www.sciencemag.org/cgi/content/full/317/5839/796/DC1

Supporting Online Material for

Improved Surface Temperature Prediction for the Coming Decade from a Global Climate Model

Doug M. Smith,* Stephen Cusack, Andrew W. Colman, Chris K. Folland, Glen R. Harris, James M. Murphy

*To whom correspondence should be addressed. E-mail: doug.smith@metoffice.gov.uk

Published 10 August 2007, *Science* **317**, 796 (2007)
DOI: 10.1126/science.1139540

This PDF file includes:

Supporting Online Material

1 Materials and Methods

1.1 Details of experimental design

Fig. S1 summarises the components used to create the decadal hindcasts assessed in this paper. The starting point was a 1300 year control simulation of HadCM3 (*S1*) (blue box in Fig. S1) in which there were no interannual variations in external forcing. From this, four simulations (*S2*) of the transient response to historical changes in external forcing (from greenhouse gases (*S3*), tropospheric and stratospheric ozone concentration, sulphur emissions, solar irradiance and volcanic aerosol) from 1860 to 2001 (green boxes) were started from initial states taken from the control simulation at 100 year intervals. NoAssim hindcast integrations (blue lines) were then initialised from each of the four transient simulations each season from 1982 to 2001, thus creating a set of 80 four-member ensembles. DePreSys hindcasts (red lines in Fig. S1 below) were initialised by assimilating atmosphere and ocean observations into one of the transient integrations (further details are given in section 1.2). Four ensemble members were initialised from consecutive days preceding and including each hindcast start date. For both NoAssim and DePreSys hindcasts (and forecasts), solar flux is specified by repeating the previous 11-year solar cycle, and volcanic aerosol is reduced exponentially with an e-folding timescale of one year (*S4*). Anthropogenic forcing is taken from the intermediate IPCC SRES B2 scenario (*S5*). We expect the uncertainty of future anthropogenic emissions to have a negligible impact on our forecasts, since projections of global T_S for the coming few decades are insensitive to a range of plausible emission scenarios (*S6*).

1.2 Details of DePreSys initialisation

Initial conditions for forecasts and hindcasts are generated by integrating HadCM3 (*S1*), relaxing oceanic temperature and salinity (T,S), and atmospheric horizontal winds (u,v), temperature

(θ) and surface pressure (p^*), to analyses of observations (Fig. S1). The atmospheric analyses are taken from the European Centre for Medium-Range Weather Forecasts (ECMWF) reanalyses (*S7*) (ERA15 from 1979-1993, ERA40 from 1994-2001, operational analysis from 2002 onwards). The ocean analyses (*S8*) are created by four-dimensional, multivariate optimal interpolation of salinity and sub-surface temperature observations and analysed sea surface temperature (SST) from HadISST (*S9*), using covariances from HadCM3. The ocean analyses are full-depth, but will be close to climatology in the deeper ocean where data are particularly sparse. Model errors cause forecasts to drift away from the observed state towards the imperfect model climate. In seasonal prediction this bias is usually removed by an *a posteriori* empirical correction computed from a series of hindcasts (*S10*). This strategy is less attractive for decadal prediction, since the smaller magnitude of the predictable signal is more likely to be masked by inaccuracies in the calculation of the bias. We therefore initialise our model with observed anomalies added to the model climate, and remove the model climate to obtain forecast anomalies. The model climate is obtained from four simulations of the twentieth century (*S2*) which include anthropogenic and natural external forcing but do not assimilate observations. The climatological period is 1941-1996 for the ocean (*S8*), and 1979-1993 (1987-2001) in the atmosphere for the period before (after) 1993 (see Supporting Text, section 2.2).

1.3 Uncertainty of differences in RMSE and bias

For the time period covered by our hindcasts, we define the true RMSE and bias for either of our forecasting systems as the value which would have been obtained from ensemble hindcasts of infinite size, run from all possible initialisation dates within the period. For a given forecast lead time, we possess estimates of the true RMSE and bias from a sample of hindcast errors, $e_i = h_i - o_i$, at N_T validation times $i = 1 \rightarrow N_T$. Here, o_i is the observed value, and h_i is the ensemble mean hindcast from N_E individual members. We wish to test the null hypothesis that differences between NoAssim and DePreSys RMSE and bias could be explained by sam-

pling errors arising from a finite N_T and N_E. The true mean difference between NoAssim and DePreSys errors can be written as $\langle d_{i,\infty} \rangle + \delta$, where $d_{i,\infty}$ is the difference between theoretical NoAssim and DePreSys hindcasts containing an infinite number of ensemble members, $\langle\rangle$ represents the mean over N_T validation points, and δ is the sampling error arising from a finite N_T. Furthermore, $d_i = d_{i,\infty} + \epsilon_i$, where d_i is the difference between NoAssim and DePreSys ensemble mean hindcasts containing N_E ensemble members, and ϵ_i is the sampling error of d_i arising from a finite N_E. We estimate uncertainties in differences between NoAssim and DePreSys RMSE and bias arising from δ and ϵ_i by Monte Carlo simulation as follows.

1. Sample uncertainties arising from a finite N_E by computing N_M ($= 5000$) Monte Carlo NoAssim hindcast errors $e_{i,m} = h_{i,m} - o_i$ ($m = 1 \rightarrow N_M$) at each validation time. Here, $h_{i,m} = h_i^{NA} + h'_{i,m}$, where h_i^{NA} is the NoAssim ensemble mean, and the Monte Carlo perturbation $h'_{i,m}$ samples the uncertainty in the difference between the NoAssim and DePreSys ensemble means arising from a finite N_E. Hence, $var(h'_{i,m}) = var(\epsilon_i) = \langle var(h_i^{NA})/N_{E,NA}^{eff} \rangle + \langle var(h_i^{DPS})/N_{E,DPS}^{eff} \rangle$, where $var()$ denotes the variance, $\langle\rangle$ represents the mean over N_T validation points, and the effective degrees of freedom (*S11*) is given by $N_E^{eff} = N_E(1-\alpha)/(1+\alpha)$, where α is the centred first serial autocorrelation of differences between individual ensemble members and the ensemble mean (see below). For unbiased estimates, $\langle e_{i,m} \rangle = \langle e_i \rangle$ and $\langle e_{i,m}^2 \rangle = \langle e_i^2 \rangle$, where e_i is the error of the NoAssim ensemble mean. It is easily shown that this requires $\langle h'_{i,m} \rangle = 0$ and $\langle e_i h'_{i,m} \rangle = \langle e_i^2 \rangle$. This is achieved by taking $h'_{i,m} = a e'_i + \varepsilon_{i,m}$, where $a = -\langle {h'_{i,m}}^2 \rangle / (2\langle {e'_i}^2 \rangle)$, $e'_i = e_i - \langle e_i \rangle$, and $\varepsilon_{i,m}$ is Gaussian white noise with variance $\langle {h'_{i,m}}^2 \rangle (1 - \langle {h'_{i,m}}^2 \rangle / (4\langle {e'_i}^2 \rangle))$.

2. Sample additional uncertainties arising from a finite N_T by computing Monte Carlo NoAssim errors $\hat{e}_{i,m} = e_{i,m} + \delta_m$ ($i = 1 \rightarrow N_T$ and $m = 1 \rightarrow N_M$). Here, δ_m is Gaussian white noise with standard deviation equal to the standard error of the sample mean of N_T differences between theoretical NoAssim and DePreSys hindcasts containing an infinite number of ensemble members. Hence, $var(\delta_m) = var(d_{i,\infty})/N_T^{eff}$, where

the effective degrees of freedom (*S11*) is given by $N_T^{eff} = N_T(1-\rho)/(1+\rho)$, and ρ is the centred first serial autocorrelation of $d_{i,\infty}$. We write $var(d_{i,\infty}) = f \times var(d_i)$, where $f = var(d_{i,\infty})/var(d_i)$, and use the Monte Carlo simulations of step 1 to estimate f as $var(\bar{d}_{i,m})/\overline{var(d_{i,m})}$. Here the overbar denotes the mean of the N_M Monte Carlo simulations computed in step 1, and $d_{i,m} = e_{i,m} - e_i^{DPS}$, where e_i^{DPS} is the error of the DePreSys ensemble mean hindcasts. For global annual mean T_S and H, we find that f is typically in the range 0.4 to 0.7. We compute ρ from $\bar{d}_{i,m}$, and obtain values typically between 0.6 and 0.8 for rolling annual mean T_S and H.

3. Compute the 5-95% confidence limits of the distributions of RMSE and bias computed from $\hat{e}_{i,m}$ ($i = 1 \rightarrow N_T$ and $m = 1 \rightarrow N_M$).

The NoAssim hindcasts are initialised from different long-running integrations which sample different realisations of low frequency internal variability. Because of this, the difference between an individual ensemble member and the ensemble mean can be significantly correlated for hindcasts started a season apart. However, most ensemble systems (including DePreSys) initialise the same observed low frequency internal variability into each member, with individual members differing only at higher frequencies based on perturbations which are typically uncorrelated between forecasts started a season or more apart. In this case, the autocorrelation (α) in step 1 would be expected to be zero.

Note that the blue shading in Fig. 1 (main paper) represents the 5-95% region in which *differences* between NoAssim and DePreSys are not significant. This is not the same as the 5-95% uncertainty of the NoAssim hindcasts, which would be achieved by taking $var(h'_{i,m}) = \langle var(h_i^{NA})/N_{E,NA}^{eff} \rangle$ in step 1, and replacing $d_{i,\infty}$ in step 2 by the mean of an infinite NoAssim ensemble.

1.4 Statistical forecasts

Forecasts of global T_S for the coming year are made by multiple linear regression using the following predictors. Interannual variability is captured by the El Niño Southern Oscillation (ENSO) (measured by the first two ENSO-related empirical orthogonal function (EOF) patterns of high pass filtered SST (*S12*) for 1911-1995) and volcanic activity (*S4*) at the end of the previous year. Decadal and longer time-scale variability resulting from human activities is predicted by historical radiative forcing from greenhouse gases, sulphate aerosol and ozone as used in recent Hadley Centre climate predictions (*S3*). The multi-decadal influence of the oceanic thermohaline circulation (*S13*) is included by a surrogate index, based on EOF2 of low pass filtered global SST (*S12*) for 1911-1995. This reflects an interhemispheric contrast in SST anomalies with emphasis on the Northern Hemisphere. Solar variability is included by an index of solar irradiance (*S14*).

2 Supporting Text

2.1 Removal of bias in initial conditions

Model errors cause forecasts to drift away from the observed state towards the imperfect model climate. DePreSys avoids this drift by assimilating observed anomalies added to the model climate (see Methods). By construction, there is no mean difference between DePreSys and NoAssim initial conditions when averaged over the whole period used to define the model and observed climatologies. However, differences in initial conditions can occur over a subset of the climatological period, or for different periods. For the ocean, the climatological period was 1941 to 1996. However, during our hindcast period (1982 to 2001), the initial upper ocean heat content (H) anomalies for the NoAssim hindcasts are consistently warmer than observed (Fig. 1c, main paper). By assimilating the observed anomalies of H, DePreSys removes this bias in initial conditions, thereby producing more skilful hindcasts.

We further investigate whether the differences in bias (and hence rmse) between DePreSys and NoAssim are caused by the initialisation of heat content anomalies by using an energy balance model (EBM) (*S15*). The EBM represents the globally-averaged vertical transport of heat in the ocean by diffusive mixing and a constant upwelling advection (with velocity 3.1 m/yr). We use a background diffusivity profile (*S1*), but increase the diffusivity by a factor of 10 within a 50m mixed layer at the surface. We initialise the EBM ocean with the mean difference between DePreSys and NoAssim initial conditions for our hindcast period (Fig. S2), and compare the evolution of ocean heat content in the upper 113m (H) computed by the EBM with the mean difference between DePreSys and NoAssim H as a function of forecast period (Fig. S3). The response of the atmosphere-ocean heat flux to a surface temperature anomaly ΔT is represented, as in other EBMs, by a standard term of the form $-\lambda\Delta T$, where λ (assumed constant) is a parameter representing the strength of surface and atmospheric feedbacks (*S15*). If we follow this approach we obtain the dotted curve in Fig. S3. This curve does show evidence of a long term influence on the hindcasts arising from the initial differences in heat content. However it also shows a steady diminution of the difference between DePreSys and NoAssim, whereas the actual DePreSys minus NoAssim H (solid curve in Fig. S3) increases for a few years before starting to reduce at longer lead times. We further investigate this by replacing the $-\lambda\Delta T$ term in the EBM with the time series of the mean DePreSys minus NoAssim surface flux (solid curve in Fig. S4). This yields the dashed curve in Fig. S3, which is a much better representation of the actual DePreSys minus NoAssim H. The cooling in DePreSys relative to NoAssim is therefore enhanced by a positive feedback in the atmosphere-ocean flux. During the first four years of the hindcasts (Fig. S4) this is driven by a reduction in surface short-wave (SW) heating, which is consistent with the positive SW cloud feedback found in experiments with the slab configuration of this model when tested using standard $\pm 2K$ SST perturbation experiments (this model is one of those giving positive feedback in Fig. 1 of (*S16*)). In summary, the EBM study confirms that the improvement in DePreSys over NoAssim bias (and hence RMSE) directly

results from initialisation of ocean heat content anomalies, however the detailed evolution of heat content is influenced by time-dependent atmospheric feedbacks, which amplify the heat content differences for forecast lead times up to four years ahead, instead of opposing them as would be expected from simple energy balance arguments.

2.2 Climatological period

The choice of climatological period in the atmosphere was limited by the availability of analyses of observations from which to make the observed means. Since our first set of hindcasts was based on the ECMWF 15-year reanalysis, we were initially constrained to use the period covered by this reanalysis (1979-1993). When the 40-year ECMWF reanalysis became available we extended our hindcasts to 2001. We retained a 15-year climatological period in the atmosphere, but used the period 1987-2001 in order to minimise the amount of data that needed to be extracted from the ECMWF archive. We would not expect the change of base period to significantly affect our hindcasts provided the difference between observed and model climatologies is similar for both periods, and indeed there is no obvious discontinuity at the switchover (Fig. 2, main paper). In future we would ideally use the same climatological period for the atmosphere and ocean, since any inconsistencies in observed atmosphere and ocean anomalies are likely to degrade the DePreSys forecasts. However, we expect this to have a minor impact, since the component of predictability on interannual to decadal timescales associated with internal variability is likely to be dominated by initial conditions in the ocean.

2.3 Confidence limits

The hindcast time-series illustrate the important ability to provide information on uncertainty by diagnosing confidence limits from the ensemble spread (red shading in Figs. 2 and 4, main paper). For example, although interannual variability (such as the warming following the El Niño of 1997) is not predicted by the ensemble mean in year nine of the hindcasts (Fig. 2b, main paper), the possibility of such events is generally captured by the 90% confidence limits

of the DePreSys hindcasts. Furthermore, the mean 5-95% confidence interval (I_{95}) is smaller for the coming year (main paper Fig. 2a, $I_{95} = 0.10°\mathrm{C}$) than for year 9 (Fig. 2b, $I_{95} = 0.18°\mathrm{C}$), consistent with an ability to predict some of the interannual variability at shorter lead times. We assess the reliability of the DePreSys confidence intervals by computing the fraction of time that the observations lie within a given confidence range. In general, the confidence limits capture the observations reasonably well (Fig. S5). However, the hindcast confidence overestimates the fraction of time that the observations lie within a given confidence range by 8% (averaged over all ranges) for both years one and nine, indicating a need to increase the ensemble spread by improving the sampling of uncertainty. Our use of initial conditions from consecutive days only samples the growth of small imperfections in the initial conditions, and could potentially be improved by adding uncertainty commensurate with the expected errors of the ocean analyses. Furthermore, uncertainties in the model formulation could be sampled either by using additional models (*S17*) or by perturbing poorly constrained parameters within the existing model (*S18*, *S19*).

2.4 Removal of bias from DePreSys forecast

We expect both NoAssim and DePreSys hindcasts of global surface temperature (T_S) to be too warm on average, especially at longer lead times, because they do not capture the cooling resulting from future volcanic eruptions. However, measured only over those hindcast cases in which there were no future volcanic eruptions in reality, we find that the NoAssim hindcasts are biased significantly warm at forecast lead times greater than two years (see Fig. S6). This bias is essentially absent in the DePreSys hindcasts for lead times up to six years, suggesting that it is caused initially by a warm bias in upper ocean heat content (see main paper). However, DePreSys does develop a warm bias in T_S beyond year six, rising to $0.07°\mathrm{C}$ by year nine. In principle, this bias may result from incorrect estimates of the applied external forcing, or from the impact of model errors on the simulated transient response to the forcing. In practice,

uncertainties in forcing seem unlikely to be the main cause (*S6*). Uncertainties in the transient response of T_S are likely to be influenced by uncertainties in the efficiency of global ocean heat uptake and climate sensitivity (*S20*). These variables are influenced by a number of different feedbacks (*S21*), each of which are themselves uncertain due to the influence of complex sub-grid scale processes which can only be represented approximately in models. For example, a major source of inter-model variation in climate sensitivity arises from changes in the radiative properties of clouds (*S22*). Hence it is not yet possible to attribute errors in predictions of T_S to a specific cause. However, we use our estimate of the bias from DePreSys hindcasts (Fig. S6) to adjust the DePreSys forecast of T_S (Fig. 4, main paper). The impact of this empirical bias correction is substantially smaller than the initial condition uncertainty captured by the ensemble spread (Fig. S7). However, our correction relies on the assumption that the bias is independent of variations in the initial conditions and the applied forcing. In future, we will attempt to account explicitly for uncertainty in the simulation of climate system processes by designing ensemble predictions which sample variations in model formulation (*S17, S18, S19*) as well as variations in initial conditions.

2.5 Initial assessment of DePreSys and NoAssim forecasts

An initial evaluation of the first 21 months of the DePreSys and NoAssim forecasts which started in June 2005 is provided in Fig. S8 in terms of the depth-weighted average ocean temperature for the upper 113m. In general, the DePreSys errors (I) are considerably smaller than the NoAssim errors (F). The difference between DePreSys and NoAssim (H) is highly anti-correlated (correlation R=-0.86) with the NoAssim errors (F) indicating that DePreSys is acting to correct the NoAssim errors almost everywhere. Some of this improvement arises through persistence of the initial heat content anomalies present at the start of the forecast. For example, the observed pattern of warm anomalies in the North Atlantic (A) is captured by both DePreSys (D) and a persistence forecast obtained by persisting the anomaly for the season preceding

the start of the forecast (B). Furthermore, successful persistence of the neutral to cool conditions in the Southern Ocean (A,B) by DePreSys (D) accounts for a large part of the cooling in DePreSys relative to NoAssim (H). Although DePreSys develops larger errors than persistence in the mid-latitude North Atlantic (I), many of the errors of persistence (E) elsewhere are corrected by DePreSys (G). In particular, DePreSys successfully predicted the development of cool conditions in the eastern tropical Pacific, consistent with an ability of climate models to predict dynamical changes associated with El Niño. The successful prediction of this change accounts for much of the remainder of the cooling in DePreSys relative to NoAssim (H).

References

S1. C. Gordon, *et al.*, *Clim. Dyn.* **16**, 147 (2000).

S2. P. A. Stott, *et al.*, *Science* **290**, 2133 (2000).

S3. T. C. Johns, *et al.*, *Clim. Dyn.* **20**, 583 (2003).

S4. M. Sato, J. E. Hansen, M. P. McCormick, J. B. Pollack, *J. Geophys. Res.* **98**, 22,987 (1993).

S5. N. Nakićenović, *et al.*, *Special report on emissions scenarios* (Cambridge University Press, 2000). 599pp.

S6. P. A. Stott, J. A. Kettleborough, *Nature* **416**, 723 (2002).

S7. J. K. Gibson, *et al.*, ERA description, *ECMWF Re-anlysis Project Report Series 1*, European Centre for Medium Range Weather Forecasts (1997).

S8. D. M. Smith, J. M. Murphy, *J. Geophys. Res.* **112**, C02022 (2007).

S9. N. A. Rayner, *et al.*, *J. Geophys. Res.* **108**, 10.1029/2002DJ002670 (2003).

S10. T. N. Stockdale, *Mon. Weath. Rev.* **25**, 809 (1997).

S11. D. S. Wilks, *Statistical methods in the atmospheric sciences* (Academic Press, San Diego, CA, 1995).

S12. C. K. Folland, D. E. Parker, A. Colman, R. Washington, *Large scale modes of ocean surface temperature since the late nineteenth century* (Springer-Verlag, 1999), chap. 4, pp. 73–102.

S13. J. R. Knight, R. J. Allan, C. K. Folland, M. Vellinga, M. E. Mann, *Geophys. Res. Lett.* **32**, L20708, doi:10.1029/2005GL024233 (2005).

S14. C. Frohlich, J. Lean, *Geophys. Res. Lett.* **25**, 4377 (1998).

S15. G. R. Harris, *et al.*, *Clim. Dyn.* **27**, 357 (2006).

S16. M. A. Ringer, *et al.*, *Geophys. Res. Lett.* **33**, L07718 (2006).

S17. T. N. Palmer, *et al.*, *Bull. Am. Meteorol. Soc.* **85**, 853 (2004).

S18. J. M. Murphy, *et al.*, *Nature* **430**, 768 (2004).

S19. D. A. Stainforth, *et al.*, *Nature* **433**, 403 (2005).

S20. S. C. B. Raper, J. M. Gregory, R. J. Stouffer, *J. Climate* **15**, 124 (2002).

S21. S. Bony, *et al.*, *J. Climate* **19**, 3445 (2006).

S22. B. J. Soden, I. M. Held, *J. Climate* **19**, 33543360 (2006).

S23. C. K. Folland, *et al.*, *Geophys. Res. Lett.* **28**, 2621 (2001).

S24. D. E. Parker, L. V. Alexander, J. Kennedy, *Weather* **59**, 145 (2004).

S25. P. D. Jones, A. Moberg, *J. Climate* **16**, 206 (2003).

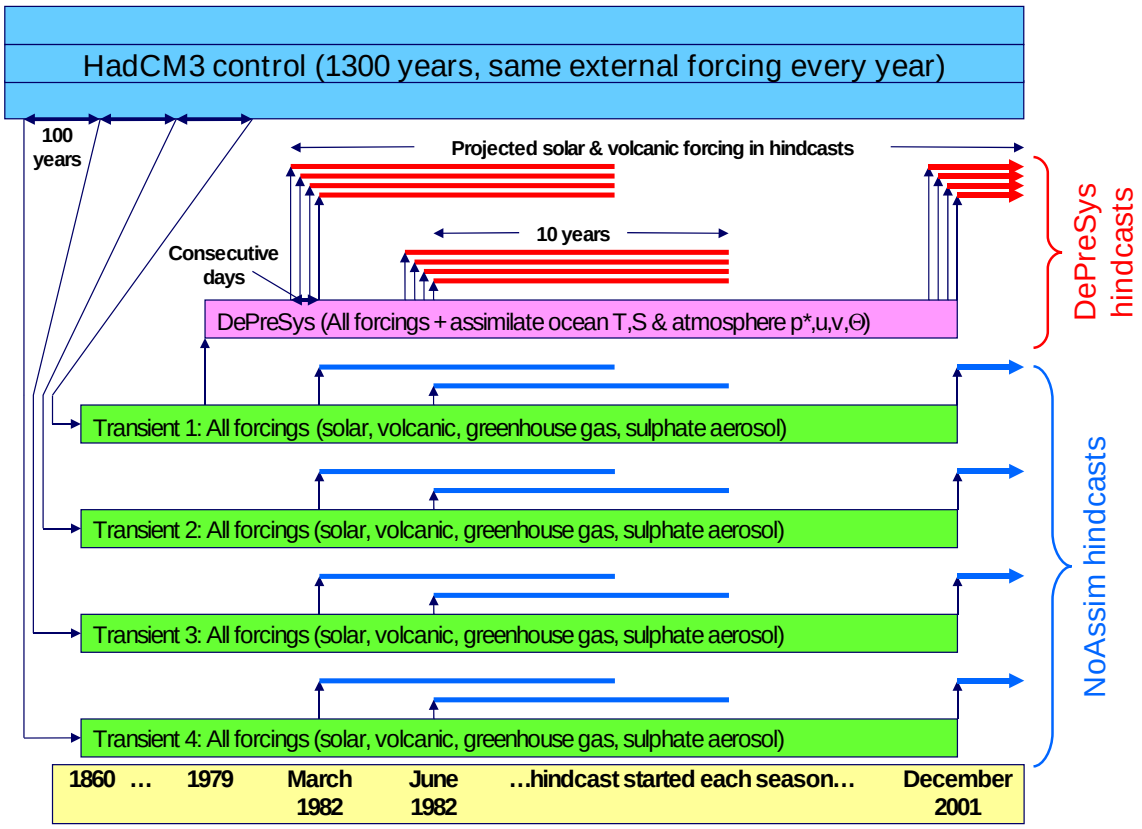

Figure S1: Schematic showing experimental method used to create NoAssim and DePreSys hindcasts.

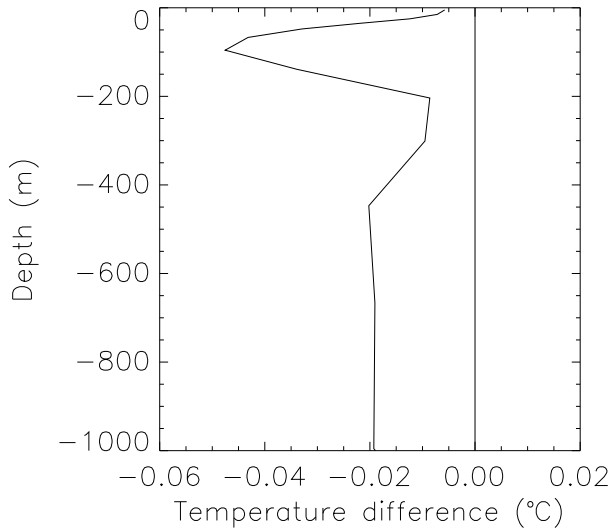

Figure S2: Difference between DePreSys and NoAssim initial conditions of globally-averaged ocean temperature for the upper 1000m averaged over our hindcast period (1982 to 2001).

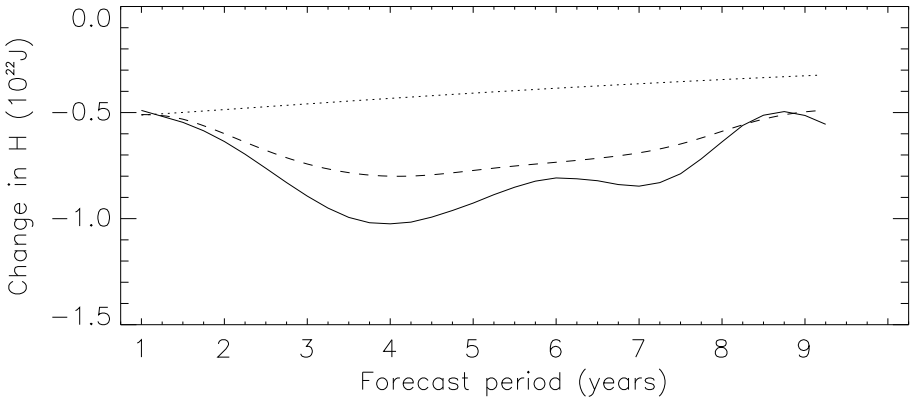

Figure S3: Comparison of the mean DePreSys minus NoAssim global annual-mean ocean heat content in upper 113m (H) averaged over all hindcasts (solid curve) with the evolution of H computed by an EBM initialised with the mean difference between DePreSys and NoAssim ocean initial conditions (see text for further details). The dotted curve shows the EBM evolution with a fixed climate feedback parameter ($\lambda = 1.12$), the dashed curve shows the EBM evolution with $\lambda = 0$ and the total surface flux difference between DePreSys and NoAssim (Fig. S4, solid curve) applied directly.

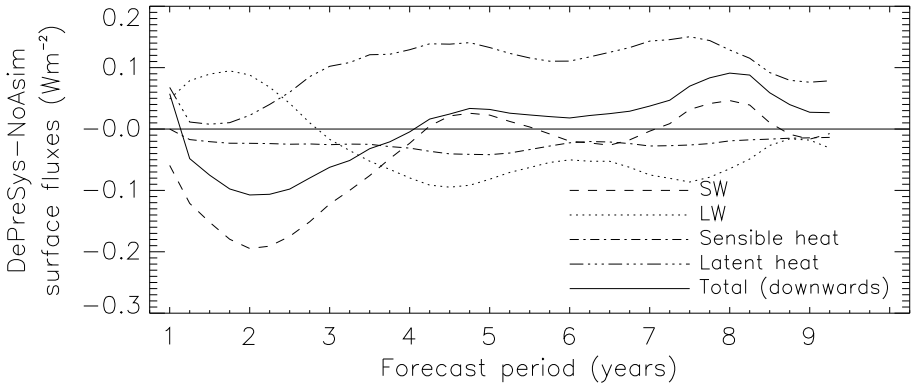

Figure S4: Difference between DePreSys and NoAssim global annual-mean surface fluxes (downwards) averaged over all hindcasts. SW=short-wave, LW=long-wave.

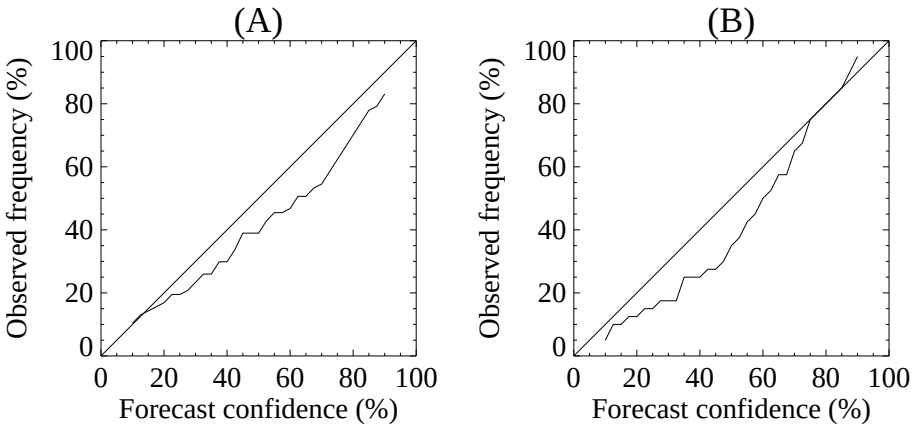

Figure S5: Reliability of the DePreSys forecast confidence limits of globally-averaged annual-mean surface temperature (T_S), assessed by computing the fraction of time that the observations lie within a given confidence range. (a) Hindcasts of the first annual mean, (b) Hindcasts of year nine (computed only for those verifying after 1993 in order to minimise the impact of Mt. Pinatubo).

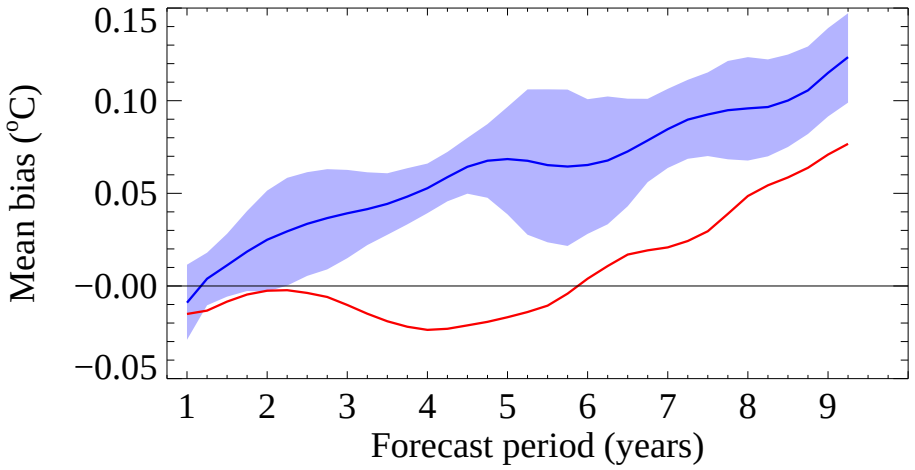

Figure S6: Bias of globally-averaged annual-mean surface temperature (T_S) anomalies (relative to 1979-2001) as a function of forecast period averaged only over those hindcast cases in which there were no future volcanic eruptions in reality. The solid red and blue curves show the bias of the ensemble mean DePreSys and NoAssim hindcasts, respectively, and the blue shading shows the 5-95% region where differences between DePreSys and NoAssim are not significant (see Methods). Observations are taken from the HadCRUT2vOA dataset (*S23, S24, S25*).

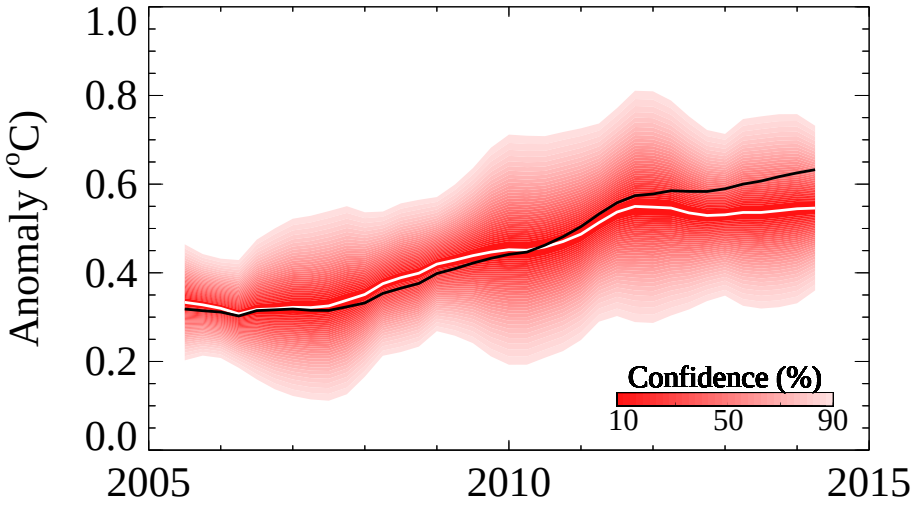

Figure S7: The impact of bias correction on the DePreSys forecast (Fig. 4, main paper) of globally-averaged annual-mean surface temperature anomaly (relative to 1979-2001). The black curve shows the uncorrected ensemble mean forecast. The white curve shows the ensemble mean forecast after removing the time-dependent DePreSys bias (Fig. S6). The confidence interval (red shading) is diagnosed from the standard deviation of the uncorrected DePreSys ensemble (see main text) and then applied to the corrected ensemble mean. Rolling annual mean values are plotted seasonally from March, June, September and December.

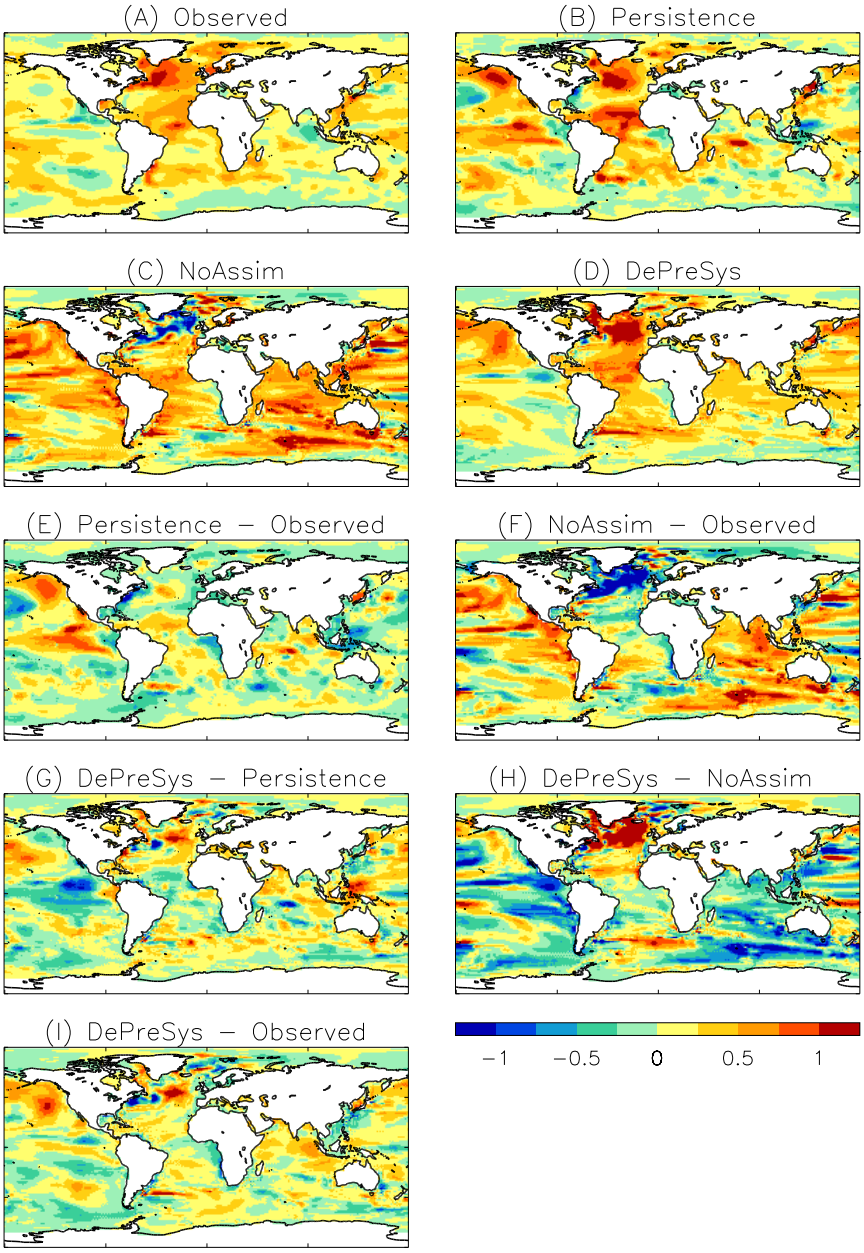

Figure S8: Assessment of the first 21 months of the DePreSys and NoAssim forecasts which started from June 2005. The maps show the depth-weighted average ocean temperature anomaly (relative to 1941-96) for the upper 113m (which is related to H) for the period June 2005 to February 2007. **(A)** Observed (*S8*). **(B)** Persistence of the seasonal mean anomaly averaged over March, April and May 2005. **(C)** NoAssim forecast **(D)** DePreSys forecast. **(E)** Error of persistence. **(F)** Error of NoAssim. **(G)** DePreSys minus persistence. **(H)** DePreSys minus NoAssim. **(I)** Error of DePreSys. Bias corrections derived from previous hindcasts were not applied to the DePreSys or NoAssim forecast values in these plots (cf Fig. 4 of the main text), as these would be small, and would not affect the differences between the DePreSys and NoAssim results.