Interannual Temperature Predictions using the IPCC multi-model ensemble mean

Thomas Laepple[*], Stephen Jewson[1], Katie Coughlin[1]

[1]Risk Management Solutions, London

Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany

E-mail: thomas.laepple@awi.de

Running title: Interannual predictions using the IPCC ensemble mean

We present a simple method to make multi-year surface temperature forecasts using the climate change simulations of the IPCC AR4 project. By calibrating the multi-model ensemble mean with current observations, we are able to make skillful interannual forecasts of mean temperatures. The method is validated using extensive hindcast experiments and is shown to perform favorably compared to a recently presented forecast method based on a global circulation model with assimilated initial conditions. Five year forecasts for the global mean temperature, the Northern Hemispheric mean temperature and the summer sea surface temperatures (SSTs) in the main development region for hurricanes (MDR) are presented.

1. Introduction

The latest report of the Intergovernmental Panel on Climate Change (IPCC) [Solomon et al., 2007] presented long-term projections of climate change into the next century. It was emphasized that most of the observed warming over the past 50 years is attributable to human activities and that the climate will likely continue to warm. Whereas the projections of the report are made on the century scale, industry and policy makers are often interested in a mid-term perspective of 1-10 years to plan their actions. Therefore there is also great interest in multi-year forecasts for the climate system.

Global seasonal-timescale climate predictions based on coupled ocean-atmosphere models are now operational in a large number of meteorological institutes [e.g. *Kanamitsu et al.*, 2002; *Mason et al.*, 1999] but interannual forecasts using these models are still in development [e.g.*Palmer et al.*, 2004]. Recently Smith et al. [2007] presented, for the first time, a mid-term, interannual global forecast which accounts for the effect of external forcing as well as internal variability. This system is based on a state of the art dynamical global climate model and takes into account the observed state of the atmosphere and ocean in order to predict the internal variability out to decadal time-scales. However, because this kind of forecast system is still developmental, the skill of the forecast needs to be weighed against the large technical and computing effort needed to implement such a system.

As an alternative, we present a very simple approach for interannual temperature forecasts using the output from the large ensemble of coupled ocean-atmosphere models which participated in experiments compiled by the Intergovernmental Panel on Climate Change (IPCC). By calibrating the model output with observed data, we use both the skill of the complex models in forecasting the anthropogenic contribution to changing temperatures and the skill of persistence, which is inherent in the temperature timeseries.

We demonstrate our prediction technique on three temperature indices: the annual global mean surface temperature (SAT) which exhibits very small interannual variability due to the large area mean, the Northern Hemispheric mean SAT, and the summer SST in the MDR region. SSTs in this Atlantic region exhibit very strong interannual to multi-decadal variability and are of special interest due to the possible connection to hurricane frequency and intensity [e.g. *Emanuel*, 2005; *Goldenberg et al.*, 2001]. Forecasts of these indices are given for a five-year outlook and the skill of interannual forecasts on 1, 5 and 9 year time-scales are compared to the skill of the Smith et al. [2007] predictions.

2. Data

We use the annual mean Land-Ocean Temperature anomaly Index for the Northern Hemispheric (NH) and Global mean Temperature (GL) provided by NASA GISS [*Hansen et al.*, 2006] (available at http://data.giss.nasa.gov/gistemp/). This dataset is a combination of land surface temperature and marine SST data from observations [*Rayner et al.*, 2003] and satellite data [*Reynolds et al.*, 2002]. The HADISST dataset [*Rayner et* *al.*, 2003] is used to extract the MDR SST index. We define this index as the July-September seasonal mean, averaged over the MDR region (15-70W,10-20N). To be consistent with the global anomaly time series we use an anomaly relative to 1951-1980. The three timeseries are shown in Figure 1a-c.

The model data consists of gridded global monthly surface air temperature and sea surface temperatures from the World Climate Research Programme's (WCRP's) Coupled Model Intercomparison Project phase 3 (CMIP3) multi-model dataset [Randal et al., 2007] (available at http://www-pcmdi.llnl.gov).

We extract mean temperatures over the seasons and regions which correspond to the observational data described above to create analogous time series for each IPCC run. The scenario 20C3M (models forced using observed concentrations of greenhouse gases/aerosols from 1850 to 2000) as well as the future IPCC-scenarios SRESA1B, SRESA2 and SRESB1 are used.

3. Forecast method

We divide the models into a set which includes historical volcanic and a set without volcanic forcing. As the volcanic forcing has a strong impact on the temperature timeseries especially on the MDR SST [*Santer et al.*, 2006], but is not predictable in the future, we only use the non-volcanic models to allow for fair hindcasts. A list of these models can be found in the Appendix of Santer et al. [2006].

The historical 20C3M simulations are merged with the future simulations which start in the year 2000. The merged simulations are then treated as continuous timeseries for the rest of the study. The models BCC-CM1 and MRI-CGCM2.3.2 as well as the SRESB1, CSIRO-Mk3.0 runs were removed from the set to avoid discontinuities in 2000 as they did not restart from the last year of the 20C3M run.

For the next decade, the differences in the forcing of the scenarios are small [*Zwiers*, 2002] so, in order to increase the size of our ensemble, we regard the scenarios as equally likely and have included runs from all three. By taking the mean over all the ensemble members of the models and over these three scenarios we are able to remove most of the internal variability of the models. The resulting non-volcanic ensemble mean timeseries are shown in Figure 1a-c together with the observed timeseries.

In order to create a prediction of a temperature timeseries for the years n+1 onwards, a bias correction is needed to shift the ensemble mean to the current state of the observed temperatures. The current state is estimated using a number of years, N, before the current date, n. The correction then involves subtracting an average of the ensemble mean values over these years (n-N,n-N+1, …, n) and adding an average of the observed values over these years.

Applying this bias correction, we then predict future temperature values from simulated values for the years n+1 onwards. We call this IPCC ensemble based method IENS.

As reference predictions, we provide an optimal persistence forecast which is the mean of the N years before the current date (we call this FLAT), and a simple persistence estimate which is the value of the year before the forecast (equivalent to N=1, we call this PERSISTENCE). By the construction of the FLAT forecast, the IENS forecast will have higher skill if the fluctuations simulated by the models are in average realistic. If the climate models predicted constant values both forecasts would be the same. We note here that a linear trend prediction, modeled using an optimized window length for the trend fit, was initially included for comparison. Although not shown here, the skill of this forecast was always less than that of the IENS method, and often less than for the FLAT method.

Obviously the forecasts IENS and FLAT depend on the calibration window length, N. The optimal N depends on the properties of the timeseries as well as on the lead time and is determined by hindcasting on the historical data where N is required to be the number of years which minimizes the root mean squared error (RMSE). Figure 2 shows the dependence of the RMSE of 5-year mean hindcasts on N. It is based on a hindcast experiment using all available data. As one 20C3M run starts in 1900, and we allow a maximum window length of 30 years, this corresponds to hincasts from 1930-2006 .

In terms of forecast error, there is an optimal calibration window, which is in this case seven years for all of the IENS hindcasts. The RMSE of the IENS methods are lower than the RMSE of the FLAT method which shows that the IPCC ensemble mean adds skill to the forecast. How can we explain the shape of the calibration window length dependence? For very short calibration windows, the mean state cannot be well estimated and has a high variance. Therefore when increasing the calibration window the RMSE decreases approximately with 1/sqrt(N). However, for long calibration periods, biases between the observations and the model mean, due to natural variability or structural errors of the models, become important and contribute to an increasing RMSE. A balance between these effects gives the minima seen in Figure 2. From this figure, we see that for longer window lengths the slopes of the error for IENS and for the FLAT predictions diverge. This is due to the fact that when the observational time series contains a trend, an average over large window lengths becomes more and more biased as the window length increases. Because the IENS ensemble can partially alleviate this bias, the RMSE for the IENS hindcasts increases at a slower rate than the RMSE for the FLAT hindcast.

4. Validation method

To compare these prediction methods, we use an estimate of the RMSE of hindcast experiments. Retrospective forecasts are calculated using the same technique as for the future forecasts. For each hindcast year, the window length for the FLAT and the IENS method are re-estimated using all data except an interval of 10 years surrounding the years to be hindcast. This is done to minimize the artificial inflation of forecast skill which occurs when the window length, N, is estimated using the same data as is used to validate the forecast. This method of independently re-estimating N for each hindcast year accounts for the uncertainty in its estimation.

Because there is a limited hindcast period, we also supply the 90% bootstrapping confidence intervals to estimate the uncertainty of the estimate of the RMSE. These confidence intervals are derived by randomly sampling (with replacement), m observed values together with their hindcast predictions. Here m is the total number of years used in the validation study for the hindcast. This is repeated 10,000 times and a RMSE is estimated each time to derive a distribution.

There are some reasons why the hindcast RMSE may be a conservative estimate of the forecast RMSE:

1.) In the case that there is no volcano in the forecast period, the error may be less than the estimate since the hindcast is performed over past periods which did include volcanoes.

2.) The mean of three scenarios is used for the forecast, but there is only one scenario for most hindcast years. Therefore, the residual of the internal variability is smaller for years after 2000, which might reduce the forecast error.

3.) We perform the validation on all available years (1930-2006) to represent the natural variability. However, one can argue that the higher ratio, of externally forced change to natural variability, in recent years will reduce the future error of the IENS approach.

There are also reasons why our hindcast RMSE may be optimistic:

1.) There is some uncertainty in the model forcing that will only be represented by the last years of the hindcast experiment.

2.) Some of the model results may be tuned to the observational period causing a lower hindcast RMSE than might be expected.

5. Results of validation and forecasts

The estimated RMSE for the different methods are compared in Figure 1d. The RMSEs, shown in Figure 1d, are slightly higher than the minimum RMSE in Figure 1a-c since these errors also include the uncertainty in the window length estimation. From this figure we see that the IENS forecast is generally more accurate than the reference methods, FLAT and PERSISTENCE. However, by bootstrapping the residuals we are able to estimate a conservative 90% confidence interval for the RMSE as shown by the error bars on the IENS value on the histogram. Using this as our guide, we have confidence that the ensemble mean forecast is significantly better than the PERSISTENCE forecast but not necessarily better than the FLAT forecast for the MDR SSTs. This is perhaps understandable if much of our prediction skill comes from the bias-correction, or estimate of the current state. The added skill due to the anthropogenic forcing modeled by the ensemble mean is most useful in the global mean and NH mean temperature where natural variability is small due to the large spatial averaging. This result collaborates well with the results from [*Lee et al.*, 2006], who found decadal climate prediction skill on the global mean temperature from changes in anthropogenic forcing.

Next we compare these predictions to a more complex model on other interannual time scales. Smith et al. [2007] used the HadCM3 model with assimilated observations to predict temperatures out to 9 years. We briefly compare the hindcasting error of their method with the hindcast errors of our simple prediction technique.

Figure 3 shows the RMSE of annual mean global temperature forecasts using the IENS, FLAT and PERSISTENCE method for lead times from 1-9 years. The hindcasts are based on 1939-2006. 1939 is chosen as the initial year for the hindcasts so that we could test window lengths of periods up to 30 years. The IENS method shows the best skill for all lead times and all three forecast methods show a decrease in skill for longer lead times. The difference between FLAT and PERSISTENCE RMSE decreases with lead time whereas the difference between IENS and FLAT increases with lead time. The reason for this is that when the bias dominates, for the FLAT and PERSISTENCE models, the better estimate of the mean state becomes less important.

As IENS predicts a trend in the right direction, the increase in RMSE with lead time is slower. In Figure 3b we show the same results but derived from a hindcast on the same years as Smith et al. [2007] used. (First validation year 1983, last validation year 2004). It can therefore be compared to Figure 1a) of Smith et al. [2007]. For this experiment, the optimal window lengths were determined on the data prior to 1983 and are therefore using completely independent data for the model choice and validation. Our simple method shows less skill for one and two year lead times compared to the assimilated forecast system (DEPRESYS) from Smith et al. [2007]. For longer lead times the RMSE compares well with that of their DEPRESYS system, and performs significantly better, according to their 90% confidence interval, than their non-assimilated reference forecast

(NOASSIM). The reduced skill of our 1-2 year forecasts may be due to the fact that the Smith et al. [2007] model has skill in predicting El Nino, and that it uses a persistence of the sulphate forcing and therefore include parts of the volcanic forcing. As we only use the “non-volcanic” ensemble for the validation, the eruption of Pinatubo in 1991 will also decrease our hindcast skill in comparison to theirs.

Smith et al. [2007] further gives the RMSE derived from hindcast experiments on different time averages of the global mean temperature, averaged over all lead times. We performed the same hindcasting experiments with IENS, again on the same years we assume Smith et al. [2007] used. Our RMSE results are 0.106 (IENS) compared to 0.105 (DEPRESYS) for annual averages, 0.059 (IENS) compared to 0.066 (DEPRESYS) for 5-year means and 0.044 (IENS) compared to 0.046 (DEPRESYS) for 9-year means. By its construction, the only multi-decadal variability that our model predicts is the persistence part. As it still performs similar to the model of Smith et al. [2007], which models natural variability for lead times larger than two years, this suggests that most of the skill of the DEPRESYS model comes from a bias correction, i.e. starting at the right state.

It should be noted that it is difficult to make such a comparison using only the time period after 1982. As the global mean temperature was dominated by a relatively linear trend in these years compared to a more regime shift-like behavior in the 1970s [*Graham*, 1994], a comparison over earlier time periods would be needed to estimate a more accurate forecast error.

The actual IENS forecast for 2007-2011 is shown in Figure 1 a-c and the forecasts are listed in Table 1. Compared to the recent decade, global mean temperature is predicted to increase more than the other temperature predictions. This is due to the model ensemble mean which predicts a stronger temperature increase in GL than in NH. One reason for this may be a slight decrease in the Atlantic Thermohaline Circulation (THC) in the models as a response to increasing CO_2 [*Gregory et al.*, 2005]. The THC reduction has a stronger effect on NH than on GL [*Knight et al.*, 2005] and would therefore partly offset the warming trend in the NH temperature timeseries. For the MDR SST, our model predicts a slight cooling compared to the last five years. The reason for this is that the last five years were exceptionally warm compared the "optimal" calibration timescale of seven years, and that the amplitude of the externally forced trend in this region is smaller than that of the GL or NH temperature trends. For this reason the RMSE of this forecast, which are given in Table 1, show that the uncertainty of the forecast in the MDR region is high.

.

6. Conclusions

Our simple technique of using the IPCC ensemble mean, bias-corrected to the current climate as a prediction for future temperatures, compares favorably with both statistical predictions and the predictions from a complex forecast model by Smith et al. [2007]. We attribute this skill to the combination of a bias-correction, which accounts for the longer-scale natural variability, and the mean of the IPCC ensemble, which, while averaging out the internal variability of each model, predicts the response due to anthropogenic forcing.

Our quinquennial forecasts, for the global and northern hemispheric mean temperatures of 2007-2011, predict unprecedented warmth. However, a slight decrease in MDR SSTs compared to the last five years is predicted. Compared to the last decade the global mean temperature is predicted to increase faster than the NH mean temperature which may be due to a slight decrease in the thermohaline circulation which some models are simulating as a response to increasing CO_2.

Since we envision that dynamical forecasting using assimilated initial conditions is actually the future for predictions on these time scales and yet acknowledge the huge technical and computing resources that this requires, we suggest that the presented simple forecasting method can serve as a benchmark for future prediction schemes.

Acknowledgements

We would like to thank the international modeling groups for providing their data for analysis, the Program for Climate Model Diagnosis and Intercomparison for collecting and archiving the model data, the JSC/CLIVAR Working Group on Coupled Modelling and their Coupled Model Intercomparison Project 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 Climate Change Research Division of the U.S. Department of Energy's Office of Science, Biological and Environmental Research (BER) program. Thomas Laepple would like to thank Gerrit Lohmann for helpful discussions and the support of this work.

References List

Emanuel, K., Increasing destructiveness of tropical cyclones over the past 30 years, *Nature*, *436* (7051), 686-688, 2005.

Goldenberg, S.B., C.W. Landsea, A.M. Mestas-Nunez, and W.M. Gray, The recent increase in Atlantic hurricane activity: Causes and implications, *Science*, *293* (5529), 474-479, 2001.

Graham, N.E., Decadal-scale climate variability in the tropical and North Pacific during the 1970s and 1980s: observations and model results, *Climate Dynamics*, *10* (3), 135-162, 1994.

Gregory, J.M., K.W. Dixon, R.J. Stouffer, A.J. Weaver, E. Driesschaert, M. Eby, T. Fichefet, H. Hasumi, A. Hu, J.H. Jungclaus, I.V. Kamenkovich, A. Levermann, M. Montoya, S. Murakami, S. Nawrath, A. Oka, A.P. Sokolov, and R.B. Thorpe, A model intercomparison of changes in the Atlantic thermohaline circulation in response to increasing atmospheric CO2 concentration, *Geophysical Research Letters*, *32* (12), -, 2005.

Hansen, J., R. Ruedy, M. Sato, and K. Lo, NASA GISS Surface Temperature (GISTEMP) Analysis, *trends: A compendium of data on global change.Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, Tenn., USA*, 2006.

Kanamitsu, M., A. Kumar, H.M.H. Juang, J.K. Schemm, W.Q. Wang, F.L. Yang, S.Y. Hong, P.T. Peng, W. Chen, S. Moorthi, and M. Ji, NCEP dynamical seasonal forecast system 2000, *Bulletin of the American Meteorological Society*, *83* (7), 1019-+, 2002.

Knight, J.R., R.J. Allan, C.K. Folland, M. Vellinga, and M.E. Mann, A signature of persistent natural thermohaline circulation cycles in observed climate, *Geophysical Research Letters*, *32* (20), -, 2005.

Lee, T.C.K., F.W. Zwiers, X.B. Zhang, and M. Tsao, Evidence of decadal climate prediction skill resulting from changes in anthropogenic forcing, *Journal of Climate*, *19* (20), 5305-5318, 2006.

Mason, S.J., L. Goddard, N.E. Graham, E. Yulaeva, L. Sun, and P.A. Arkin, The IRI seasonal climate prediction system and the 1997/98 El NiÃ±o event, *Bull.Amer.Meteor.Soc*, *80* (9), 1853-1873, 1999.

Palmer, T.N., A. Alessandri, U. Andersen, P. Cantelaube, M. Davey, P. Delecluse, M. Deque, E. Diez, F.J. Doblas-Reyes, H. Feddersen, R. Graham, S. Gualdi, J.F. Gueremy, R. Hagedorn, M. Hoshen, N. Keenlyside, M. Latif, A. Lazar, E. Maisonnave, V. Marletto, A.P. Morse, B. Orfila, P. Rogel, J.M. Terres, and M.C. Thomson, Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER), *Bulletin of the American Meteorological Society*, *85* (6), 853-+, 2004.

Rayner, N.A., D.E. Parker, E.B. Horton, C.K. Folland, L.V. Alexander, D.P. Rowell, E.C. Kent, and A. Kaplan, Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, *Journal of Geophysical Research-Atmospheres*, *108* (D14), -, 2003.

Reynolds, R.W., N.A. Rayner, T.M. Smith, D.C. Stokes, and W.Q. Wang, An improved in situ and satellite SST analysis for climate, *Journal of Climate*, *15* (13), 1609-1625, 2002.

Santer, B.D., T.M.L. Wigley, P.J. Gleckler, C. Bonfils, M.F. Wehner, K. AchutaRao, T.P. Barnett, J.S. Boyle, W. Bruggemann, M. Fiorino, N. Gillett, J.E. Hansen, P.D. Jones, S.A. Klein, G.A. Meehl, S.C.B. Raper, R.W. Reynolds, K.E. Taylor, and W.M. Washington, Forced and unforced ocean temperature changes in Atlantic and Pacific tropical cyclogenesis regions, *Proceedings of the National Academy of Sciences of the United States of America*, *103* (38), 13905-13910, 2006.

Smith, D.M., S. Cusack, A.W. Colman, C.K. Folland, G.R. Harris, and J.M. Murphy, Improved surface temperature prediction for the coming decade from a global climate model, *Science*, *317* (5839), 796-799, 2007.

Zwiers, F.W., Climate change - The 20-year forecast, *Nature*, *416* (6882), 690-691, 2002.

Figure captions including Figures:

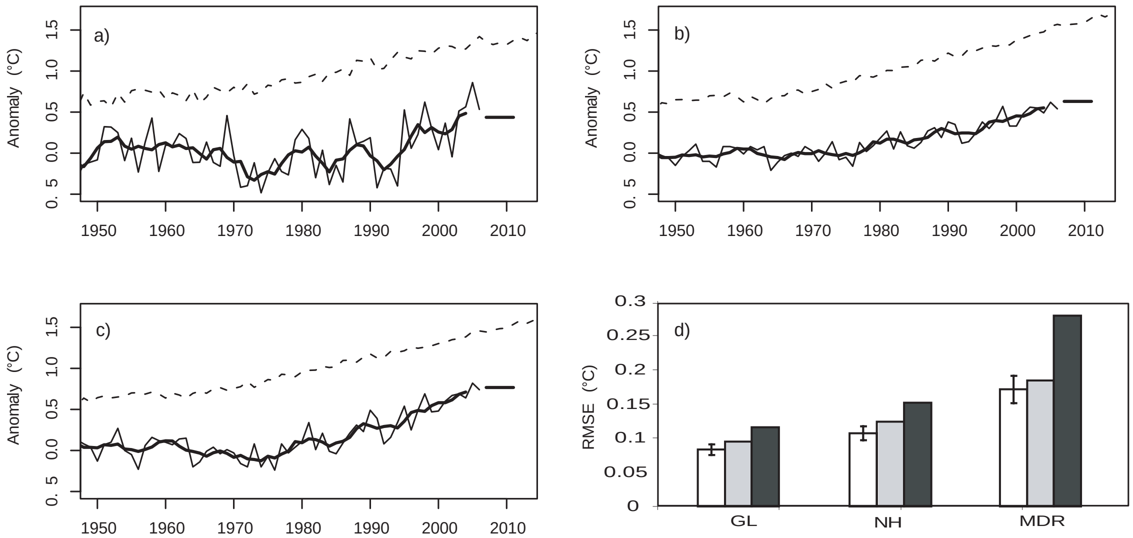

Figure 1. 1a-c show the observed timeseries (thin line), the 5year mean of the observed timeseries (thick line) and the ensemble mean of the non-volcanic model runs (dashed line). The corresponding indices are MDR SST (a), the NH temperature (b) and GL temperature (c). All timeseries are anomalies from 1951-1980 and the ensemble mean timeseries is shifted by 0.75K for easier visual comparison with the observations. Additionally the 2007-2011 forecast of the IENS method is shown as horizontal thick line. The RMSE associated with each prediction using IENS (white), FLAT (gray) and PERSISTENCE (black) is shown in 1d. The error bar on the IENS RMSE value is the 90% bootstrap confidence interval.

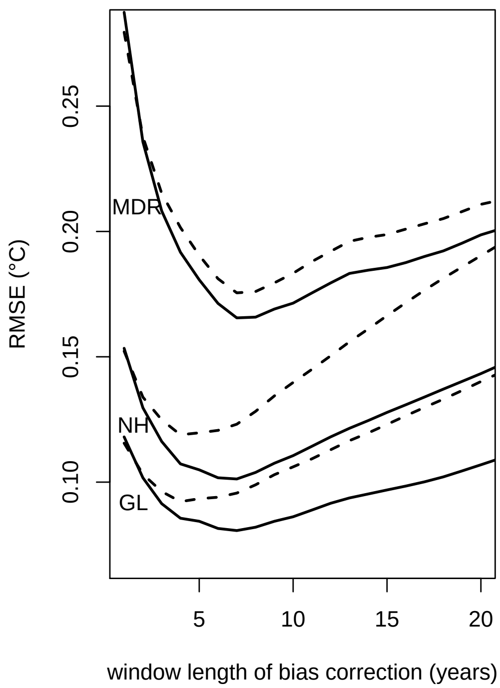

Figure 2. Impact of the bias correction window length on the hindcast skill. The RMSE of the GL temperature, the NH temperature and the MDR SST five year means are shown for the IENS method (continuous line) and for the FLAT method (dashed line)

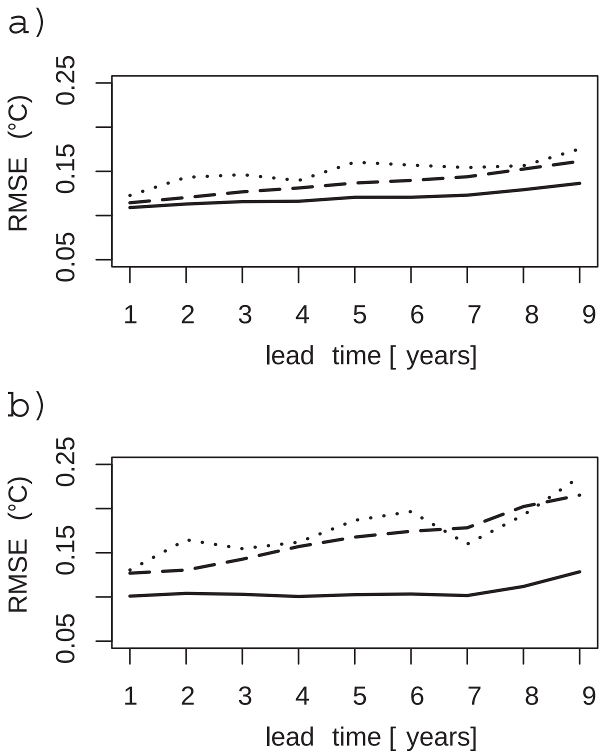

Figure 3. Dependence of the hindcast skill on lead time. RMSE for annual global mean temperature are shown. a) using IENS-forecast (continuous), FLAT forecast (long dashed) and using PERSISTENCE forecast (dotted). b) as in a) but the validation years are restricted to 1982-2004 to allow for a direct comparison with Figure 1a of Smith et al. [2007].

Tables including captions:

Table 1. The predictions for the 2007-2011 surface temperature mean from the IENS technique. Additionally the estimated RMSE of the forecast and the optimal calibration window length used are given.

	GL SAT	NH SAT	MDR SST
Forecast, relative to 1951-1980 (˚C)	0.63	0.77	0.44
Forecast error, RMSE (˚C)	0.084	0.107	0.171
Calibration window length (years)	7	7	7