding errors in the unperturbed model ensures that all variables are given equal weight. The relative r.m.s.e. is plotted in Fig. 2b. Note that because we do not have an explicit and adequate noise model (ε_{js}^2 does not account for correlations, for example), these 'scores' cannot be interpreted explicitly in terms of likelihood, but nevertheless provide an indication of the relative merits of different model control climates.

For the CMIP II data the $(m_i - o_i)^2$ term is reduced by the variance of the mean to compensate for the greater variability found in models with dynamic oceans.

Received 4 November; accepted 20 December 2004; doi:10.1038/nature03301.

1. Schneider, S. H. What is dangerous climate change? *Nature* **411,** 17–19 (2001).
2. Reilly, J. *et al.* Uncertainty in climate change assessments. *Science* **293,** 430–433 (2001).
3. Wigley, T. M. L. & Raper, S. L. Interpretation of high projections for global mean warming. *Science* **293,** 451–454 (2001).
4. Allen, M. R. & Stainforth, D. A. Towards objective probabilistic climate forecasting. *Nature* **419,** 228 (2002).
5. Allen, M. R. & Ingram, W. J. Constraints on future changes in climate and the hydrological cycle. *Nature* **419,** 224–232 (2002).
6. Palmer, T. N. Predicting uncertainty in forecasts of weather and climate. *Rep. Prog. Phys.* **63,** 71–116 (2000).
7. Smith, L. What might we learn from climate forecasts? *Proc. Natl Acad. Sci. USA* **99,** 2487–2492 (2002).
8. Kennedy, M. C. & O'Hagan, A. Bayesian calibration of computer models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **63,** 425–450 (2001).
9. Murphy, J. M. *et al.* Quantifying uncertainties in climate change from a large ensemble of general circulation model predictions. *Nature* **430,** 768–772 (2004).
10. Knutti, R., Stocker, T. F., Joos, F. & Plattner, G. K. Constraints on radiative forcing and future climate change from observations and climate model ensembles. *Nature* **416,** 719–723 (2002).
11. Forest, C. E. *et al.* Quantifying uncertainties in climate system properties with the use of recent climate observations. *Science* **295,** 113–117 (2002).
12. Allen, M. R., Stott, P. A., Mitchell, J. F. B., Schnur, R. & Delworth, T. L. Quantifying the uncertainty in forecasts of anthropogenic climate change. *Nature* **417,** 617–620 (2000).
13. Stott, P. A. & Kettleborough, J. A. Origins and estimates of uncertainty in predictions of twenty-first century temperature rise. *Nature* **416,** 723–726 (2002).
14. Cubasch, U. *et al.* in *Climate Change 2001, The Science of Climate Change* (ed. Houghton, J. T.) Ch. 9 527–582 (Cambridge Univ. Press, Cambridge, 2001).
15. Allen, M. R. Do-it-yourself climate prediction. *Nature* **401,** 627 (1999).
16. Stainforth, D. *et al.* Distributed computing for public-interest climate modeling research. *Comput. Sci. Eng.* **4,** 82–89 (2002).
17. Hansen, J. A. *et al.* Casino-21: Climate simulation of the 21st century. *World Res. Rev.* **13,** 187–189 (2001).
18. Stainforth, D. *et al.* Climate*prediction*.net: Design principles for public-resource modeling. In *14th IASTED Int. Conf. on Parallel And Distributed Computing And Systems (4–6 November 2002)* 32–38 (ACTA Press, Calgary, 2002).
19. Stainforth, D. *et al.* in *Environmental Online Communication* (ed. Scharl, A.) Ch. 12 (Springer, London, 2004).
20. Giorgi, F. & Francisco, R. Evaluating uncertainties in the prediction of regional climate change. *Geophys. Res. Lett.* **27,** 1295–1298 (2000).
21. Palmer, T. N. & Raisanen, J. Quantifying the risk of extreme seasonal precipitation events in a changing climate. *Nature* **415,** 512–517 (2002).
22. Collins, M. & Allen, M. R. Assessing the relative roles of initial and boundary conditions in interannual to decadal climate predictability. *J. Clim.* **15,** 3104–3109 (2002).
23. Pope, V. D., Gallani, M., Rowntree, P. R. & Stratton, R. A. The impact of new physical parameterisations in the Hadley Centre climate model – HadAM3. *Clim. Dyn.* **16,** 123–146 (2000).
24. Gregory, J. M. *et al.* An observationally-based estimate of the climate sensitivity. *J. Clim.* **15,** 3117–3121 (2002).
25. Andronova, N. G. & Schlesinger, M. E. Objective estimation of the probability density function for climate sensitivity. *J. Geophys. Res.* **106,** 22605–22612 (2001).
26. Morgan, M. G. & Keith, D. W. Climate-change — subjective judgments by climate experts. *Environ. Sci. Technol.* **29,** 468–476 (1995).
27. Smith, L. A. in *Disentangling Uncertainty and Error: On the Predictability of Nonlinear Systems* (ed. Mees, A. I.) Ch. 2 (Birkhauser, Boston, 2000).
28. Covey, C. *et al.* An overview of results from the coupled model intercomparison project. *Glob. Planet. Change* **37**(1–2), 103–133 (2003).
29. McAveney, B. J. *et al.* in *Climate Change 2001, The Science of Climate Change* (ed. Houghton, J. T.) Ch. 8, 471–524 (Cambridge Univ. Press, Cambridge, 2001).
30. Williams, K. D., Senior, C. A. & Mitchell, J. F. B. Transient climate change in the Hadley centre models: the role of physical processes. *J. Clim.* **14,** 2659–2674 (2001).

Supplementary Information accompanies the paper on **www.nature.com/nature**.

Acknowledgements We thank all participants in the 'climate*prediction*.net' experiment and the many individuals who have given their time to make the project a reality and a success. This work was supported by the Natural Environment Research Council's COAPEC, e-Science and fellowship programmes, the UK Department of Trade and Industry, the UK Department of the Environment, Food and Rural Affairs, and the US National Oceanic and Atmospheric Administration. We also thank Tessella Support Services plc, Research Systems Inc., Numerical Algorithms Group Ltd, Risk Management Solutions Inc. and the CMIP II modelling groups.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.A.S. (d.stainforth1@physics.ox.ac.uk).