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

Major Milestone MM1.2

Provision of a "first generation" ensemble prediction system (Version 1) for use in RT2.

Due date of deliverable: 31 August 2006
Actual submission date: 8 September 2006

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

Organisation name of lead contractor for this deliverable:

European Centre for Medium Range Weather Forecasts (ECMWF)
Met Office - Hadley Centre for Climate Prediction and Research (METO-HC)

ENSEMBLES

Major Milestone MM1.2

RT number: RT1

A. Weisheimer (ECMWF), J. Murphy (METO-HC), M. Collins (METO-HC)

with contributions from the partners
IfM-Kiel, CNRM, CERFACS, KNMI, UOXFDC, FUB

Due: month 24

Provision of a "first generation" ensemble prediction system (Version 1) for use in RT2.

1. Introduction

This report aims at describing the first version of the ensemble-based modelling system for the prediction of climate change at different time scales from seasons to decades and beyond, provided by the partners of ENSEMBLES RT1. The system consists of global coupled climate models developed in Europe for use in the generation of multi-model simulations of future climate in RT2.

The focus of the first release of the ensemble prediction system has been twofold: i) the provision of a tested ensemble prediction system for seasonal-to-decadal (s2d) forecast ranges using different methods to represent the inherent uncertainties, and ii) the assembly of modelling strategies on centennial time scales, including the currently available Earth System model component modules.

After a short introduction and motivation for the use of the main techniques to generate ensemble forecasts on these time scales in section 2, sections 3 and 4 of the report discuss the s2d and centennial systems, respectively, in some more detail. A summary of the key developments and recommendations for future work are given in section 5.

2. Generation of ensembles

The non-linear chaotic nature of the climate system makes dynamical climate model forecasts sensitive to small perturbations introduced by both the initial state of forecasts, and variations in model formulation. Thus, individual forecasts with one fixed model are of limited value. Instead, ensembles of forecasts are used to assess the range of possible evolutions of future climate on a range of timescales. Four main techniques for a “first generation” of ensembles for climate forecasting have been explored in RT1 over the last two years:

1) Uncertainties in the initial conditions of each single model are accounted for by generating an ensemble of slightly different atmospheric and ocean analyses (*Palmer, 2000*). The perturbations of the initial conditions are either of an optimal statistical nature or based on insight into the dynamics of the physical system.

2) Model error in climate forecasts occurs because climate models cannot in principle simulate every single aspect of the climate system with arbitrary detail. The multi-model method partially samples errors that occur due to structural inadequacy in individual climate models (e.g. different model formulations and approximations, systematic biases) by using different coupled models (*Palmer et al., 2004; Weisheimer and Palmer, 2005*). This approach relies on the fact that global climate models have been developed somewhat independently at different climate institutes, using different numerical schemes to represent the dynamics and applying different parameterizations of physical processes.

3) There are uncertainties in the specification of the parameters that are used in the parameterizations in climate models. Many of the physical parameters in the models either do not have a direct equivalent in the real climate system, or their numerical values are not precisely known. By perturbing these parameters within a single model, errors in specific model formulations can be accounted for *(Murphy et al., 2004; Stainforth et al., 2005*).

4) Due to the coarse finite spatial resolution of our climate models, the representation of processes on spatial scales smaller than the truncation scales and their feedback on larger scales remains subject to considerable uncertainty. The impact of unresolved scales can be approximated by stochastic perturbations of the physical tendencies in the model (*Palmer, 2001*; *Shutts, 2005*).

In s2d forecasting, the combination of the initial-condition ensemble methodology (1) with the multi-model concept (2) leads to the multi-model ensemble approach. Together with the perturbed physical parameter (3) and stochastic parameterisation (4) strategies it forms the basis of the ensemble prediction system on seasonal, interannual and decadal time scales. For the centennial prediction system, the multi-model concept (2) and the perturbation of physical parameters (3) have been explored and provide the core of the first-generation ensemble prediction system for centennial forecast ranges.

3. Seasonal-to-Decadal predictions

Seasonal time scale dynamical climate predictions are made routinely at a number of operational meteorological centres around the world, using comprehensive coupled models of the atmosphere, oceans, and land surface (*Stockdale et al. 1998; Mason et al., 1999*). In contrast to seasonal forecasting, interannual and decadal forecasting are at their earlier stages. The major advantage of s2d predictions compared to forecasts on longer time scales, where verification *per se* is impossible, is that the quality of the s2d forecast systems can be estimated based on statistical analyses of sets of hindcast simulations. In view of future seamless weather and climate prediction systems (*Palmer and Hagedorn, 2006*), assessments of the forecast performance on s2d time ranges can, and should, be considered as an indication for potential predictability on much longer time scales.

In the past, seasonal prediction either relied on one particular forecast model (*Stockdale et al. 1998)* or made use of the multi-model ensemble approach (*Palmer et al., 2004*). However, more recently other approaches to represent model uncertainty have been developed, as discussed section 2. The ENSEMBLES first generation s2d ensemble prediction system includes three different methods to tackle the problem of model uncertainty: the multi-model ensemble technique, the perturbation of physical parameters and stochastic parameterisations of sub-grid processes. In order to compare the three approaches and to assess their relative merits in an s2d framework, a set of common simulations (so-called ‘stream 1’ simulations) for all three of them was defined, performed and analysed. Coordinated ensemble model integrations over 7 and 14 months, using 9 initial-condition ensemble members, have been done for the stream 1 hindcast period 1991-2001, initialised each 1st of May and November, respectively. Furthermore, a first common set of 9-member 10-year long integrations for two contrasting decades has been run. These simulations started 1st of November 1965 and 1994.

This section first summarises developments in methodological techniques (3.1), initialisation procedures (3.2) and system assembly (3.3), before discussing the actual results of stream 1 simulations in 3.4.

3.1 Development of techniques to represent of modelling uncertainties in ensemble predictions

The multi-model ensemble forecast system for s2d predictions, see 3.3 “Assembly of a multi-model ensemble prediction system”, has been successfully installed at ECMWF’s IBM supercomputer. It has capabilities to run, in addition, perturbed parameterisations and stochastic physics. Unified output and archival routines were developed and documentation is available at http://www.ecmwf.int/research/EU_projects/ENSEMBLES/index.html.

Climate models for all time scales usually use bulk-formula parameterisations of processes at the unresolved scales of motion. However, the resulting dissipation of near grid-scale motion prevents scale interactions, which can lead to significant errors in our current models simulating even the large-scale circulation patterns correctly. A new approach to ensemble climate prediction based on sampling stochastic parameterisation uncertainties has been developed and applied for one of the coupled models, namely the IFS/HOPE model (ECMWF). The **C**ellular **A**utomaton **S**tochastic **B**ack**S**cattering (CASBS) scheme (*Shutts, 2005*) introduces streamfunction perturbations on the near-grid scale and is motivated by the notion that a fraction of the dissipated energy will scatter upscale and inject kinetic energy into the resolved flow. For this purpose, a cellular automaton is utilised to generate a spatially and temporally correlated pattern that is weighted by the flow-dependent dissipation rate from numerical dissipation and friction from deep convection and gravity/mountain wave drag.

CASBS version 1.0 has been tested in s2d hindcast simulations for the 1991-2001 period, based on start dates in November and May, using IFS/HOPE with the atmospheric model cycle CY29R2. For the decadal simulations starting in November 1965 and 1994 a revised version of CASBS (v1.1) was used, while at the same time the latest changes to the atmospheric IFS cycle (CY30R1) have been introduced in the coupled IFS/HOPE system. The forecast skill of these preliminary simulations compared to other approaches to model uncertainty (multi-model and perturbed physical parameters) has been evaluated and documented, see “Assembly of a multi-model ensemble prediction system” below and ENSEMBLES Milestone M1.2. Work has been carried out to further develop the stochastic physics scheme. An updated version CASBS (v1.2) is currently being tested in the medium range ensemble forecasting system at ECMWF.

An s2d system based on perturbed physical parameters (PPE, **P**erturbed **P**hysics **E**nsembles) was developed by the Met Office/Hadley Centre and has been implemented on the ECMWF computer. The s2d PPE system, hereafter referred to as DePreSys_PPE (**De**cadal Climate **Pre**diction **Sys**tem_PPE), employs a subset of perturbed physics HadCM3 model versions used in the METO-HC centennial prediction system (see section 4 and *Collins et al., 2006*), but designed to produce s2d hindcasts. It differs from the original DePreSys in several ways. The main changes are: 1) modelling uncertainties are sampled through the application of multiple perturbations to atmosphere model parameters; 2) flux adjustments have been applied at the atmosphere-ocean interface; 3) a fully interactive sulphur cycle scheme has been implemented replacing the partial sulphur cycle scheme used previously; and 4) the external radiative forcing agents (trace gases, ozone, aerosols) have been updated with recent data sets.

The system consists of eight member perturbed physics parameter versions, picked to represent a wide range of climate sensitivities (2.6 – 7.1°C) and a wide range of ENSO amplitudes (Nino3.4 SST anomaly standard deviation ranging from 0.5 – 1.2°C), plus one standard parameter HadCM3 version. In each model version, flux adjustments are applied to correct for local SST errors and to prevent model drift. The earlier perturbed initial condition version of the METO-HC s2d system, which does not employ flux adjustments, is available for comparison. This version is hereafter referred to as DePreSys_Orig.

An anomaly assimilation technique was used to generate initial conditions for the PPE integrations. ERA-40 provided prognostic variables for the assimilation of the atmosphere, whereas monthly salinity and temperature analyses were used for the ocean (*Smith et al., 2006*). Forecast anomalies for each model version are produced by removing the relevant model climatology, obtained from an independent simulation of 20th century climate.

A small set of perturbed initial condition ensemble hindcasts for May start dates for years 1991-92 and November start dates for years 1993-94 have been completed as of now, using the version of HadCM3 with standard parameter settings, and including the flux adjustment and sulphur cycle updates indicated above. These experiments are referred to as DePreSys_ICE. These experiments will later be repeated and extended to a wider range of start dates using improved initialisation, and will eventually contribute to the multi-model ensemble.

3.2 Development of improved methods for representation of initial condition uncertainties in s2d ensemble predictions with a special focus on the ocean modules

At the start of the ENSEMBLES project, significant efforts were made to examine the ocean hindcast sets that were generated for the EU FP5 project ENACT to find the best ocean analysis schemes for coupled climate model hindcast production. In the following, we summarise the efforts made during the first 24 months of ENSEMBLES towards improved initialisation strategies in the ocean for the coupled s2d predictions.

At the METO-HC an extended ocean *in situ* dataset for 1958-2004 of all available ocean observations (now including over 7 million observations) was provided in May 2005. It is a significant upgrade since the ENACT version as the previously detected warm bias in the XBT observations due to double-correction and errors associated with the background climatology had been reduced and the criteria for the rejection of salinity observations were improved (*Ingleby and Huddleston (2006*). The data have been made widely available to all partners; see http://www.hadobs.org.

In four of the coupled models contributing to the s2d multi-model ensemble, model uncertainties in the initial state are represented through an ensemble of nine different ocean initial conditions. ECMWF provided an improved set of wind stress and SST perturbations to be used by the partners. A combination of four SST and two wind stress perturbations was used to construct a total of eight perturbed initial conditions plus one unperturbed condition. Two documents describing how the perturbations have been generated, and the impact of the new wind stress perturbations on the ECMWF ocean analysis system, have been prepared and can be downloaded at http://www.ecmwf.int/research/EU_projects/ENSEMBLES/exp_setup/ini_perturb/index.html. These perturbations were used at ECMWF to produce ocean analyses for the IFS/HOPE model covering the stream 1 period 1991-2001 plus 1965 and 1994.

In order to produce initial conditions for the ENSEMBLES s2d hindcasts using the GloSea model, the wind stress and SST perturbations provided by ECMWF were combined with GloSea data assimilation development and the new set of ocean observations, described above, resulting in a new three member ocean analysis ensemble for 1987-2001 and 1961-1966. The ocean model analysis scheme was tuned to improve the mean state and variability. The most significant impact in the sub-surface ocean analysis has been the reduction the spurious salinity variability in the tropical Atlantic, a problem in the previous version caused by assimilating temperature observations only at the Pirata moorings. The improvement has been achieved through a combination of using water-mass preservation (*Troccoli and Haines, 1999*) and the assimilation of salinity observations. Careful calibration of the ocean ensemble perturbations was made to ensure that subsurface intra-ensemble anomalies were representative of known uncertainties.

CERFACS provided a new set of ocean analyses based on the variational assimilation system developed for ENACT. These analyses have been used for the ocean initialisation in the stream 1 simulations for the period 1991-2001. Perturbed sets of ocean model analyses and simulations are obtained by perturbing the surface wind stress, SSTs, and observations used for assimilation. The perturbations for the surface forcing fields were those provided by ECMWF, as described above, while the perturbations for the observations are defined so that they have statistical properties consistent with the observation-error covariances used in the assimilation algorithm. Ensemble experiments with 16 members have been done over the period 1987-1990. The ensemble method is being used to provide flow-dependent estimates of the background error variances in a 3D-VAR assimilation system. These new ensemble variance estimates are compared with the currently used parameterized variances (weakly flow-dependent) and the best system will be further used in ENSEMBLES.

During ENACT KNMI developed a prototype ensemble Kalman filter (EnKF, **En**semble **K**alman **F**ilter) data assimilation system for the MPI-OM1 ocean model. System documentations is available in *Leeuwenburgh (2005a; 2005b)*. It was concluded that the analysis scheme of the KNMI EnKF set-up was promising, because of the use of a robust analysis scheme, a rather realistic data constraint error model and the capability to assimilate temperature and salinity. It has therefore been decided to set up a comparable system for the OPA ocean model based on the MPI-OM-1 results and an analogue analysis scheme for OPA has been implemented in the ECMWF environment. EnKF systems result in larger ensembles than can be handled in hindcast runs in ENSEMBLES; therefore, a method is needed to reduce the size of the ensembles. A “tubing” method for ensemble reduction in the EnKF system has been tested. In this method, the ensemble is characterised by its mean and by several tubes extending in state space so that they reach the members most remote from the ensemble mean. Here only the ensemble member closest to the ensemble mean and the extreme members of each tube are considered.

The previous data assimilation system at INGV has been upgraded through an implementation of seasonally dependent (versus stationary) bivariate EOFs computed at each model grid point. A paper on this work is under review (*Bellucci et al., 2006*). The system has been adapted to the ocean model to be used in ENSEMBLES stream 2 simulations. The new *in situ* ocean dataset, produced by METO-HC and described above, has been implemented and a global ocean analysis covering the period 1957-2001 was produced with the latest system. A comparison of the upper 300 and 3000 m heat content in the new analysis and the one produced using the previous data base of ocean subsurface temperature and salinity revealed that the largest differences are found in the equatorial region and in the southern hemisphere.

IfM-Kiel is working towards a multivariate version of the IfM 3D-VAR data assimilation system for the ocean. Studies on the MPI-OM1 ocean model initialisation revealed large initial condition errors in the North Atlantic. In particular, the strength of the meridional overturning circulation and its variability is far too strong. These errors appear related to the nudging of full SST as opposed to anomalies. However, these experiments demonstrate that the nudging strategy may indeed provide a method for initialising the thermohaline circulation, and future experiments are planned assimilating anomalies instead of full SSTs.

3.3 Assembly of a multi-model prediction system

A multi-model initial-condition ensemble system for climate prediction on seasonal, interannual and decadal time scales has been developed. It is installed on a single supercomputer at ECMWF and allows for efficient implementation of computer-intensive ensemble experiments, while ensuring model diversity. The ENSEMBLES s2d multi-model system is based on the experiences of the FP5 DEMETER system and the operational multi-model seasonal forecast system EUROSIP installed at ECMWF.

Since the end of the DEMETER project, a significant amount of work has been devoted to upgrade many of the climate model components and to port the models to the newer supercomputer architecture. The new multi-model initial-condition ensemble is currently built from five coupled climate models, namely IFS/HOPE (ECMWF), ARPEGE4.5/OPA (CNRM), ARPEGE/OPA (CERFACS), GloSea (METO-HC), and ECHAM5/MPI-OM1 (IfM-Kiel). Simulations from the METO-HC DePreSys_ICE system to contribute to the multi-model initial-condition ensemble are currently under way.

In the following, the developments of the individual coupled models, which were used to perform the common set of s2d stream 1 integrations, are discussed. Table 1 at the end of this report provides as brief summary of the contributing models, whereas further details on the multi-model assembly can be obtained from the ENSEMBLES Deliverable D1.4 (delivered month 18), available from http://www.ecmwf.int/research/EU_projects/ENSEMBLES/gen_info/deliverables.

The METO-HC GloSea forecasting system as initially developed in DEMETER has been further enhanced to include the time varying evolution of greenhouse/trace gases, ozone concentrations, volcanic aerosols and the interannual solar cycle. In addition, a method to correct for erroneously warm air temperatures in the ECMWF ERA-40 re-analysis (*Allan et al., 2004*) and their impact on sea-ice simulations in the ocean-only model has been devised.

The ARPEGE/OPA model at CNRM uses the most recent cycle 24T1 of ARPEGE-IFS version 4.6 with improvement in the vertical diffusion and the gravity wave drag, and increased vertical resolution with 91 levels with an extension in the stratosphere and an improved ozone scheme for the atmosphere and OPA8.2 as the ocean model. Detailed information about the model dynamics and physics can be found in http://www.cnrm.meteo.fr/hiretycs/div/arp4cae.pdf. OPA8.2 uses a different grid than the ORCA model in DEMETER and a free surface dynamics. Both atmosphere and ocean are coupled with the GELATO sea-ice model. The initial atmospheric state has been improved by using a nudging 44-year simulation (ARPEGE driven by ERA-40 6h by 6h).

CERFACS has run the DEMETER system initialised with new initial conditions issued from an ocean reanalysis using an early version of the 3D-VAR system developed in ENACT, see discussion in 3.2 above.

At ECMWF the latest version of the IFS atmospheric model component (CY29R1) coupled to the HOPE ocean model has been used. For the initialisation of the atmosphere, perturbations based on singular vectors have been applied in a similar way as in the operational medium range ensemble forecasts. As boundary forcings the evolution of greenhouse gases, a climatological annual cycle of five types of aerosol and interannual solar activity have been used. The model version has the capability to run in stochastic physics mode, that is using the newly developed stochastic backscatter parameterisation CASBS.

IfM Kiel used the latest version of the Max-Planck-Institute for Meteorology (MPIfM) climate model to perform the s2d hindcasts. Ocean and atmospheric initial conditions are generated from three coupled runs for the period 1950 until present, in which model SSTs are strongly damped to observed (NCEP) SSTs. As boundary forcings the evolution of greenhouse gases, natural and anthropogenic sulphate aerosol, and interannual solar activity have been used. The effect of volcanic aerosols is also included by varying optical depth.

Further to these individual fixed-parameter model developments towards a multi-model ensemble, the METO-HC is also developing perturbed physics and perturbed initial condition ensemble systems, both based on their orginal DePreSys System for interannual and decadal prediction. The new systems, denoted here as DePreSys_ICE, and DePreSys_PPE are part of the ENSEMBLES techniques to represent modelling uncertainties and will eventually also contribute to the multi-model ensemble.

The DePreSys system is based on the HadCM3 climate model, and was originally developed prior to ENSEMBLES. The updated versions used here are based on a version of HadCM3 including an enhanced representation of the atmospheric sulphur cycle and flux adjustments to restrict the development of regional biases in sea surface temperature and salinity (*Collins et al.*, 2006). In order to create initial conditions for the hindcasts, HadCM3 was run in assimilation mode from December 1989 to November 2001. During this integration, the atmosphere and ocean were relaxed towards analyses of observations, which were assimilated as anomalies with respect to the model climate in order to minimise climate drift when the assimilation is switched off. The assimilation integration was itself started from an initial state taken from a simulation of 20th century climate including variations in radiative forcing derived from observed changes in well-mixed trace gases, ozone, sulphate aerosol, total solar irradiance and volcanic aerosol. The forcings in the assimilation and hindcast integrations are the same, switching from values based on observations to values based on the SRES A1B emissions scenario after the year 2000. One exception is that during hindcasts solar irradiance was estimated by repeating the previous 11-year solar cycle and volcanic aerosol was specified to decay exponentially from the initial value with a time scale of one year. This was done to avoid assuming accurate knowledge of future variations in these forcings.

In order to ensure coordinated data encoding, archiving and efficient dissemination, common lists of output variables and formats for the s2d stream 1 integrations have been agreed upon among the partners as part of the s2d activities in RT2A WP2A.4. The list of variables, which includes daily, monthly and model-level data, is available from http://www.ecmwf.int/research/EU_projects/ENSEMBLES/data/common_variables.html. The conventions to encode and archive the atmospheric and oceanic output are documented at http://www.ecmwf.int/research/EU_projects/ENSEMBLES/data/. All stream 1 s2d simulations have been archived in MARS, the Meteorological Archival and Retrieval System at ECMWF, and are already available to the ENSEMBLES partners. A summary of the full set of data available can be found at http://www.ecmwf.int/research/EU_projects/ENSEMBLES/table_experiments/index.html. Furthermore, it is planned to disseminate the s2d data in GRIB and NetCDF formats via a public data server including an OpeNDAP access and a link to the KNMI Climate Explorer (http://climexp.knmi.nl).

3.4 Results

This section summarises some selected key properties of the modeling systems constituting version 1 of the s2d ensemble prediction system obtained from the stream 1 simulations. The analysis of these integrations is still ongoing and it is expected that the preliminary results will be updated and further substantiated during the future course of the project. Please note that a preliminary assessment of the various ensemble methodologies was made for the ENSEMBLES milestone report M1.2, delivered in February 2006, and a further update of the work is going to be prepared for deliverable D1.8, due month 30. A comprehensive set of diagnostics for the seasonal-to-interannual and the decadal simulations, which is updated regularly, is available on the RT1 website under http://www.ecmwf.int/research/EU_projects/ENSEMBLES/results/index.html.

An illustration of the performance of these forecast systems on seasonal and interannual timescales over the tropical Pacific is given in Figure 1. It displays for the multi-model, perturbed physics and stochastic physics ensembles the root mean squared error (RMSE) of the ensemble mean and the ensemble standard deviation of the SST over the Nino3 (5°N-5°S,150°W-90°W) region over lead time for the May and November start dates for the stream 1 period 1991-2001. The multi-model ensemble as shown is formed from four of the models. Data from one model could not be included in the analysis so far as their archiving is not completed by the time of writing. The data for the perturbed physics ensemble are generated by the DePreSys_PPE system and will be replaced by an updated version of that system in the future. In this preliminary analysis the multi-model ensemble has the smallest RMS error for lead times up to 10 months and both start dates in May and November (Fig. 1a and b). At the same time, the spread of the ensemble matches the error reasonably well, which can be considered as desirable feature of a good prediction system. The multi-model also performs remarkably better for the analysed variable than a simple statistical forecast based on persistence. The perturbed physics ensemble has a larger RMS error than the ensemble spread for all lead times in the May start date simulations (Fig. 1b). For the November start dates, however, they match better and get close to the corresponding multi-model scores for lead times above 4 months (Fig. 1d). In the early months of the November forecasts, DePreSys_PPE seems to perform worse than persistence. The stochastic physics simulations show a reduction in RMS error and increase in the ensemble spread compared to the control forecast done with the same forecast model for the May start dates (Fig. 1e). This nicely demonstrates one of the benefits of this approach in generating ensemble spread. In November, however, the impact can hardly be distinguished from the control scores (Fig. 1f). A test on the impact of the different ensemble sizes in the three systems revealed that the results are qualitatively robust. A wide collection of similar analyses carried out for different variables, regions and the anomaly correlation coefficient over lead time can be found in http://www.ecmwf.int/research/EU_projects/ENSEMBLES/results/index.html.

The benefits of the stochastic physics approach are not only found in the reduction of the ensemble mean RMSE and increase in ensemble spread. In addition, systematic errors, computed as differences between the model and observed climatologies, of different variables are reduced. As an example, Figure 2 shows the impact of CASBS on global seasonal mean (December-February, DJF) precipitation and 500-hPa geopotential hight fields (Z500). While the control IFS/HOPE integrations have an excess of precipitation over the tropical oceans (Fig. 2a), the CASBS version reduces this wet bias as well as other systematic errors over the continents (Fig. 2b). 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. The lower panels of Figure 2 show the systematic errors in Z500. The control simulation shows large errors over the North Pacific (Fig. 2c), a long-standing problem in simulating the extratropical circulation. The new stochastic physics scheme seems to be able to reduce this error, though not fully eliminate it (Fig. 2d). Note that, as a result of on-going development, is it expected that in future versions of CASBS the strong bias over the southern polar regions will be significantly reduced. For the boreal summer period, the climate also improves in terms of winds and precipitation, in particular for the Southeast Asian monsoon.

Consistent with the systematic reduction of the circulation errors over the North Pacific in CASBS, the model performance in terms of simulating blocking has also improved. Figure 3a shows the mean Northern Hemisphere blocking frequency computed with the Z500-based *Tibaldi and Molteni* (1990) index for DJF. 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. This is in agreement with previous experiments carried out with preliminary versions of the stochastic physics scheme at ECMWF (*Palmer et al., 2005*; *Jung et al.*, *2005*). The total wave number spectrum of the Z500 anomalies for DJF for the control and stochastic physics integrations are shown in Figure 3b. CASBS improves the energy spectrum for almost all wave numbers, most remarkably for the large scales (wave numbers 3-8 and above wave number 12), which provides an indication for the ability of the stochastic backscatter scheme to enhance the upscale energy transfer from the smaller to the larger scales of atmospheric motion.

Turning to the perturbed physics ensembles, a comparison of the ENSO-related s2d hindcast skill shows that the DePreSys_PPE system is less skilful than the original system, referred to here as DePreSys_Orig (Fig. 4). A persistence forecast beats the DePreSys_PPE system at short lead times of up to two months, though not at longer lead times. Reasons for the reduction in skill at short lead times are currently under investigation. One possibility is that there is a mismatch between the PPE model versions and the ocean analyses, as the covariance statistics used in the ocean analyses has been derived from the standard parameter setting run without flux adjustments. It has recently been found that the assimilation integrations for DePreSys_PPE employed the solar and volcanic forcing datasets intended for the hindcasts (rather than the correct dataset of observed historical variations), which may have contributed to initialisation problems. In addition, it is known that flux-adjustments (used in DePreSys_PPE but not in DePreSys_Orig) influence the intrinsic behaviour of the ENSO mode in HadCM3, resulting in a weakening of the coupling of the atmosphere and subsurface thermocline in the east Pacific (*Spencer et al., 2006*). Further assessment and development of the perturbed physics ensembles is needed, in particular to resolve the above initialisation issues, and hence provide an improved basis for assessing the potential for improving the predictions of both initial condition and multi-model ensembles with this technique.

Figure 5 shows decadal hindcasts of global and European mean annual mean near-surface temperature for the 1994 start date, for several multi-model ensemble members, plus the stochastic and perturbed physics systems. The absolute value of this quantity for the forecast ensemble is shown. It would be preferable to remove the effects of climate drift, in order to isolate any potentially predictable long-term anomalous signal associated with the initial conditions. However, this cannot be done here, as we do not have available forecast climatologies as a function of lead time. Such climatologies are used in seasonal simulations (e.g., DEMETER, *Palmer et al., 2004*), but it would be prohibitively expensive to generate them on the decadal time scale, at least for the high resolution models which form the multi-model ensemble. On the global scale (Fig. 5a), one of the models has a significant model-climate drift, whereas the other hindcasts produce reasonable trends and variability. Research is needed to assess whether these individual hindcast experiments have useful forecast information in a multi-model sense.

Figure 6 shows the Nino3 SST anomalies for all the multi-model ensemble members (also for the November 1994 case) 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. This highlights the non-stationary performance of the systems and suggest that other systematic errors not being sampled may be dominant (e.g. no coupled model is adequately representing all the key climate process including the quasi-biennial oscillation, the Madden-Julian oscillation and the El Nino-Southern Oscillation) and also that the hindcast sample is statistically quite small.

Since DePreSys_PPE is initalised from anomalies, forecast variables are more or less free of climate drift, and can therefore be expressed as anomalies formed relative to a previously available long-term climate simulation. Figure 7 shows the performance in predicting the annual global mean surface air temperature anomalies is encouraging at short lead times (up to five years), as it compares reasonably well with the hindcasts of the DePreSys_Orig system available prior to ENSEMBLES (upper panel, Fig. 7). However the increased RMSE values at longer lead times reflect the relatively larger warming bias (lower panel, Fig. 7) predicted by the DePreSys_PPE system in this particular decade. This relatively increased rate of decadal warming may not necessarily indicate a fundamental error in the system, but could just reflect the faster warming nature of the contributing member versions. The averaged climate sensitivity (CS) of the slab-model versions corresponding to the flux adjusted transient versions included in the s2d PPE system is 4.1°C, whereas the standard parameter setting version has CS of 3.5°C. Different levels of radiative forcing from the updated forcing inputs and interactive sulphur cycle may also contribute to a faster warming rate.

For interannual to decadal predictions, a key question is whether the initialisation of models with analyses of observations improves skill at extended lead times, just as it does for seasonal predictions. In the lower panel of Figure 7, we assess this by comparing the averaged bias of global mean temperature predictions from the DePreSys_PPE s2d simulations (red curve) against that from the long-term historical climate simulations (initialised from pre-industrial conditions) from which the s2d ensemble was derived (blue curve). For the first year, the DePreSys_PPE bias is smaller, as would be expected, given the initialisation of observed values. In later years the DePreSys_PPE runs show a steadily increasing positive bias (as discussed above), however it is encouraging that this bias remains consistently below that found in the historical simulations. This suggests that the initial conditions continue to influence the predictions throughout the decadal period of the hindcasts, illustrating the potential for improved long-term predictions associated with initialisation from observations. Further work is underway to understand these results in greater detail.

Another positive aspect of the PPE results is that the DePreSys_PPE forecast anomalies produce a wider spread (upper panel, Fig. 8) relative to the corresponding initial condition ensemble (the new DePreSys_ICE system), while maintaining a comparable averaged forecast bias at all lead times. This is a desirable quality of any ensemble prediction system, as it increases the possibility of enveloping extreme observed conditions. However, note that these features may not be statistically significant as only four start dates have been used in the spread and bias calculations.

4. Centennial predictions

The principal difference between making predictions on centennial time scales and making predictions on seasonal-decadal predictions (discussed above) is the lack of independent verification cycles over which the ensemble predictions may be assessed. There is no way of directly assessing, by comparison with observations, whether the ensemble spread is too wide, too narrow, or in some way biased. There is as yet no simple way of adjusting the ensemble generation technique, or post-processing the ensemble output, to produce a spread which is consistent with previous forecasts (though RT1 plans to pioneer work in searching for links between unverifiable centennial predictions, and verifiable shorter term predictions, during the second phase of ENSEMBLES) . Yet the consequences of producing pdfs which are biased or which do not encompass the full range of possibilities may be great. Mitigating action against climate change may not be perceived to be required or adaptation strategies may be ill-formed.

To avoid such possibilities, there are a number of desirable attributes that centennial-time scale ensembles should aim towards. While it is difficult to express the attributes mathematically, they may be described in generic terms thus.

- They should realistically sample a wide range of the uncertain processes, feedbacks and climate forcings relevant to centennial time-scales.

- They should be relatively large to sample potential interactions between difference sources of uncertainty.
- Where predictions are required on regional scales or for “exotic” variables such as extreme rainfall, the models used to generate the ensembles should be able to simulate those variables in a physically consistent way.

The challenge in centennial ensemble prediction is to generate such ensembles making optimal use of resources.

Ensemble generation is, however, only the first part of the process of making probabilistic predictions of future climate change. The ensemble output must be manipulated to produce probability distribution functions (pdfs) of user-relevant variables. Many techniques, which utilise output from a hierarchy of models, have been proposed and applied to probabilistic estimates of both equilibrium global temperature change (e.g. *Murphy et al. 2004*; *Piani et al. 2006*; *Frame et al., 2006*) and transient global temperature change (e.g. *Stott et al., 2006*; *Harris et al., 2006*). The basic approach is to estimate some prior distribution of the forecast variable of interest and then to use observations to constrain the prediction. The constraint works by adjusting the prior probabilities according to a comparison with some observed aspect of climate and/or climate change.

Because studies use a range of different prior assumptions about prediction variables, and because of the range of different observational constraints employed, different studies have produced different probabilistic predictions of the same variable. In the particular case of equilibrium change, such differences are exacerbated because of the tendency for predictive pdfs to differ most in the region of the upper tail; that is the tail of the distribution, which determines the probability of the most extreme or “dangerous” climate change. In estimates of the pdf of transient global mean temperature change, this problem is less severe (*Frame et al., 2006*) but nevertheless, for predictions of regional climate change and for predictions of more exotic variables, different prior assumptions and the application of different observational constraints are likely to have a leading-order influence on the predictive pdfs. Thus, the sensitivity of predictions to assumptions and the comparison of pdfs produced using diverse techniques is a critical issue, and will be the subject of future activities in ENSEMBLES.

4.1 Progress in the First 2 Years of ENSEMBLES

ENSEMBLES has adopted a two-stream approach to centennial ensemble generation:

Perturbed Physics Ensembles: The perturbed physics approach involves perturbing parameters and, in some cases, switching between existing options, in a single model structure. The Hadley Centre model structure (which we denote HadCM3) has been extensively used to generate large perturbed physics ensembles of climate change simulations under idealised equilibrium and transient forcing scenarios (*Murphy et al., 2004*; *Collins et al., 2006*). The use of distributed computing techniques and the generous donation of spare CPU time on home computers by members of the public has facilitated the generation of some very large equilibrium-experiment ensembles under the climateprediction.net project (*Stainforth et al., 2005*) and transient simulations are underway. In addition, preliminary perturbed physics experiments with the EGMAM model have been performed. A preliminary assessment of the perturbed physics approach is given in ENSEMBLES Milestone M1.3.

Multi-Model Ensembles: While the perturbed physics approach allows the production of large ensembles, and for the specification of prior assumptions about model parameters, the full range of feedbacks and patterns of regional climate change may not be adequately sampled. The effects of this “structural uncertainty” may be assessed by examining existing and new models developed as part of the ENSEMBLES project and elsewhere, for example the archive of model output collected as part of the IPCC Fourth Assessment Report. Progress on model development within ENSEMBLES is reported in Major Milestone MM1.1.

Comparison of the output from both ensembles reveals some consistent behaviour for both global mean fields (Fig. 9) and even at large regional scales (Fig. 10). Such comparisons also highlight differences, which must be recognised when using the ensembles to make probabilistic predictions.

4.2 The First-Generation Centennial Ensemble Prediction System

For predictions on centennial time scales, the first-generation ENSEMBLES system comprises:

- *Hadley Centre model perturbed physics ensembles*: Relatively large ensembles of the HadCM3 model in which perturbations are made to uncertain parameters and schemes in the atmospheric component have been performed. Experiments simulate both equilibrium and transient conditions under pre-industrial and future concentrations of greenhouse gases and other forcing agents. Further simulations with perturbations to parameters in other model components will be performed by both the Hadley Centre and the climateprediction.net project in the final three years of ENSEMBLES.
- *An experimental perturbed physics ensemble:* This has been performed with the physical EGMAM model from FUB. Perturbations have been made to atmospheric-component parameters and an initial assessment has been made of present day climate simulations. This study will be extended to look at climate change scenarios.
- *The multi-model ensemble*: This comprises a range of models of varying complexity, from physical models of the atmosphere and ocean to more complex models with representations of Earth-systems processes such as aerosol chemistry and the carbon cycle. In order to sample a wide range of possible model structures, it is important that output from models that are not formally part of ENSEMBLES, are used. Restricting attention to ENSEMBLES models may result in an underestimate of structural uncertainty (e.g. Fig. 9 and 10).

The system thus comprises information from state-of-the-art perturbed physics ensemble and multi-model approaches. It represents a basis from which we may produce a wide range of probabilistic predictions of climate variables at global and large regional scales and will aid in the understanding of uncertainties in centennial climate predictions in other RTs.

5. Summary and outlook

5.1 Seasonal to Decadal Predictions

The ENSEMBLES first generation multi-model initial-condition ensemble system for climate prediction on seasonal, interannual and decadal time scales has been developed. It includes three different methods to tackle the problem of model uncertainty: the multi-model ensemble technique, the perturbation of physical parameters and stochastic parameterisations of sub-grid processes. The system is successfully installed on the IBM supercomputer at ECMWF and allows for efficient implementation of computer-intensive ensemble experiments, while ensuring model diversity. Unified output and archival routines are provided.

In order to compare the three approaches mentioned above and to assess their relative merits in an s2d framework, a set of common simulations (stream 1) for all three of them was defined, performed and analysed. The coupled forecast were initialised using improved methods to represent initial uncertainty in the ocean and atmosphere.

In the preliminary analysis presented above, the multi-model ensemble has the smallest ensemble mean RMS error for lead times up to 10 months and both start dates in May and November. At the same time, the spread of the ensemble matches the error reasonably well. For the analysed variables (mainly SSTs in the tropics), the multi-model performs remarkably better than a simple statistical forecast based on persistence.

Benefits of the stochastic physics approach on seasonal time scales are found in reducing of the ensemble mean RMSE and increasing in ensemble spread. In addition, systematic errors in the forecasts of tropical precipitation, extratropical geopotential height and blocking frequency are found to be reduced.

First results of an s2d perturbed physics ensembles showed promising results. Further assessment of these simulations is needed, in particular to improve the initialisation, and hence provide a better basis for assessing the potential for improving the predictions of both initial condition and multi-model ensembles with this technique.

A first set of decadal hindcasts have been produced by all models. Preliminary quality assessments highlight the non-stationary performance of the systems and suggest that other systematic errors not being sampled may be dominant (e.g. no coupled model is adequately representing all the key climate process including the quasi-biennial oscillation, the Madden-Julian oscillation and the El Nino-Southern Oscillation) and also that the hindcast sample is statistically quite small. Research is needed to assess whether these individual hindcast experiments have useful forecast information in a multi-model sense. A key recommendation for future development of the s2d system is thus that coupled climate model performance should be enhanced and ways sought to account for missing areas of forecast uncertainty, whilst also continuing to develop the various streams of complementary ensemble methodologies.

For interannual to decadal predictions, a key question is whether the initialisation of models with analyses of observations improves skill at extended lead times, just as it does for seasonal predictions. Results from the perturbed physics ensemble integrations suggest that the initial conditions continue to influence the predictions throughout the decadal period of the hindcasts, illustrating the potential for improved long-term predictions associated with initialisation from observations.

It is planned to address these questions in the second stage of the ENSEMBLES s2d integration stream, called stream 2. This new set of hindcasts will not only take into account a larger sample of start dates per year (1st February, 1st May, 1st August, and November) for the seasonal and interannual integrations, but also cover a longer hindcast period (1960-2005) as well as more decadal runs. Details of the proposed stream 2 plans can be found at http://www.ecmwf.int/research/EU_projects/ENSEMBLES/exp_setup/stream2.html.

5.2 Centennial Predictions

Much of the work to date has been in solving many of the technical issues that have arisen in building the first generation system and in making preliminary predictions of global mean climate change under idealised forcing scenarios. The focus of years 3-5 of ENSEMBLES will be in both extending the system and applying it to produce regional predictions which may be used by stakeholders, including those in other Research Themes. In addition, work will be undertaken to evaluate and understand the predictions made by the system.

Plans for years 3-5 include:

- Extension of the perturbed physics approach to other components of the HadCM3 model. Specifically, perturbations will be made to parameters in the ocean component and the sulphur cycle component by both the Hadley Centre and climateprediction.net groups. If deemed feasible, perturbations to the carbon cycle component will also be attempted.
- Application of the perturbed physics approach to other modelling structures.
- Use of interim estimates of pdfs of regional climate change in assessing the impacts of climate change (Deliverable 1.7).
- Application and comparison of different techniques to produces probabilistic predictions of large-scale regional climate change that utilise the HadCM3 component of the first generation system. Comparison of predictions made using different approaches to specifying prior assumptions and utilising different observational constraints will be made.
- Accounting for the effects of structural uncertainty in probabilistic predictions by applying the Bayesian approach of *Rougier (2006)*. This will allow us to produce probabilistic predictions combining (for the first time) the information from perturbed physics and multi-model ensembles, and promises to be a major development relative to the current state of the art.
- Recommendations for the specification of post-ENSEMBLES centennial probabilistic predictions (Major Milestone, due month 60).

5.3 System Integration

The major advantage of s2d predictions compared to centennial forecasts, where verification *per se* is impossible, is that the quality of the s2d forecast systems can be estimated based on statistical analyses in sets of hindcast simulations. However, in view of aiming to investigate the prospects for moving toward a single seamless weather and climate prediction system in the future, assessments of the forecast performance on s2d time ranges can, and should, be linked to the evaluation of potential predictability on much longer time scales. For example, it is planned to extend some of the s2d runs with the perturbed physics scheme to a century ahead and compare them to the standard centennial experiments.

References

Allan, R.P., M.A. Ringer, J.A. Pamment and A. Slingo, 2004, Simulation of the earth's radiation budget by the European Centre for Medium-Range Weather Forecasts 40-year reanalysis (ERA40). *J. Geophys. Res.*, **109,** D18107, doi:10.1029/ 2004JD004816.

Bellucci A., S. Masina, P. Di Pietro and A. Navarra, 2006, Using temperature-salinity relations in a global ocean implementation of a multivariate data assimilation scheme. *Mon. Wea. Rev.* submitted

Collins, M., B.B.B. Booth, G. Harris, J.M. Murphy, D.M.H. Sexton and M.J. Webb, 2006, Towards quantifying uncertainty in transient climate change. *Clim. Dynamics*, **27 (2-3)** 127–147.

Frame DJ, Booth BBB, Kettleborough JA, Stainforth DA, Gregory JM, Collins M, Allen MR, 2005, Constraining climate forecasts: The role of prior assumptions. *Geophys. Res. Letts*., **32 (9)** L09702.

Harris G.R., Sexton D.M.H. , Booth B.B.B., Collins M., Murphy J.M., Webb M.J., 2006, Frequency distributions of transient regional climate change from perturbed physics ensembles of general circulation model simulations. *Clim. Dynamics,* **27 (4)**: 357-375.

Ingleby, B. and M. Huddleston, 2006, Quality control of ocean temperature and salinity profiles - historical and real-time data, *J. Marine Systems*, in press.

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.

Leeuwenburgh, O., 2005a, Implementation and testing of an Ensemble Kalman Filter assimilation system for the Max Planck Institute Ocean General Circulation Model, Technical Report TR-274, KNMI, available on request from KNMI.

Leeuwenburgh, O., 2005b, Validation of an Ensemble Kalman Filter system for OGCM initialisation assimilating temperature, salinity, and surface height measurements, *Mon. Wea. Rev.*, submitted.

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 and D.A. Stainforth, 2004, Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, **430**, 768-772.

Palmer, T.N., 2000: Predicting uncertainty in forecasts of weather and climate. *Rep. Prog. Phy.*, **63**, 71-116.

Palmer, T.N., 2001: A Nonlinear Dynamical Perspective on Model Error: A Proposal for Non-local Stochastic-Dynamic Parametrisation n Weather and Climate Prediction. *Quart. J. R. Meteorol. Soc.,* **127**, 279-304.

Palmer, T.N., A. Alessandri, U. Andersen, P. Cantelaube, M. Davey, P. Délécluse, M. Déqué, E. Díez, F.J. Doblas-Reyes, H. Feddersen, R. Graham, S. Gualdi, J.-F. Guérémy, R. Hagedorn, M. Hoshen, N. Keenlyside, M. Latif, A. Lazar, E. Maisonnave, V. Marletto, A. P. Morse, B. Orfila, P. Rogel, J.-M. Terres, M. C. Thomson, 2004. Development of a European multi-model ensemble system for seasonal to inter-annual prediction (DEMETER). *Bull. Amer. Meteorol. Soc.*, **85**, 853-872.

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.

Palmer, T.N. and R. Hagedorn, 2006, Predictability of Weather and Climate, Cambridge University Press.

Piani, C., D.J. Frame, D.A. Stainforth, M.R. Allen, 2005, Constraints on climate change from a multi-thousand member ensemble of simulations. *Geophys. Res. Lett.,* **32**, L23825.

Rougier, J.C., 2006, Probabilistic Inference for Future Climate Using an Ensemble of Climate Model Evaluations, *Climatic Change*, to appear.

Salas Melia, D., 2002. A global coupled sea ice-ocean model. *Ocean Modelling*, **4**, 137-172.

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

Smith, D., J. Murphy, S. Cusack, 2006, An objective ocean temperature and salinity analysis using covariances from a global climate model. *J. Geophys. Res*, submitted.

Spencer, H., R. Sutton, J.M. Slingo, 2006, El Nino in a coupled climate model: sensitivity to changes in mean state induced by heat flux and wind stress corrections. *J. Climate*, submitted.

Stainforth, D. A., T. Aina, C. Christensen, M. Collins, N. Faull, D. J. Frame, J. A. Kettleborough, S. Knight, A. Martin, J. M. Murphy, C. Piani, D. Sexton, L. A. Smith, R. A. Spicer, A. J. Thorpe, M. R. Allen, 2005, Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature*, **433**, 403-406.

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.

Stott P.A., Mitchell J.F.B., Allen M.R., Delworth T.L., Gregory J.M., Meehl G.A., Santer B.D, 2006, Observational constraints on past attributable warming and predictions of future global warming. *J. Climate*, **19 (13)**: 3055-3069.

Tibaldi, S. and F. Molteni, 1990, On the operational predictability of blocking. *Tellus*, **42A**, 343-365

Troccoli A. and K. Haines, 1999, Use of temperature-salinity relation in a data assimilation context. *J. Atmos. Ocean. Tech.*, **16**, 2011-2025.

Webb, M. J. , C. A. Senior, K. D. Williams, M. D. H. Sexton, M. A. Ringer, B. J. McAvaney, R. Colman, B. J. Soden, N. G. Andronova, S. Emori, Y. Tsushima, T. Ogura, I. Musat, S. Bony, and K. Taylor, 2006, On uncertainty in feedback mechanisms controlling climate sensitivity in two GCM ensembles. *Clim. Dynamics*, **27 (1)**: 17-38.

Weisheimer, A. and T.N. Palmer, 2005, Changing frequency of occurrence of extreme seasonal temperatures under global warming. *Geophys. Res. Lett.*, **32**, L20721, doi:10.1029/2005GL023365.

Tables

	CERFACS	ECMWF	CNRM	METO-HC DePreSys	METO-HC GloSea	IfM-Kiel
atmosphere component	ARPEGE	IFS	ARPEGE	HadAM3	HadAM3	ECHAM-5
resolution	T63 31 levels	T95 40 levels	T63 31 levels	2.5º x 3.75º 19 levels	2.5º x 3.75º 19 levels	T63 31 levels
atmosphere initial conditions	ERA-40	ERA-40	ERA-40	ERA-40 anomalies	ERA-40	coupled run relaxed to observed SSTs
reference	Déqué 2001	Palmer et al., 2004	Déqué 2001	Pope et al. 2000	Pope et al. 2000	Roeckner et al. 2003
ocean component	OPA 8.2	HOPE-E	OPA 8.2	HadCM3 OGCM	GloSea OGCM, based on HadCM3	MPI-OM1
resolution	2.0º x 2.0º 31 levels	1.4º x 0.3º-1.4º 29 levels	182 GP x 152 GP 31 levels	1.25º x 1.25º 20 levels	1.25º x 0.3º-1.25º 40 levels	1.5º x 1.5º 40 levels
ocean initial conditions	ocean analyses forced by ERA-40	ocean analyses forced by ERA-40	ocean analyses forced by ERA-40	coupled run relaxed to analyses of temperature and salinity anomalies	ocean analyses forced by ERA-40	coupled run relaxed to observed SSTs
reference	Delecluse and Madec 1999	Wolff et al. 1997	Madec et al. 1998	Gordon et al. 2000	Gordon et al. 2000	Marsland et al. 2003
ensemble generation	windstress and SST perturbations	windstress and SST perturbations	windstress and SST perturbations	lagged initial conditons and perturbed model parameters	windstress and SST perturbations	ensemble coupled initialisation runs

Table 1: Combinations of atmosphere and ocean models used by the modelling groups. The resolution of the models and the initialization strategy is outlined as well.

Figures

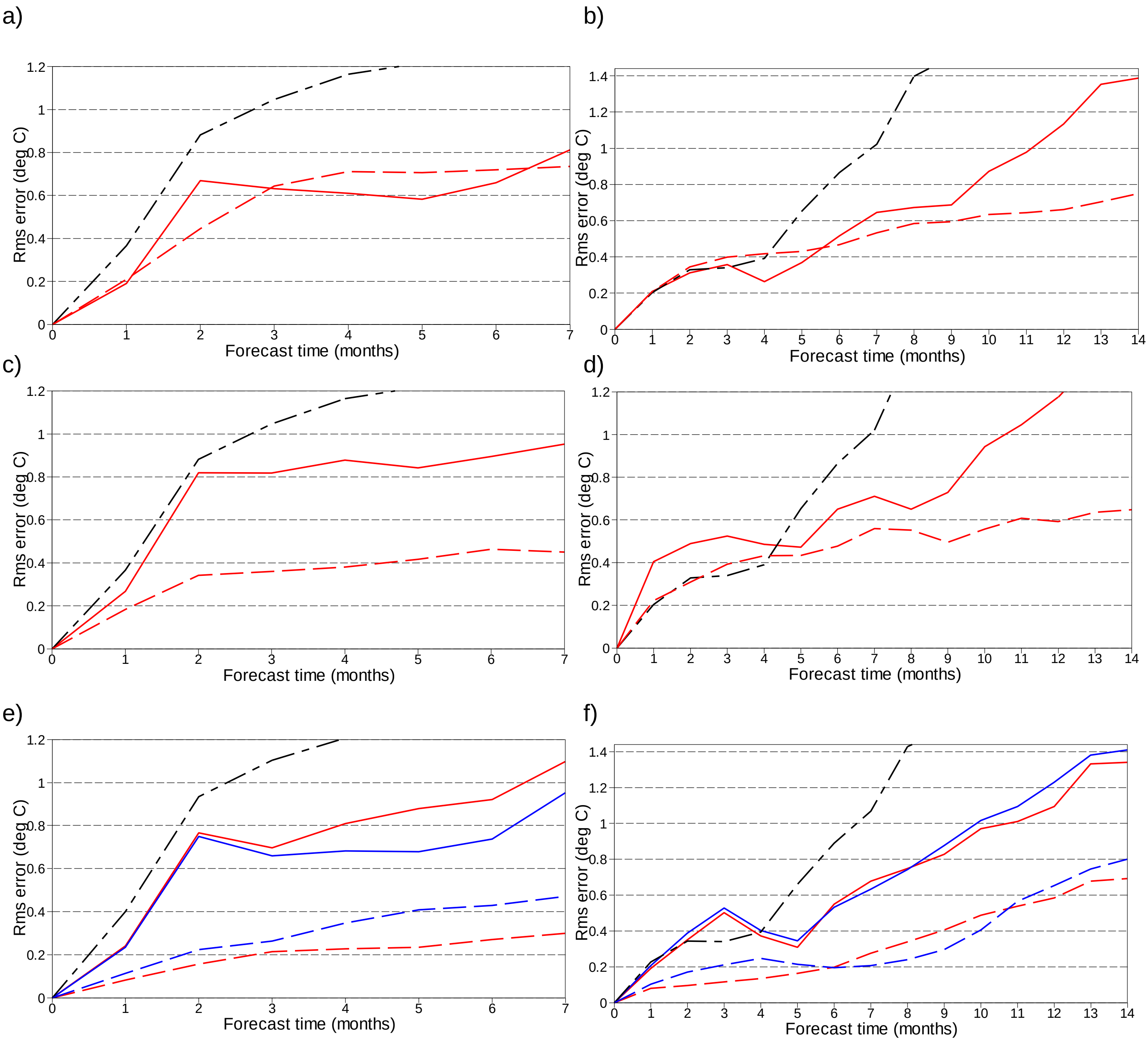


Figure 1: Seasonal-to-internanual RMS error of the ensemble mean (solid lines) and spread of the ensemble as measured by its standard deviation (dashed lines) over lead time for Nino3 SST hindcasts from 1991 to 2001 starting on May 1st (left) and November 1st (right). For comparison, the RMS errors of a persistence forecast are shown with the black dashed lines. Top row: multi-model ensemble consisting of the following models: IFS/HOPE control, GloSea, ARPEGE4.5/OPA and ECHAM5/MPI-OM1. Each of them has been run for an ensemble of nine initial conditions. Middle row: DePreSys_PPE perturbed physics ensemble based on nine ensemble members. Bottom row: IFS/HOPE control (red) and IFS/HOPE stochastic physics using CASBS1.1 (blue), each of them based on a nine-member ensemble.

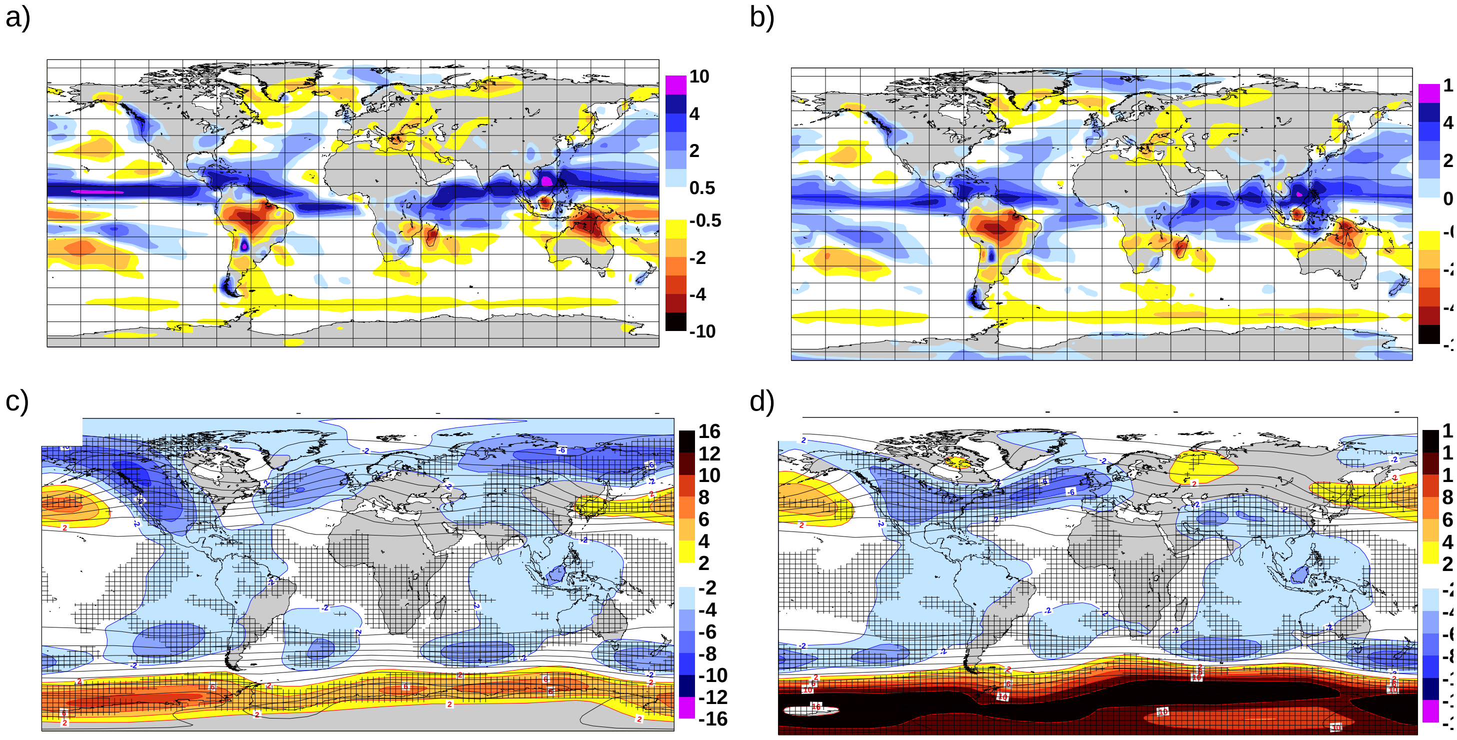
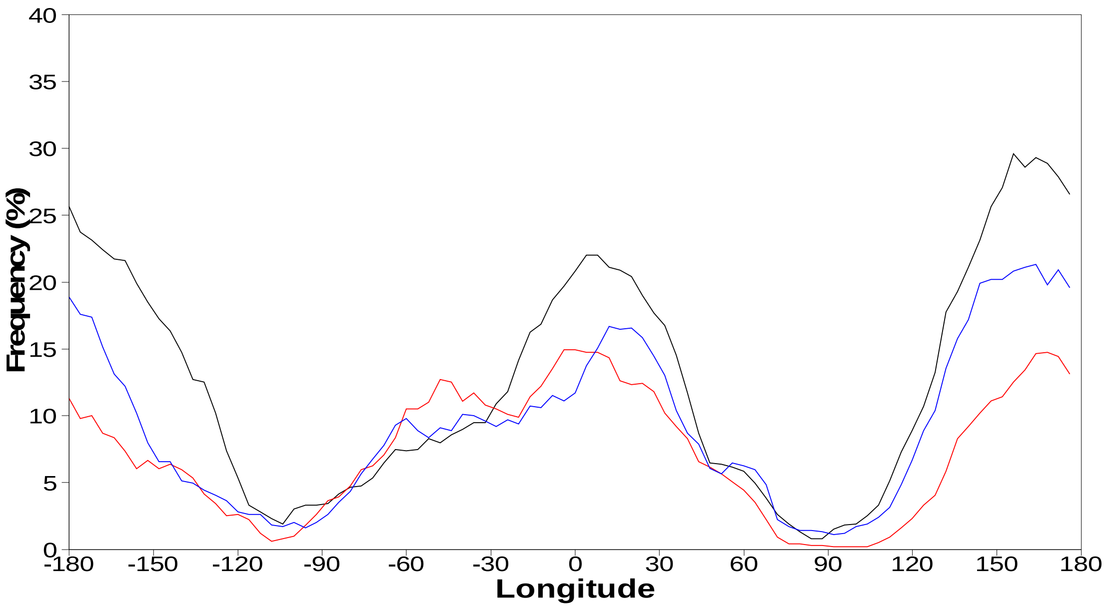


Figure 2: Seasonal mean (DJF) systematic errors in the control and stochastic physics simulations for stream 1 seasonal hindcasts from the 1991-2001 November start dates. a) IFS/HOPE control minus GPCP precipitation field (in mm/day). b) IFS/HOPE CASBS minus GPCP precipitation field (in mm/day). c) IFS/HOPE control minus ERA-40 Z500 field (in m). d) IFS/HOPE CASBS minus ERA-40 Z00 field (in mm).

a)



b)

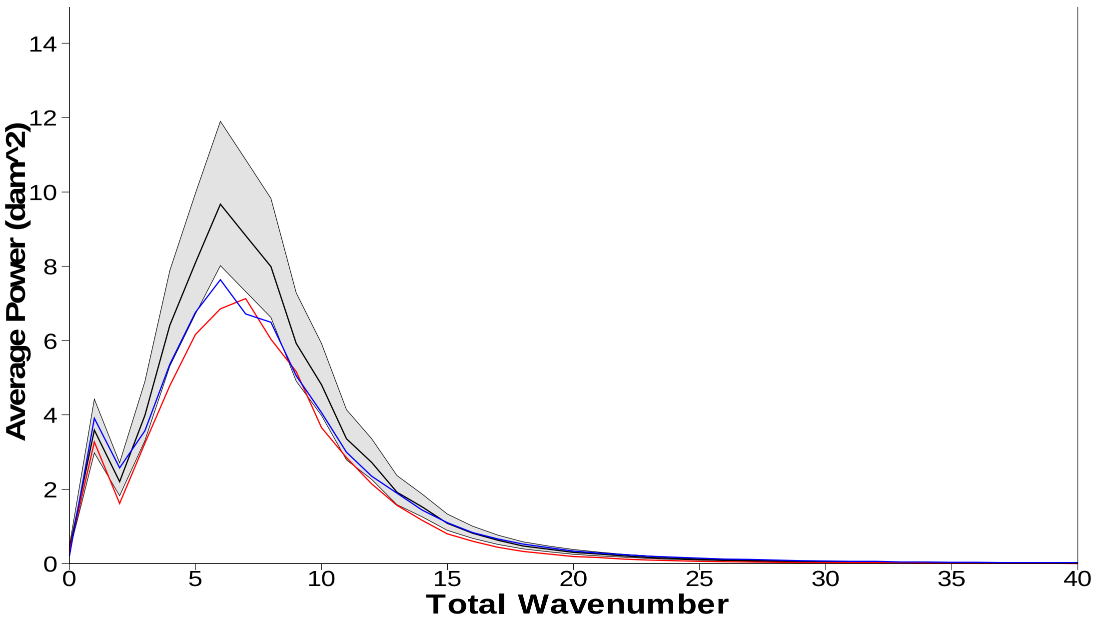


Figure 3: a) Blocking frequency after *Tibaldi and Molteni* (1990) during DJF for 1991-2001 November start hindcasts. b) Total wavenumber spectrum of the Z500 anomalies. IFS/HOPE control (red), IFS/HOPE stochastic physics using CASBS1.1 (blue) and ERA-40 (black). The grey area in panel b) indicates the uncertainty range in the ERA-40 data due to sampling uncertainty.

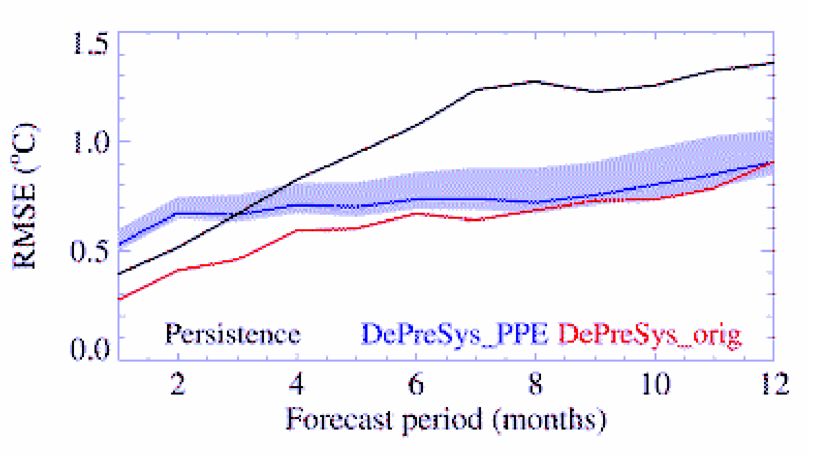


Figure 4: RMS error of DePreSys_PPE ensemble mean monthly hindcasts as a function of the lead time for Nino3 surface air temperatures, averaged over the 22 stream 1 hindcast experiments. Also shown for comparison is the error from the original DePreSys_Orig system for the same set of years (with 5-95% confidence ranges shaded). Note that the start dates for the DePreSys_Orig simulations, which were run prior to ENSEMBLES, were 1st of June and 1st of December, whereas they were 1st of May and 1st of November for the DePreSys_PPE experiments. DePreSys_Orig consisted of four ensemble members with initial dates lagged by one day, whereas DePreSys_PPE consists of nine ensemble members using perturbed parameters.

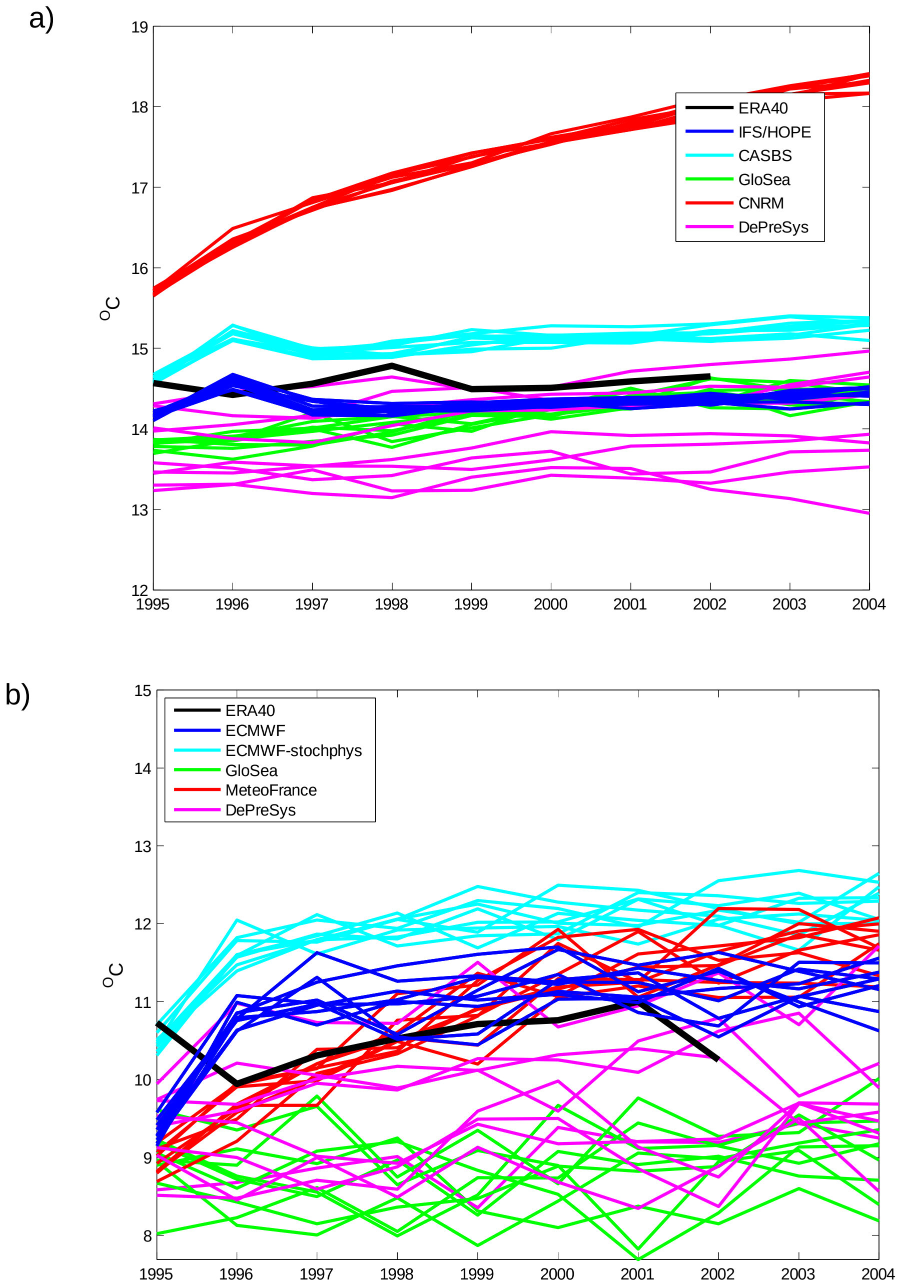


Figure 5: a) Global annual mean near-surface temperature of all available decadal simulations starting on 1st November 1994. b) The same as a), except for European annual mean values instead of global means. The individual models are IFS/HOPE control (dark blue), IFS/HOPE stochastic physics using CASBS1.2 (light blue), GloSea (green), CNRM (red), DePreSys_PPE (magenta). ERA-40 (black) covers the years up to 2001.

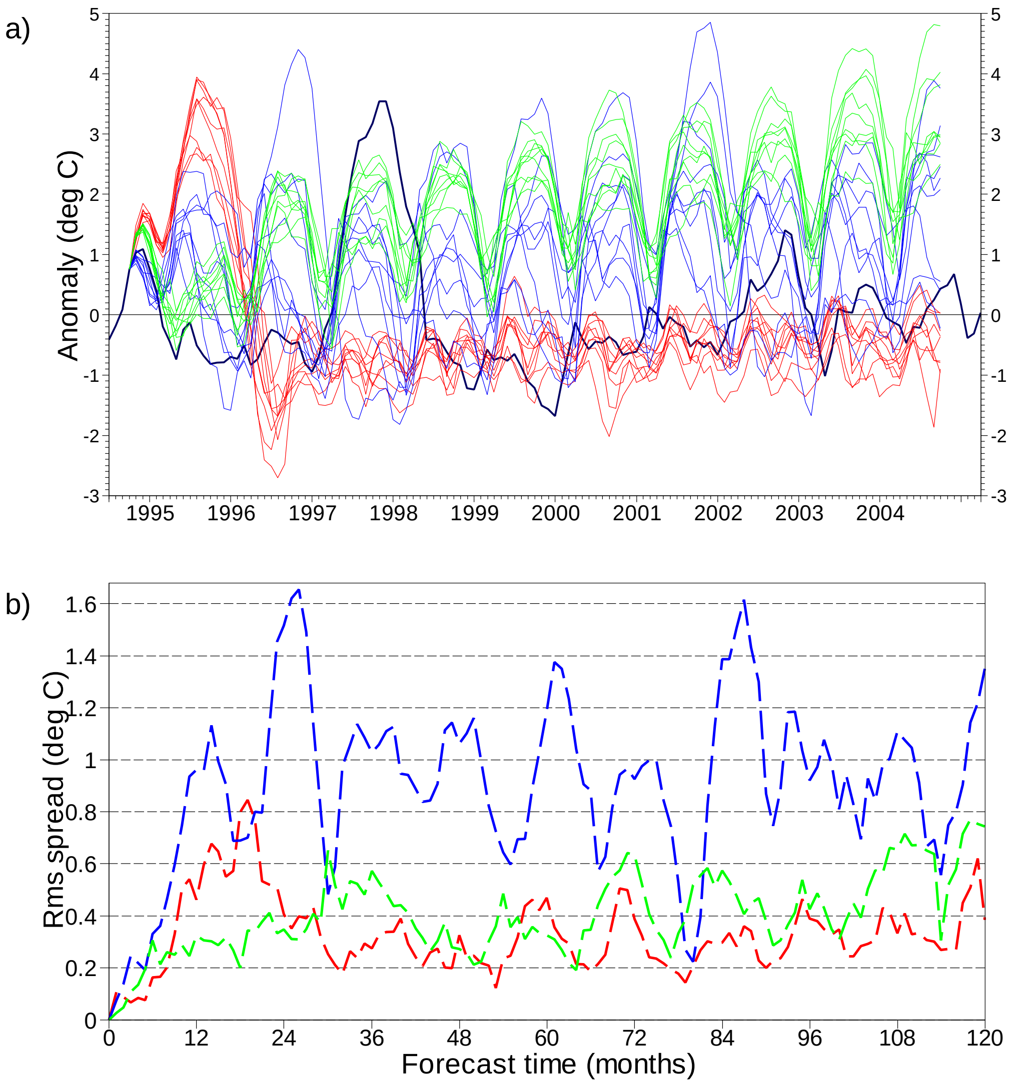


Figure 6: a) Nino3 SST anomalies of the decadal multi-model initial condition ensemble started in November 1994. The individual models are IFS/HOPE control (red), GloSea (blue) and CNRM (green). The black line shows the observed anomalies with respect to the observed climatology. The same observed climatology is used to compute the model anomalies. b) Nino3 SST spread of the individual single-model ensembles as measured by its standard deviation.

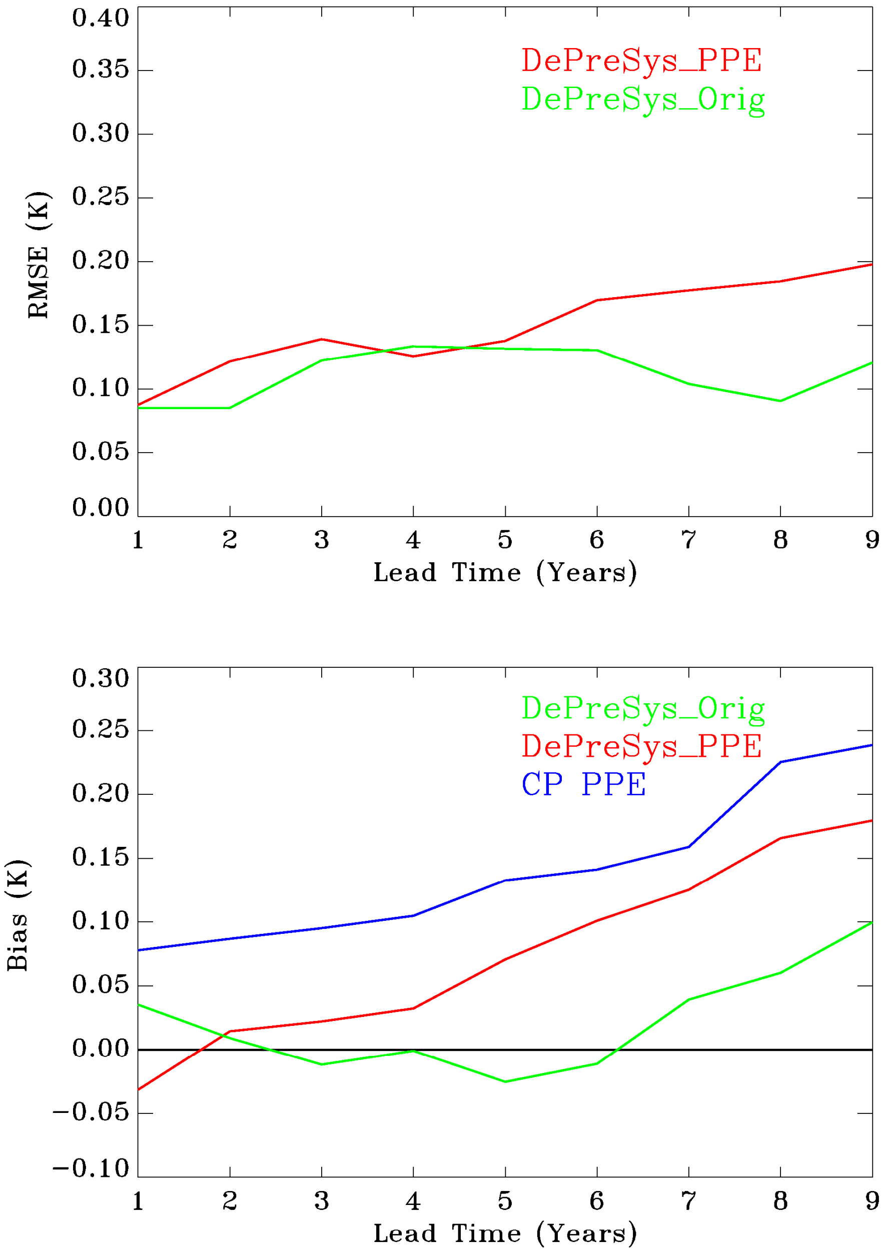


Figure 7: Decadal RMS error (upper panel) and average bias (lower panel) of the DePreSys_PPE and DePreSys_Orig ensemble means for annual global mean surface air temperature anomalies. The DePreSys_PPE system produced 9-member ensembles with 22 start dates, while DePreSys_Orig generated 8-member ensembles from 22 start dates. The lower panel also shows the average bias in the historical climate simulations of corresponding PPE members used in the centennial predictions of section 3 (blue curve).

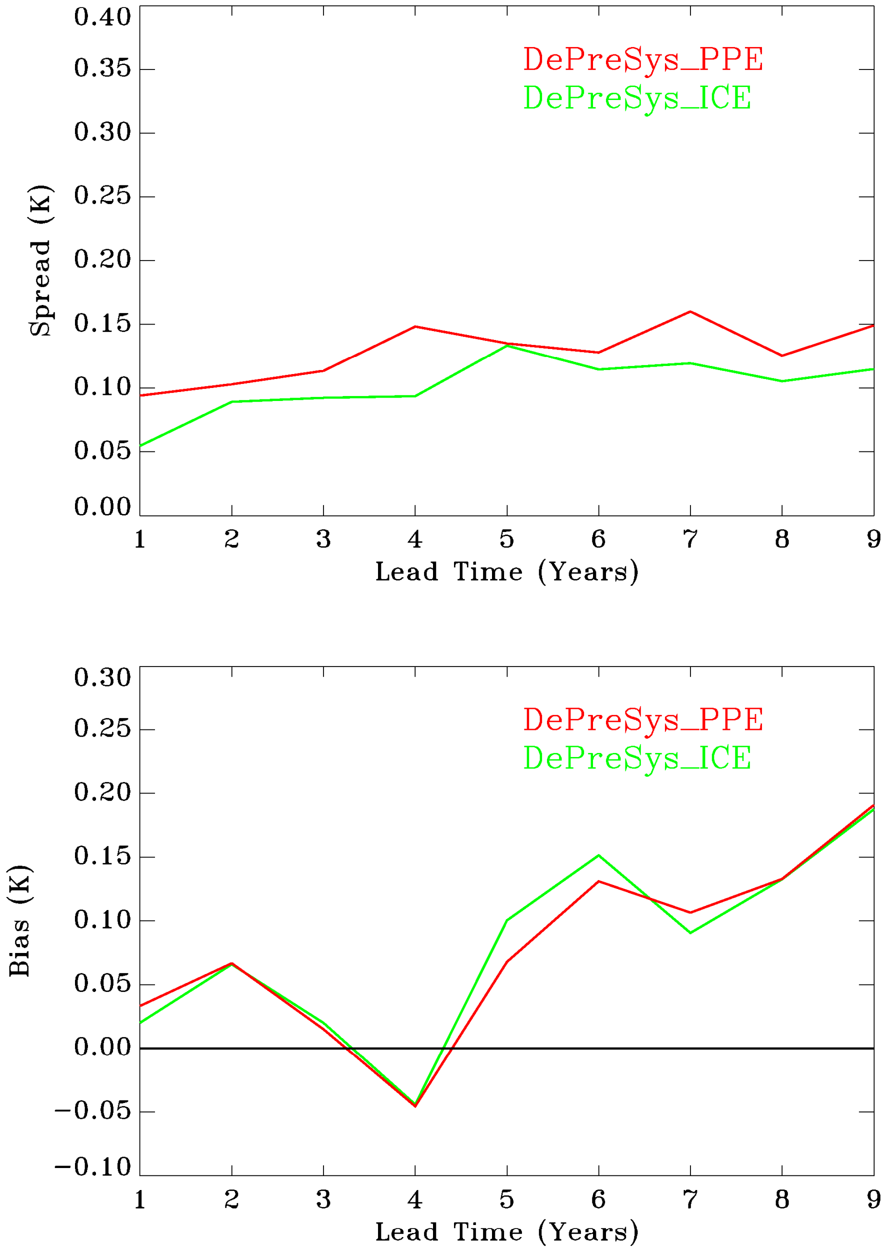


Figure 8: Average spread (upper panel) and bias (lower panel) for DePreSys PPE members and corresponding initial condition members. Four start dates (May 1991-92 and November 1993-94) have been used.

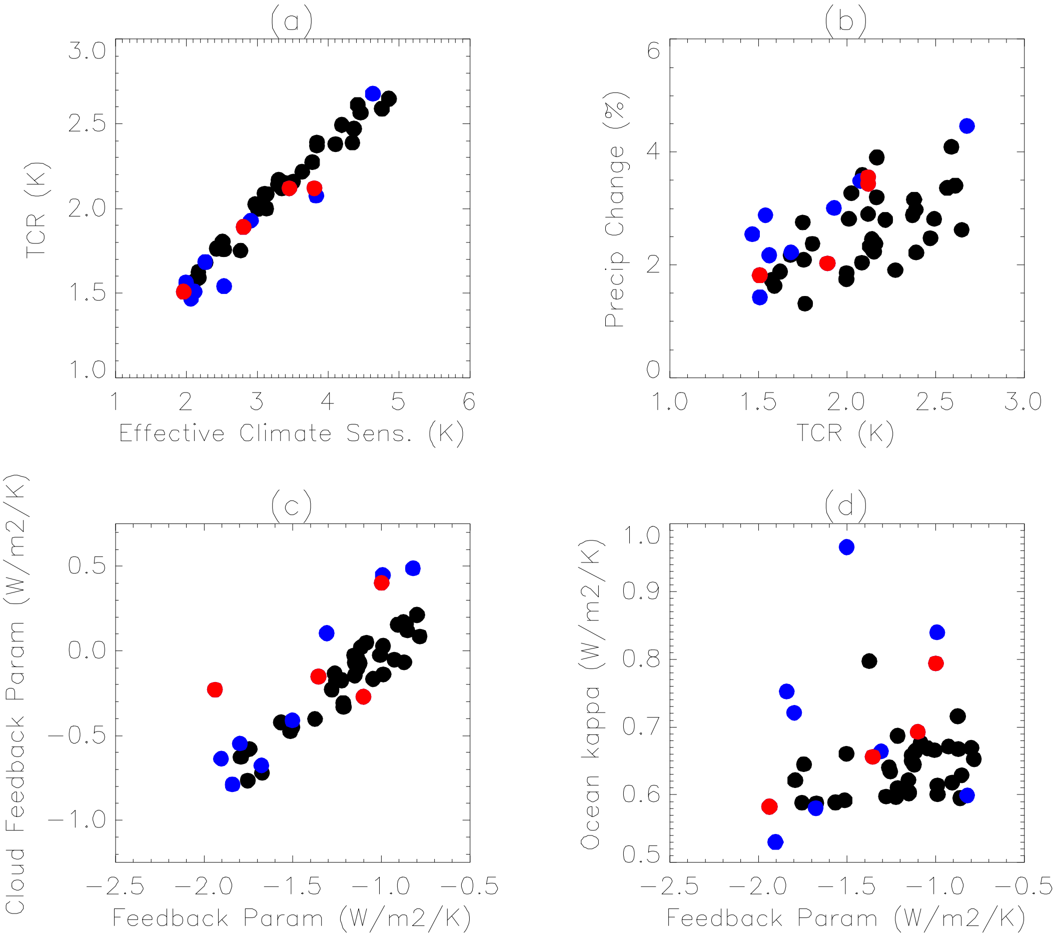


Figure 9: A comparison of global mean quantities computed from HadCM3 centennial perturbed physics experiments (black dots) and from the centennial multi-model ensemble (coloured dots with red indicating ENSEMBLES models and blue indicating other models from the IPCC AR4 archive). In each case, quantities are computed from the difference between the 20-year average at the time of CO_2 doubling in a 1%/year CO_2 scenario and the corresponding 80-year control experiment with fixed CO_2. (a) The effective climate sensitivity plotted against the Transient Climate Response (TCR – the global mean temperature change at 2xCO2). (b) TCR against percentage precipitation change. (c) The total climate feedback parameter against the cloud feedback parameter showing cloud feedbacks to be the main driver of uncertainty in temperature change. (d) The total feedback parameter against the effective ocean heat uptake efficiency.

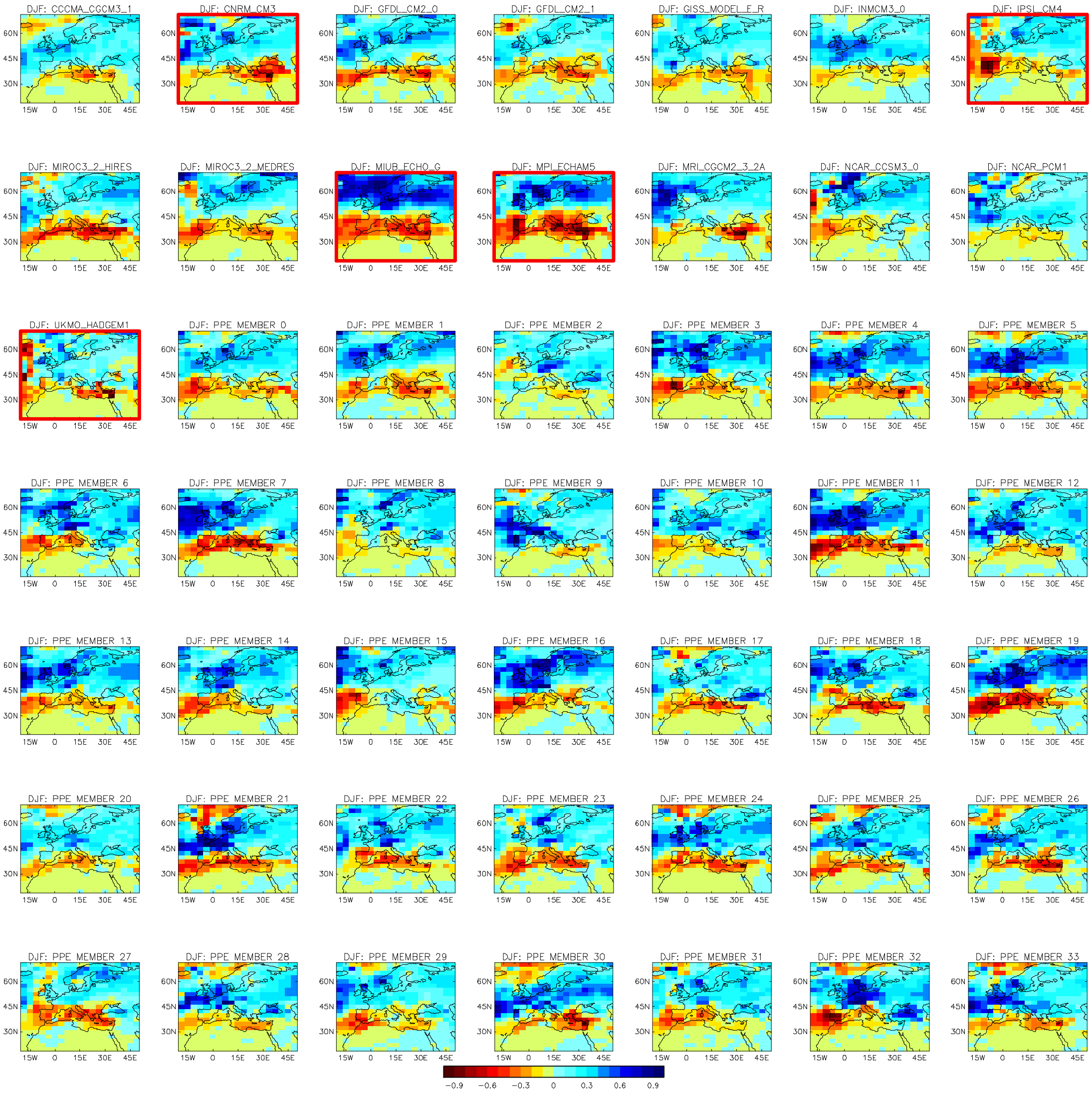


Figure 10: December-February precipitation change (mm/day) at 2xCO$_2$ from the 1%/year CO$_2$ increase scenario experiments. The first 15 panels show responses from the multi-model ensemble, with model names indicated and ENSEMBLES models highlighted with a red border. The remaining panels are from an ensemble of 34 perturbed physics versions of HadCM3. The consistency of the large-scale pattern of N. European wetting and S. European drying is noteworthy.