Project no. GOCE-CT-2003-505539

Project acronym: ENSEMBLES

Project title: ENSEMBLE-based Predictions of Climate Changes and their Impacts

Instrument: Integrated Project

Thematic Priority: Global Change and Ecosystems

Deliverable Reference Number and Title
M1.2: Preliminary assessment of the relative merits of the multi-model approach, the perturbed parameter approach, and the stochastic physics approach, to representing model uncertainty in seasonal to decadal forecasts. Recommendations to the ENSEMBLES project concerning the design of the production ensemble system

Due date of deliverable: 28 February 2006
Actual submission date: 9 March 2006

Start date of project: 1 September 2004 Duration: 60 Months

Organisation name of lead contractor for this deliverable: ECMWF

Revision [draft, 1, 2, ..]

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	x
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the Consortium (including the Commission Services)	

ENSEMBLES

Milestone M1.2

WP number: 1.4

Participants: ECMWF, METO-HC, IfM-Kiel, CNRM, CERFACS

Due: month 18

Preliminary assessment of the relative merits of the multi-model approach, the perturbed parameter approach, and the stochastic physics approach, to representing model uncertainty in seasonal to decadal forecasts. Recommendations to the ENSEMBLES project concerning the design of the production ensemble system

1. Introduction

Despite the fact that predictable climate signals can arise from atmosphere-land-ocean interaction, the climate system is intrinsically chaotic. This implies that the predicted day-to-day evolution is necessarily sensitive to initial conditions (*Palmer*, 1993). In practice, the impact of such sensitivity can be determined by integrating forward in time ensembles of forecasts of a model, the individual members of the ensemble differing by small perturbations to the starting conditions. The phase-space dispersion of the ensemble is expected to give a quantifiable flow-dependent measure of the underlying predictability of the flow.

However, if uncertainties in initial conditions are the only perturbations represented in a climate forecast ensemble, the resulting measures of predictability will not be reliable. Climate models are not perfect. Model error in climate forecasts occurs because climate models cannot in principle simulate every single aspect of the climate system with arbitrary detail. For example, the way in which processes on spatial scales smaller than the resolved grid scales are represented in climate models is uncertain. Likewise, many of the physical parameters used in climate models either don’t have a direct equivalent in the real climate system, or their numerical values are not precisely known.

At present, there is no underlying theoretical formalism from which a probability distribution of model uncertainty can be estimated (*Palmer et al.*, 2005) and more pragmatic approaches must be sought. One such approach relies on the fact that global climate models have been developed somewhat independently at different climate institutes, using different numerical approaches to represent the climate dynamics and applying different parameterizations of physical processes. An ensemble comprising such quasi-independent models is referred to as a multi-model ensemble. Another approach to represent model uncertainty is based on including a stochastic term in the physical tendencies of the model, a method known as “stochastic physics” (*Palmer*, 2001). In particular, a cellular automaton backscatter scheme (CASBS; *Shutts*, 2005) backscatters energy to scales above the truncation limit using a total energy dissipation function involving contributions from numerical diffusion, mountain drag and deep convection. A third approach, the perturbed-parameter method (*Murphy* et al., 2004), creates an ensemble of simulations by changing either one or a set of parameters of the numerical schemes used to represent the sub-grid scale processes in a given model. The relative merits of all these methods are being investigated in ENSEMBLES by performing co-ordinated experiments.

Seasonal time scale dynamical climate predictions are now made routinely at a number of operational meteorological centres around the world, using comprehensive coupled models of the atmosphere, oceans, and land surface (e.g. *Stockdale et al.*, 1998; *Mason et al.*, 1999). In contrast to seasonal forecasting, interannual and decadal forecasting are at their earliest stages (*Boer*, 2000), although preliminary assessments indicate that there are signs of ensemble-mean skill in multi-annual time scales (*Smith et al.*, 2006), partly due to the future impact of the increase of greenhouse gases in the atmosphere. Recent results show that the effects of anthropogenic climate forcing needs to be considered in all these forecast systems (*Doblas-Reyes et al.*, 2006). To assess how the different approaches to dealing with model uncertainty can improve seasonal-to-decadal (s2d) forecast quality, the following task was defined in ENSEMBLES:

Task 1.5.1

Seasonal and decadal-timescale ensemble integrations will be made using a) the multi-model ensemble system, b) the perturbed parameter system, c) the stochastic physics system. The seasonal integrations will be 6 months long, and made over a number of start dates at different times of year, and for ENSO and non-ENSO periods. The decadal integrations will be 10 years long and made over two contrasting decades from the 20^{th} century (e.g. 1960s and 1990s). ECMWF ERA-40 data will be used to provide atmospheric initial conditions and atmospheric verification.

This report describes the co-ordinated experiment designed to address the task 1.5.1. The main goal of this task is to provide a dataset that offers a preliminary assessment of the relative merits of the three approaches mentioned above to representing model uncertainty in s2d forecasts and to allow the drafting of recommendations to the ENSEMBLES project relevant to the design of the ensemble system. A comprehensive report will follow in month 30 as deliverable D1.8.

2. Experimental setup

A coordinated forecast experiment, which has been labelled "stream 1" or "pre-production" of the seasonal-to-decadal (s2d) ENSEMBLES integrations, has been planned. In this experiment, the following 9-member ensemble integrations were specified:

- Seasonal: 7 months long, 1st of May start date, 1991-2001 period.
- Annual: 14 months long, 1st of November start date, 1991-2001 period.
- Two 10-year integrations starting on the 1st of November 1965 and 1994

The forecast systems being used in the experiments are IFS/HOPE and IFS/HOPE CASBS (ECMWF), GloSea and DePreSys (METO-HC), ECHAM5/MPI-OM1 (IfM-Kiel), ARPEGE4.5/OPA (CNRM) and ARPEGE3/OPA (CERFACS). A brief summary of the different coupled models can be found in the deliverable D1.4. For each model, except HadCM3, uncertainties in the initial state are represented through an ensemble of nine different ocean initial conditions. In addition, IFS/HOPE uses atmospheric singular vectors. It should be noted that every DePreSys ensemble simulation is 10 years long.

All the simulations are carried out on the ECMWF super-computer and the model output is available from the ECMWF mass storage system (MARS). Although most of the integrations planned for the stream 1 experiment have been performed, not all the model output has been archived in the common repository as yet. All data available in MARS at the date of writing have been used in this preliminary assessment. In particular, the integrations used in the assessment are:

- Multi-model: seasonal, annual and one 10-year (1994 start date) simulations from IFS/HOPE, ARPEGE/OPA and GloSea models.
- Stochastic physics: seasonal, annual and one 10-year (1994 start date) simulations from IFS/HOPE. The 10-year simulation with start date November 1965 confirms the results described for the 10-year run started in 1994.
- Perturbed parameters: one 10-year (1994 start date) simulation from DePreSys.

A complete set of hindcasts with a larger multi-model and perturbed-parameter ensembles will be used in a thorough assessment as will be described in the deliverable D1.8, due by month 30.

Anomalies are computed, unless stated otherwise, by removing the model climatology in cross-validation for each initial month and lead time. A similar process is used to produce the verification anomalies. The main verification data set used in this system is ERA-40 (*Uppala et al.*, 2005), although Reynolds sea surface temperature (SST) and GPCP precipitation are used for these two variables.

3. Seasonal and annual simulations

Given the small sample of perturbed-parameter integrations, only the multi-model and stochastic physics approaches were considered in the assessment of the seasonal and annual ensemble hindcasts. The 11 years of hindcasts over the period 1991-2001 allow for an estimate of the forecast quality, but they are also used to assess the impact of the stochastic physics in the model climate.

As an illustration of the performance of these forecasts systems over the tropical oceans, Figure1 displays measures of deterministic forecast quality of SST over the Nino3 (5°N-5°S,150°W-90°W) region. Although the root mean square error (RMSE) of the ensemble mean is depicted, the conclusions hold for the anomaly correlation coefficient too. A comparison of the stochastic physics (solid blue lines in the top row panels) with the multi-model RMSE shows the superiority of the latter for the hindcasts started in May and for the early period of those started in November. The spread (measured as the average standard deviation of the ensemble members with regard to the ensemble mean) shows larger values for the multi-model, especially for the ensembles started in May. Interestingly, the spread of the multi-model approaches the RMSE, a desirable feature in a reliable prediction system (*Jolliffe and Stephenson*, 2003). It is likely that the smaller ensemble size of the stochastic physics ensemble with respect to the multi-model ensemble (9 versus 27) favours the latter in this comparison (*Hagedorn et al.*, 2005), so that a further comparison with ensembles of the same size needs to be carried out. Although the stochastic physics ensemble has slightly worse forecast quality in terms of RMSE and spread than the multi-model, Figures 1a & b illustrate one of the benefits of this approach: a uniform increase in spread for some of the hindcasts (e.g. the ensemble started in May) that gives values closer to the RMSE.

However, the benefits of the stochastic physics approach are not only found in the ensemble spread. Figure 2 shows the systematic error, computed as differences between the model and observed climatologies, of the SST over the regions Nino3 and Nino4 (5°N-5°S,160°E-150°W). The stochastic physics alleviates the excessive warming over the two regions, especially in Nino4. Although the improvement is not as large for the hindcasts started in November (not shown), CASBS improves the climate of the ocean surface in the IFS/HOPE model. This important result also holds for some atmospheric variables, as Figure 3 demonstrates. While the control IFS/HOPE integrations have an excess of precipitation in DJF (with respect to GPCP) over the tropical oceans, the CASBS version reduces this wet bias as well as other systematic errors over the continents. One of the reasons for the better performance of the CASBS version of the model is that the inter-tropical convergence zone (ITCZ) in the tropical Pacific becomes less narrow. Along with this, low-level winds over the region become more realistic, which helps explain at the same time the more realistic SST climate, forced by the direct effect of the surface wind stress on the upper ocean dynamics. For the boreal summer period, the climate also improves in terms of winds and precipitation, in particular for the Southeast Asian monsoon.

As mentioned above, the improvements in the stochastic physics experiment are not only found over the tropics, but also in the extra-tropics. Figure 4 shows the mean blocking frequency computed with the *Tibaldi and Molteni* (1990) index for winter using the hindcasts started in November. While there are no substantial changes over the North Atlantic region, the frequency increases in the stochastic physics experiment up to 30% over the Pacific region, making the simulation far more realistic than the control. The improvement over the Pacific is consistent with a reduction in 500 hPa geopotential height bias in the region (not shown), a result in agreement with previous experiments carried out with preliminary versions of the stochastic physics (*Palmer et al.,* 2005; *Jung et al.*, 2005). Similar improvements have been found in the summer circulation and in twin (uncoupled) experiments carried out with observed SSTs. However, the stochastic physics does not appear to improve all aspects of the model climate. For example the zonal wind bias over the high latitude winter hemispheres is increased with the current version of CASBS.

In summary, for seasonal and annual integrations the multi-model predictions suggest a better performance than the single-models and than the stochastic-physics ensemble, especially in terms of spread. Figure 5 illustrates the kind of predictions the multi-model and the single-model ensembles provide. The 10-month Nino3 SST forecasts displayed in the figure show some skill in predicting the variability, with an overall agreement between the different single-model ensemble means and the multi-model lying inside the envelope created by the three models. However, this is not always the case. In some instances a strong disagreement between the single models has been found, even when the forecast signal is particularly strong. In those cases the multi-model represents a compromise between three very uncertain forecast systems.

4. Decadal simulations

Figure 6a shows the evolution of the global annual mean near-surface temperature for all available decadal simulations starting in 1994, together with the 2m temperature from the ECMWF Re-Analysis ERA-40 considered as verification. Note that ERA-40 is only available until mid 2001. The simulations from CNRM, IFS/HOPE and GloSea are based on perturbations of the initial conditions only and thus these three models form the current decadal multi-model ensemble. The stochastic physics simulations were done implementing CASBS to the IFS/HOPE model as in the seasonal and annual integrations. The DePreSys model system performed simulations with varying values of selected physical parameters in HadCM3 and so constitutes the third methodological approach of this comparison, the perturbed physics ensemble.

The time series of the five model ensembles in Figure 6a reveal distinctive qualitative features in the long-term behaviour of the different coupled climate models. Whereas one model simulates a strong, but unrealistic global warming trend over the 10 years, the remaining models are in general in better agreement with the verification.

One of the most obvious characteristics, however, is the large difference in spread between the perturbed physics simulations and the other single-model and stochastic physics ensembles. In order to compare the mean errors and spread of the three approaches, Figure 6b shows the ensemble mean errors together with the standard deviations as a measure of spread for the multi-model ensemble, the stochastic physics ensemble and the perturbed parameter ensemble.

Due to the unrealistic warming in one of the models forming the multi-model, the spread of the multi-model is the largest among the three methods and is almost a factor of 2 larger that the ensemble-mean error. Given the fact that there are only three models contributing to the multi-model and that there is no account of the relative performance of the individual models in the construction of the multi-model, the unweighted raw multi-model is not considered to be a realistic representation of the underlying model uncertainty. In addition, it is not yet possible to remove the effects of systematic errors in the simulations of the multi-model ensemble members, as model climatologies are not available for the interannual to decadal time scale, in contrast to the seasonal to annual predictions of section 3. More models and a larger set of decadal simulations that allows for model calibration would be needed to address these points.

The impact of the stochastic physics on the decadal simulations manifests itself in a warmer global mean state compared to a control ensemble simulation with the same climate model (IFS/HOPE), while the spread remains basically unchanged and is comparable to the small spread of the single-model ensembles. After year 3 of the hindcast, the ensemble mean error of the stochastic physics is the smallest compared to the multi-model and perturbed physics runs and reaches a similar level to the perturbed physics ensemble toward the end of the simulation (see Figure 6b).

As mentioned above, the perturbed physics ensemble creates a large spread, approximately a factor of 5 larger than that for the stochastic physics. Whereas in the first half of the simulation the ensemble-mean perturbed physics error is the largest, it becomes closer to the level of spread and to the stochastic physics error towards the end of the simulation. All three ensemble forecast systems show large errors in forecasting the 1997 El Niño SSTs in the western Pacific (see Figure 7b for the Nino4 region).

Qualitatively similar conclusions can be drawn from a comparison of the spread of the three approaches for SSTs in different equatorial ocean regions, as shown in Figure 7. The spread of the multi-model is largest, while the stochastic physics ensemble appears to be underdispersive.

One interesting feature in the eastern tropical Pacific (Figure 7a showing the Nino3 region) is the drop of the ensemble-mean error for all three methods from 2000 onwards. Whether this can be related to the spun-up state of the models or to the lack of observed variability remains open until further verification has been carried out and/or longer simulations become available.

As could be seen from Figure 7a, the multi-model develops the largest error and spread after the fourth year. In order to understand the reasons for this behaviour, Figure 8 shows the Nino3 SST anomalies for all the multi-model ensemble members together with the spread for each of the single-model ensembles. As can be seen there are quite substantial systematic differences between the three models. IFS/HOPE simulates an initial warming in the first ca. 18 months, consistent with the warm bias found in the seasonal hindcast experiments, and then cools down to a level below the mean observed value. However, the GloSea and CNRM models have, on average over the simulation period, a warmer response with a strong seasonal cycle in the Nino3 SST anomalies. The spread of the IFS/HOPE and CNRM models is comparable in magnitude, whereas GloSea contributes to the multi-model ensemble spread with a substantially larger single-model ensemble spread.

As mentioned above, the climate models used for long-range forecasts develop systematic biases with respect to observations. These biases can dominate the total error of a forecast. The standard method of correcting the bias is to post-process predictions to remove an estimate of the systematic error diagnosed from a large set of quasi-independent hindcasts. The necessary integrations are beyond the scope of ENSEMBLES stream 1, so bias-corrected hindcasts cannot yet be formulated in the multi-model and stochastic physics cases. However, the DePreSys system uses a method of bias removal based on the multiannual mean of a long climate simulation, rather than a large hindcast set. The required climate simulations have already been carried out for WP 1.6, so it is possible to assess the skill of bias-corrected versions of the DePreSys forecasts.

Figure 9a displays a plume of bias-corrected global mean surface air temperature anomalies for the November 1994 start date. The DePreSys ensemble consistently predicts positive anomalies (relative to a baseline period of 1958-2001), and a secular warming trend is also apparent, consistent with the expected impact of the radiative forcings included in the integrations. Encouragingly, the ensemble distribution encompasses the observed anomalies at most lead times. For example, the ensemble spread is wide enough to indicate the possibility of the record warm anomalies associated with the 1997/8 El Niño, even though skilful deterministic predictions of the event would not be possible at multiannual lead times. In the absence of bias correction, however, the observed time series appears as an outlier of the DePreSys ensemble distribution (Figure 6a), because the impact of the model systematic errors dominates the smaller signals associated with the evolution of each ensemble member with respect to its own long term climate.

Figure 9b shows the RMSE of the ensemble mean, and the ensemble standard deviation, of bias-corrected DePreSys anomalies of global mean surface air temperature, based on results from five start dates (May 1991-94 and Nov 94). Values of the error and spread are reduced substantially when bias removal is applied (cf Figure 6b), confirming the need to remove the effects of systematic errors in order to quantify the predictability of the system. For example, the RMSE typically exceeds the average ensemble spread in Figure 9b by ~50%, close to the ratio which would be expected from a system in which the forecast errors are consistent with the ensemble spread. Figure 9c shows a corresponding plot for predictions of surface air temperature over the Nino3 region: the smallest values of error and spread are found in year 1, consistent with previous results showing a degree of deterministic predictability in ENSO predictions at this lead time (Smith et al., 2006). The increased errors for years 4-7 reflect the influence of the 1997/8 El Niño, which occurs at different lead times in different hindcasts.

5. Conclusions and recommendations

On the seasonal and annual time scales the preliminary results presented in this report suggest that the multi-model ensemble forecasting system performs slightly better than the single-model ensemble and the approach using stochastic physical parameterizations. Integrations with the perturbed parameter approach are not available yet, so that a full comparison cannot be presented.

As for the decadal integrations, the three systems have been compared for a single start date in this preliminary assessment. The mean error of each of the three methods is of comparable magnitude. The perturbed parameter ensemble shows the advantage of a better balance between ensemble mean error and spread. However, larger hindcast data sets are needed for more conclusive statements. A larger sample of five start dates has been assessed for the perturbed parameter approach. The results show the importance or removing systematic simulation biases, since these are often larger than the climate anomalies being predicted. Bias-correcting the global surface temperature yields predicted anomalies showing approximate consistency between forecast errors and ensemble spread.

The stochastic physics approach shows promising results and some positive impacts in representing climate processes and reducing systematic errors. In the light of this result further research is planned in the coming months.

On the basis of this preliminary assessment, our recommendations to the ENSEMBLES project concerning the design of the production ensemble system are:

- To carry out a more comprehensive assessment of the three methodological approaches to address the problem of model uncertainty
- To continue with the research on the implementation of all three approaches
- To carefully design, on the basis of the above, the stream 2 experiments

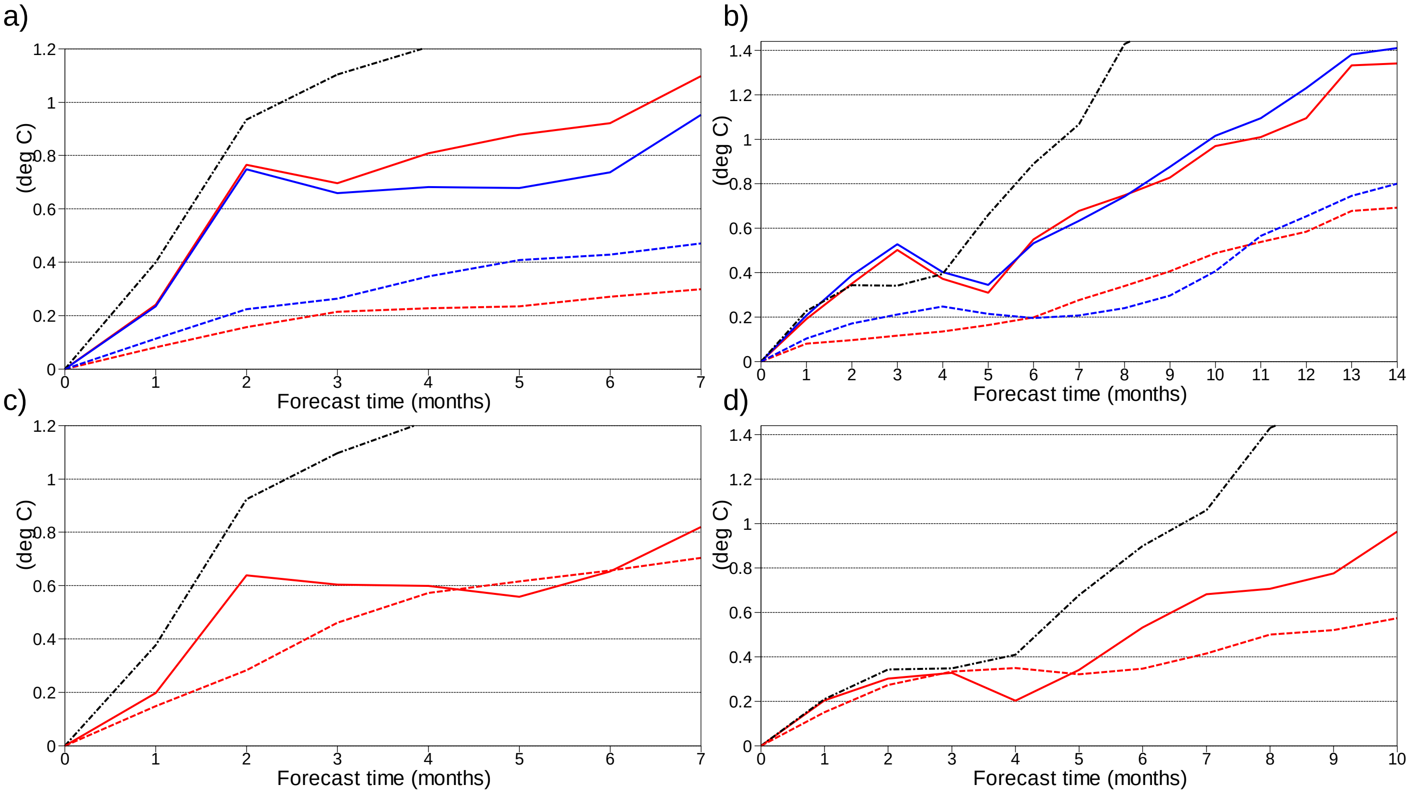

Figure 1: RMS error of the ensemble mean (solid lines) and spread of the 9-member ensembles as measured by its standard deviation (dashed lines) over lead time for Nino3 SST hindcasts from 1991 to 2001 starting on May 1st (left) and November 1st (right). Top row: IFS/HOPE control (red) and IFS/HOPE stochastic physics using CASBS1.1 (blue). Bottom row: multi-model ensemble consisting of IFS/HOPE control, GloSea and CNRM. For comparison, the RMS errors of a persistence forecast are shown with the black dashed lines.

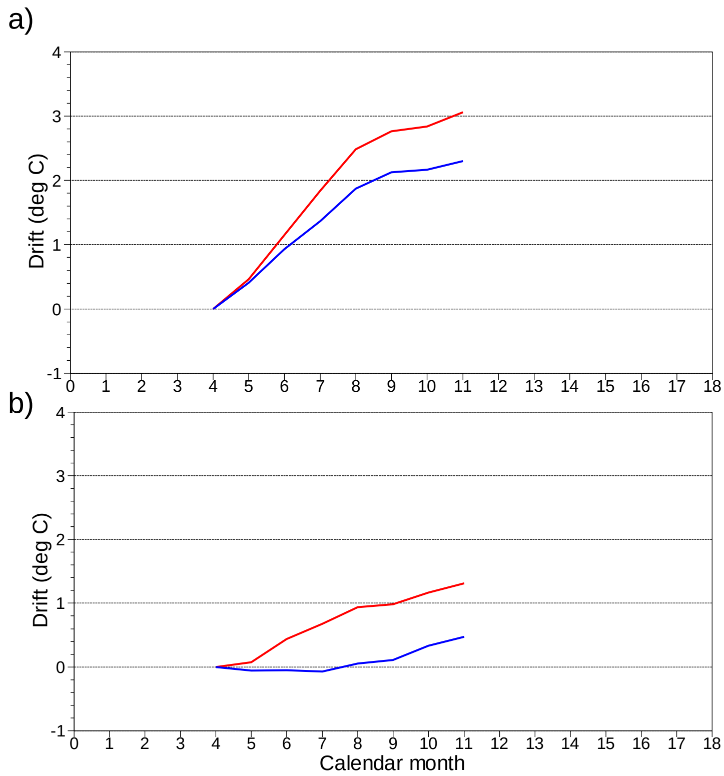

Figure 2: Mean SST drift over lead time for SST hindcasts from 1991-2001 starting on May 1st in the Nino3 (a) and Nino4 (b) regions for the two IFS/HOPE model versions: control (red) and stochastic physics using CASBS1.1 (blue).

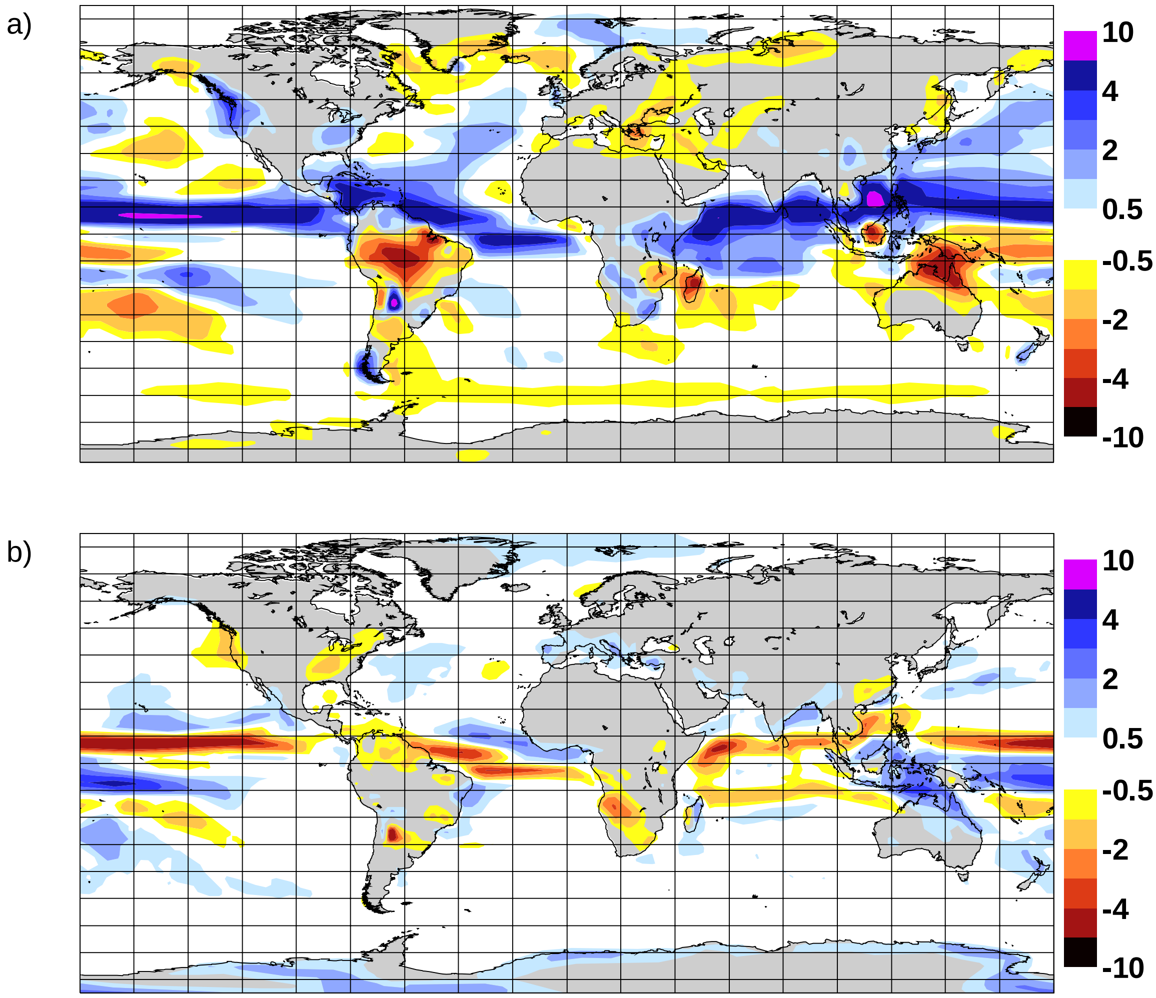

Figure 3: Differences in DJF mean precipitation (in mm/day) between IFS/HOPE control and GPCP (a) and between IFS/HOPE stochastic physics using CASBS1.1 and IFS/HOPE control (b) for 1991-2001 from the November start hindcasts.

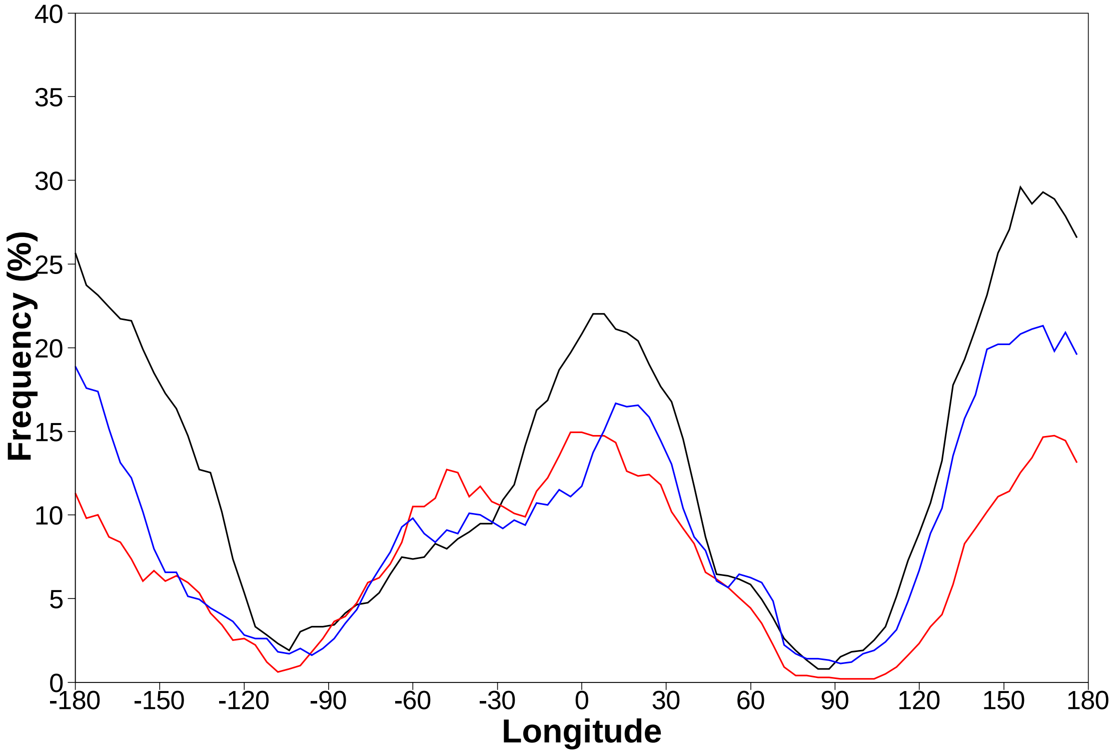

Figure 4: Blocking frequency after *Tibaldi and Molteni* (1990) during DJF for 1991-2001 November start hindcasts of IFS/HOPE control (red), IFS/HOPE stochastic physics using CASBS1.1 (blue) and ERA-40 (black).

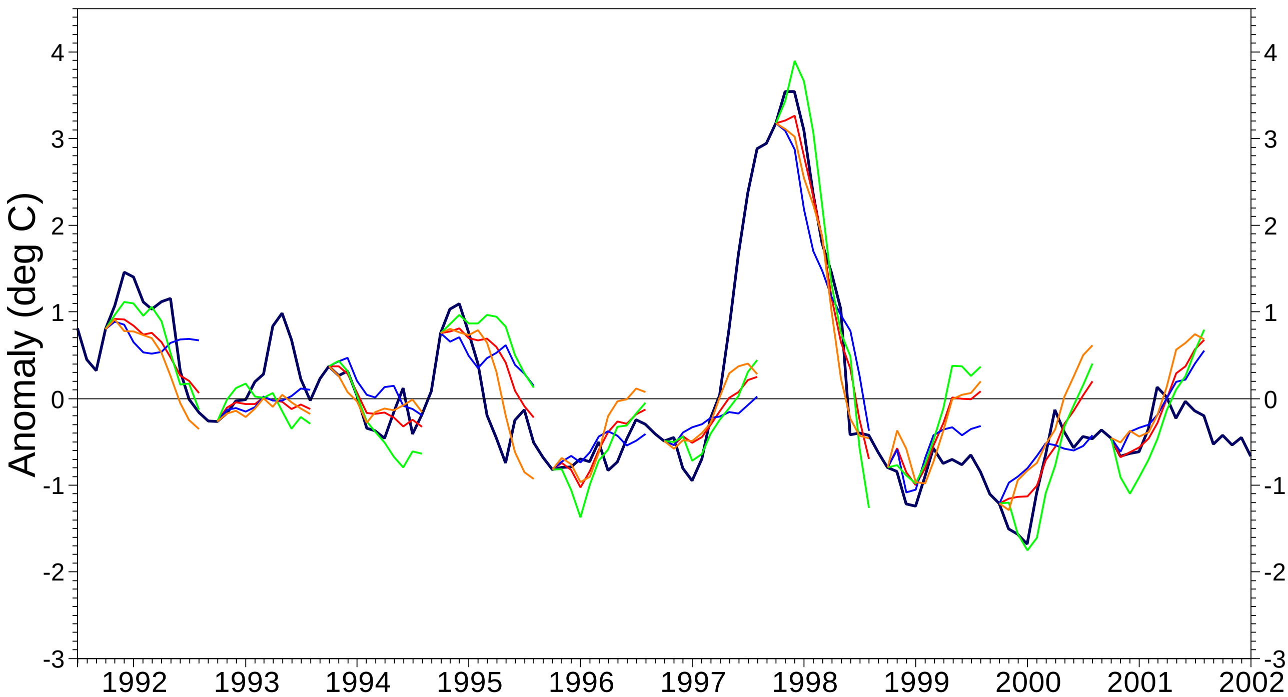

Figure 5: Multi-model (red) and single-model ensemble-mean Nino3 SST anomalies for 1991-2001 November start hindcasts. The individual model ensemble means are IFS/HOPE control (blue), GloSea (green) and CNRM (orange). The thick dark line shows the observed anomalies with respect to the climatology.

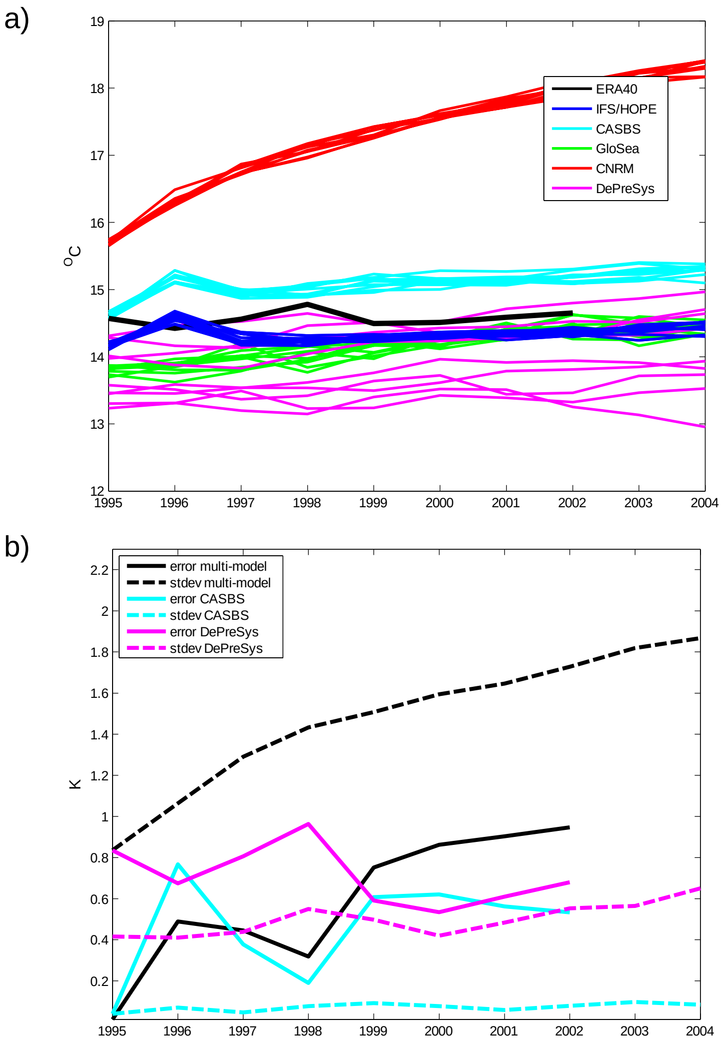

Figure 6: a) Global annual mean near-surface temperature of all available decadal simulations starting on 1st November 1994. The individual models are IFS/HOPE control (dark blue), IFS/HOPE stochastic physics using CASBS1.2 (light blue), GloSea (green), CNRM (red), DePreSys perturbed parameters (magenta). ERA-40 (black) covers the years up to 2001. b) Absolute error of the ensemble mean (solid lines) and spread of the ensemble as measured by its standard deviation (dashed lines) for the three different ensemble systems: multi-model initial condition ensemble combining IFS/HOPE control, GloSea and CNRM (black), stochastic physics ensemble based on IFS/HOPE CASBS1.2 (light blue) and perturbed physics ensemble based on DePreSys (magenta).

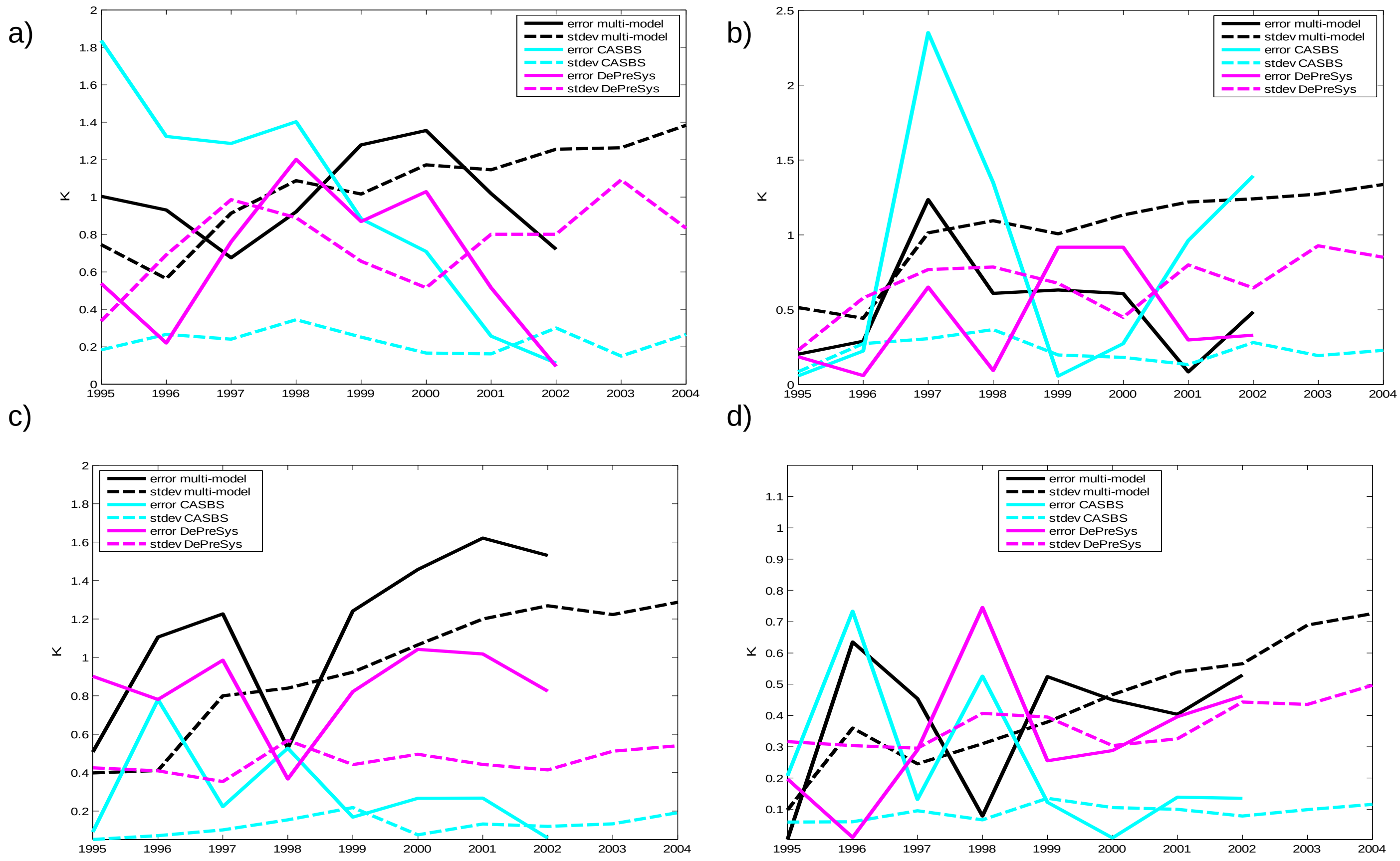

Figure 7: Absolute SST error of the ensemble mean (solid lines) and spread of the ensemble as measured by its standard deviation (dashed lines) for the three different ensemble systems and 4 different oceanic regions. The regions are a) Nino3, b) Nino4; c) equatorial Atlantic, d) equatorial Indian Ocean. The lines correspond to the multi-model initial condition ensemble combining IFS/HOPE control, GloSea and CNRM (black), stochastic physics ensemble based on IFS/HOPE CASBS1.2 (light blue) and perturbed physics ensemble based on DePreSys (magenta).

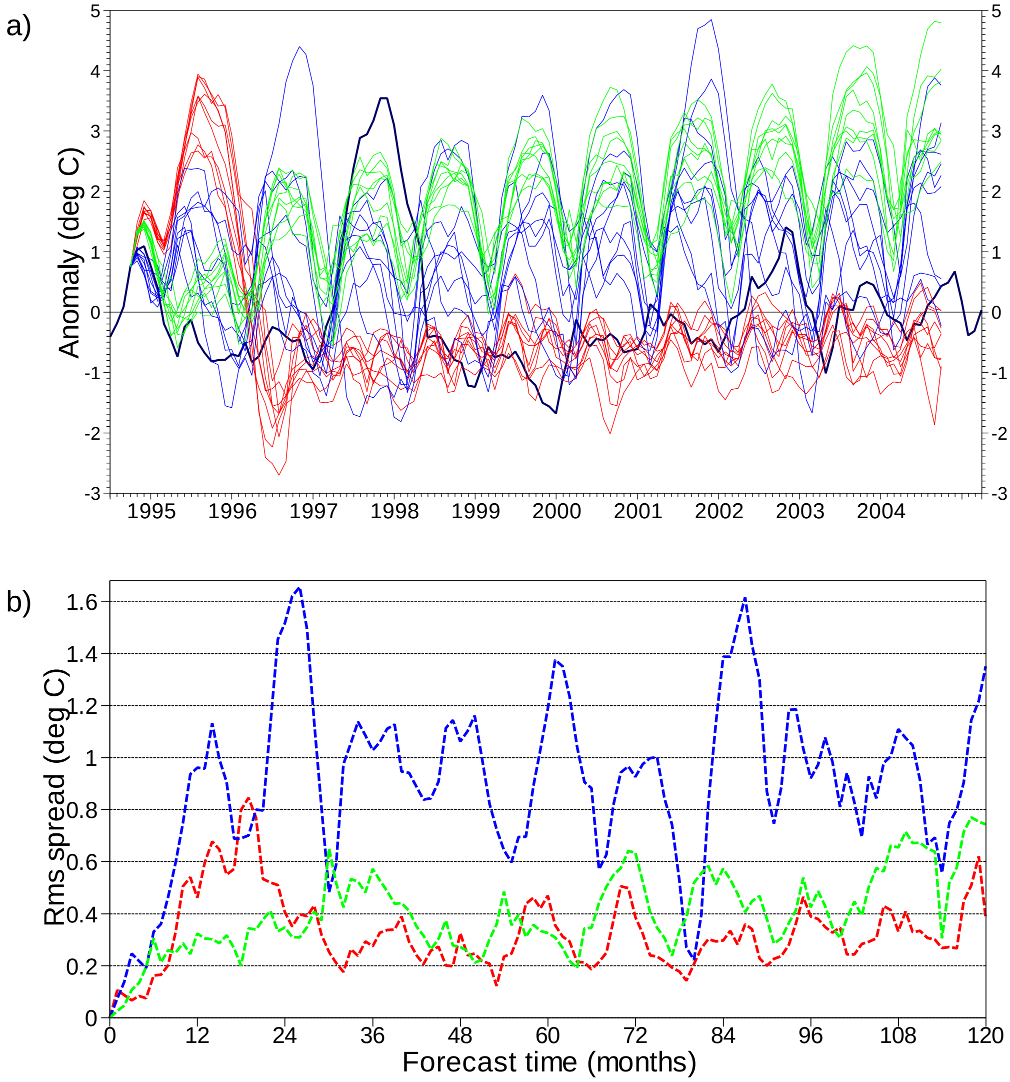

Figure 8: a) Nino3 SST anomalies of the decadal multi-model initial condition ensemble started in November 1994. The individual models are IFS/HOPE control (red), GloSea (blue) and CNRM (green). The black line shows the observed anomalies with respect to the observed climatology. The same observed climatology is used to compute the model anomalies. b) Nino3 SST spread of the individual single-model ensembles as measured by its standard deviation.

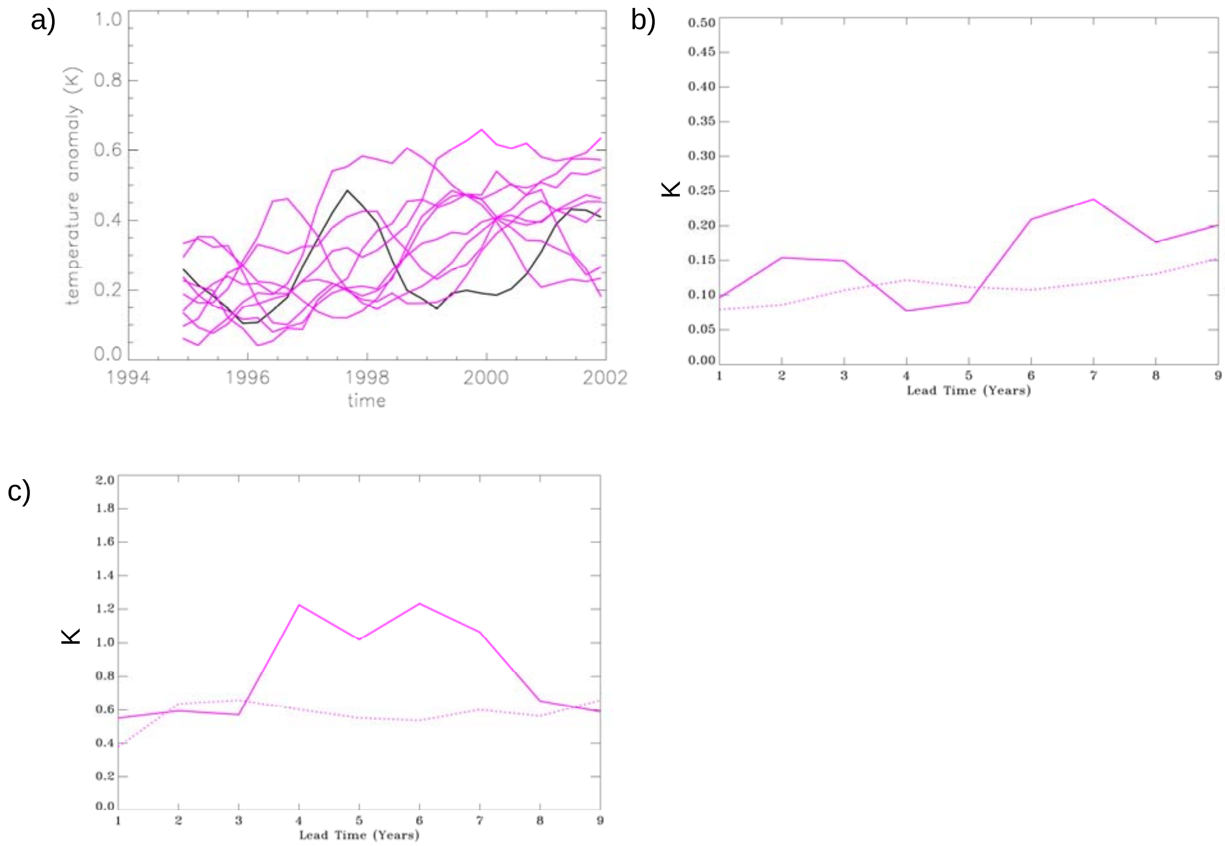

Figure 9: a) Global annual-mean near-surface temperature anomalies predicted by the DePreSys ensemble members from the 1st November 1994 start date (magenta lines), and observed (black line, based on ERA-40). Anomalies are calculated relative to 1958-2001. b) RMS error of the ensemble-mean (solid line), and ensemble standard deviation (dotted line), for hindcasts of global annual mean near-surface temperature anomalies from DePreSys. Results are averaged over five start dates, May 1991-94 and Nov 1994. c) As (b), for predictions of annual mean near-surface air temperature over the Nino3 region.

References:

Boer, G. J., 2000. A study of atmosphere-ocean predictability on long time scales. *Climate Dyn.*, **16**, 469-472.

Doblas-Reyes, F.J., R. Hagedorn, T.N. Palmer and J.-J. Morcrette, 2006. Impact of increasing greenhouse gas concentrations in seasonal ensemble forecasts. *Geophys. Res. Lett.*, in press.

Hagedorn, R., F.J. Doblas-Reyes and T.N. Palmer, 2005. The rationale behind the success of multi-model ensembles in seasonal forecasting. Part I: Basic concept. *Tellus A*, **57**, 219-233.

Jung T., T.N. Palmer, G.J. Shutts, 2005. Influence of a stochastic parameterization on the frequency of occurrence of North Pacific weather regimes in the ECMWF model. *Geophys. Res. Lett.*, **32**, L23811, doi:10.1029/2005GL024248.

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

Murphy, J. M., D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, D. A. Stainforth, 2004. Quantifying uncertainties in climate change from a large ensemble of general circulation model predictions. *Nature*, **430**, 768-772.

Palmer, T.N., 1993. Extended-range atmospheric prediction and the Lorenz model. *Bull. Amer. Meteor. Soc.*, **74**, 49-65.

Palmer, T.N., 2001. A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models. *Quart. J. Roy. Meteor. Soc*., **127**, 279-304.

Palmer, T.N., G.J. Shutts, R. Hagedorn, F.J. Doblas-Reyes, T. Jung and M. Leutbecher, 2005. Representing model uncertainty in weather and climate prediction. *Annual Review of Earth and Planetary Sciences*, **33**, 163-193.

Smith, D. M., A. W. Colman, S. Cusack, C. K. Folland, S. Ineson and J. M. Murphy, 2006. Predicting surface temperature for the coming decade using a global climate model. *Nature*, submitted.

Stockdale, T.N., D.L.T. Anderson, J.O.S. Alves and M.A. Balmaseda, 1998. Global seasonal rainfall forecasts using a coupled ocean-atmosphere model. *Nature*, **392**, 370-373.

Shutt, G., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. R. Meteorol. Soc*., **131**, 3079-3102.

Uppala, S.M., Kållberg, P.W., Simmons, A.J., Andrae, U., da Costa Bechtold, V., Fiorino, M., Gibson, J.K., Haseler, J., Hernandez, A., Kelly, G.A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R.P., Andersson, E., Arpe, K., Balmaseda, M.A., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B.J., Isaksen, L., Janssen, P.A.E.M., Jenne, R., McNally, A.P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N.A., Saunders, R.W., Simon, P., Sterl, A., Trenberth, K.E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J. 2005: The ERA-40 re-analysis. *Quart. J. R. Meteorol. Soc.*, **131**, 2961-3012.