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 MM1.3

Specification of a "second generation" ensemble prediction system (Version 2).

Due date of deliverable: 31 August 2009
Actual submission date: 13 January 2010

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 MM1.3

J. Murphy, M. Collins, F. Doblas-Reyes., T. Palmer and A. Weisheimer

Specification of a "second generation" ensemble prediction system (Version 2).

1. Introduction

A report for ENSEMBLES Major Milestone 1.2 was produced in August 2006 (Weisheimer et al., 2006), summarising progress achieved in the development of an ensemble-based modelling system for the prediction of climate variability and change at seasonal to decadal (s2d) and multidecadal lead times, provided by the partners of ENSEMBLES RT1. This “first generation” system was based on the first stream of climate model s2d hindcasts and centennial climate change projections produced during the first two years of the project. It included an initial methodology for s2d prediction based on three different methods to account for uncertainties in the representation of physical processes in models. The system also included a set of longer term projections forming part of a systematic strategy for characterising uncertainties in longer term projections, based on relatively large ensembles carried out within a single model framework augmented by results from a multi-model ensemble of alternative models.

Section 2 recaps the three methods for characterising modelling uncertainties employed in the project, and sections 3 and 4 describe developments during the past three years leading to the second-generation system for s2d and multidecadal projections, including a selection of key results. These demonstrate the capabilities of the latest prediction system, which provides better projections on seasonal to annual time scales, a first demonstration of the benefits of sampling modelling uncertainties in decadal projections, and probabilistic climate change projections for Europe accounting more comprehensively than previously for uncertainties in current knowledge of key Earth system feedbacks and their uncertainties. Priorities for future research are discussed in section 5, recognising that the progress made during ENSEMBLES raises significant new challenges. These include the potential to combine more fully the different methods for sampling modelling uncertainties considered during the project, and to develop a single “seamless” prediction system covering seasonal, decadal and multidecadal time scales (e.g. Palmer et al., 2008; Hurrell et al., 2009).

2. Techniques for sampling modelling uncertainties

For a given scenario of future changes in natural and man-made external forcing agents (solar variability, volcanic eruptions, changes in emissions or concentrations of greenhouse gases, aerosols and their precursors), uncertainty in model projections arises from internal climate variability and uncertainty in prediction of the response to external forcing. Uncertainty due to internal variability can be reduced by starting projections from estimates of the current observed state of the climate system, and strategies for achieving this have been implemented in the s2d part of the new ensemble prediction system (section 3). However, modelling uncertainties also contribute to uncertainty in projections of internal variability, as well as to externally-forced natural variability and longer term climate change. Uncertainty in model formulation arises due to the inability of climate models to simulate every single aspect of the climate system with arbitrary detail. Climate models have limited spatial and temporal resolution, so that physical processes that are active at smaller scales (e.g. convection, orographic wave drag, cloud physics, mixing) must be parameterized using semi-empirical relationships.

In ENSEMBLES, three approaches to address model uncertainty in seasonal-to-decadal predictions have been explored:

1. The **multi-model method** empirically samples errors that occur due to structural inadequacy in individual climate models by using models with different formulations and parameterizations (Palmer et al. 2004; Weisheimer et al., 2009). 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.
2. Given that some of the most important model uncertainties are in the specification of the parameters that are used in the physical parameterizations (Murphy et al., 2004; Stainforth et al., 2005), the **perturbed-parameter approach** samples model uncertainty by creating ensembles of alternative variants of a single model in which multiple uncertain parameters are perturbed.
3. Due to the finite spatial resolution of climate models, the representation of processes on spatial scales smaller than the truncation scales, and their feedback onto larger scales, remains subject to considerable uncertainty. The impact of unresolved scales can be approximated by **stochastic physics** elements that either act as perturbations to the physical tendencies or via energy backscatter processes from the sub-grid scales to the resolved scales (Palmer, 2001; Palmer et al., 2009a).

In s2d prediction all three approaches were explored in version 1 of the ensemble prediction system, and this is again the case in the second version. Updates implemented for the second stream of hindcasts, which covered improved initialisation as well as changes to the methods for sampling modelling uncertainties, are briefly summarised in section 3. For centennial predictions, version 1 of the system consisted of an interim staging post towards a systematic methodology based on a set of perturbed parameter ensembles variants of a single climate model (HadCM3), augmented by results from a multi-model ensemble. Version2 (described in section 4) consists of the completed methodology, configured for probabilistic prediction through the use of a specialist statistical framework suitable to combine the available model projections with a set of observational constraints derived from key metrics of historical climate.

3. Seasonal-to-Decadal hindcasts

Major Milestone 1.2 (Weisheimer et al., 2006) summarised a first stream of coordinated multi-model s2d hindcasts, based on the IFS/HOPE (ECMWF), ARPEGE/OPA (Meteo-France), GloSea (UK Met Office), DePreSys (also UK Met Office) and ECHAM5/OM1 (IfM-GEOMAR Kiel) models, using nine ensemble members per model with perturbed initial conditions. Hindcasts were initialised each year from 1st May and November and run for seven and fourteen months respectively (see Doblas-Reyes et al., 2009). In addition, two experimental decadal hindcasts were performed using each model, started from 1st November 1965 and 1994. Parallel seasonal and annual hindcasts were also performed using an implementation of stochastic physics in one of the participating models (IFS/HOPE), using the Cellular Automaton Stochastic BackScatter parameterisation of Berner et al. (2008). This re-injects a fraction of the kinetic energy dissipated in the model back into the resolved flow (Shutts, 2005), on spatial scales determined by a simple cellular automaton. A further set of seasonal to annual hindcasts was performed using a nine-member perturbed parameter ensemble of variants of the DePreSys system, based on HadCM3. This system was also used to perform a larger set of decadal hindcasts, by extending the hindcasts from each of the 22 start dates to a decade ahead.

3.1 Stream 2 s2d hindcast experiments

Version 2 of the s2d ensemble prediction system is based on a second stream of hindcasts, consisting of a more comprehensive set of integrations covering the period 1960-2005:

- Seasonal: 7-months long, once a year starting in February, May, August and November.
- Annual: 14-months long, once a year starting in November.
- Decadal: Ten hindcasts, started from 1 November 1060, 1965, 1970,…, 2005.

The multi-model ensemble for the stream 2 experiments consisted of global coupled atmosphere-ocean climate models from the UK Met Office (UKMO), Météo France (MF), the European Centre for Medium-Range Weather Forecasts (ECMWF), the Leibniz Institute of Marine Sciences at Kiel University (IFM-GEOMAR) and the Euro-Mediterranean Centre for Climate Change (CMCC-INGV) in Bologna (Italy), see Weisheimer et al., 2009. Each of the models contributed to the seasonal experiment, and all except CMCC-INGV performed the annual hindcasts. Several groups used new or upgraded versions of their models, compared to those used in the stream 1 hindcasts. All models included the main radiative forcings. The atmosphere and ocean were initialized using realistic estimates of their observed states and each model was run from an ensemble of nine initial conditions (three for the decadal hindcasts). Table 3.1 summarizes the main model components and their initialization strategies.

The **stochastic physics approach** was also updated (Palmer et al., 2009a), and has been used in test mode for a subset of the seasonal hindcasts with ECMWF's coupled model. The approach is based on the idea of a stochastic representation of the equations of motion at the computational level and as such focuses on uncertainty related to unresolved processes. Conventional physical parametrization schemes describe the effects of subgrid-scale processes in models of weather and climate by deterministic bulk formulae which depend on the local resolved-scale variables. However, through the upscale cascade of energy, the neglected unresolved subgrid-scale variability can have an impact on the larger scales in the model and thus contribute to model errors on different spatial and temporal scales. Stochastic physical parametrization ensembles provide a methodology for representing model uncertainty due to variability of the unresolved scales. ECMWF has recently revised its stochastically perturbed parameterization tendency (SPPT) scheme and developed the stochastic backscatter scheme (SPBS). SPPT applies univariate Gaussian perturbations to the wind, temperature and humidity tendencies of physical processes in the form of multiplicative noise with a smoothly varying pattern in space and time. A two-scale version of the perturbations with a shorter characteristic spatio-temporal scale on the order of 6 hours and 500 km together with a longer characteristic spatio-temporal scale of 30 days and 2500 km has been used. The SPBS scheme is based on the idea of backscatter of kinetic energy from unresolved scales. It is formulated in terms of a spectral streamfunction forcing field estimated from the numerical, convective and orographically induced dissipation rate and uses vertical phase correlations. A preliminary set of seasonal hindcasts for the May and November start dates over the period 1991-2005 have been completed and results of which will be discussed in this report. ECMWF plans to extend these hindcasts to the full Stream 2 period and start dates as soon as the detailed settings of the scheme will have been finalised.

The **perturbed parameter approach** was based on the HadCM3 DePreSys system (as in the stream 1 hindcasts), with a number of corrections and refinements to the initialisation strategy. The nine model variants consisted of the version of HadCM3 with standard parameter settings, plus eight variants with multiple perturbations to 31 parameters controlling key surface and atmospheric processes, perturbed to sample a wide range of simulated outcomes for climate sensitivity and the amplitude of ENSO variability, while providing plausible simulations of present-day mean climate. These are nine of the 17 HadCM3 variants used to sample atmospheric process uncertainties in the centennial climate change projections of section 4.2 (see also Collins et al., 2009). This ensemble was used to produce a larger set of decadal hindcasts, initialsed every November from 1960-2005. Parallel “No_Assim” hindcasts employing identical specifications of external forcing (from major greenhouse gases, emissions of sulphate aerosol precursors, solar variability and major volcanic eruptions) were also produced, in which initial states were taken from randomly selected model states, rather than analyses of observations.

partner	**atmospheric model and resolution**	**ocean model and resolution**	**initialization**	
			atmosphere and land	ocean
ECMWF	IFS CY31R1; T159/L62	HOPE; 0.3º-1.4º/L29	ERA-40/oper. analysis, atmospheric singular vectors	wind stress perturbations to generate ensemble of ocean reanalyses; SST perturbations at initial time
UKMO	HadGEM2-A; N96/L38	HadGEM2-O; 0.33º-1º/L20	ERA-40/oper. analysis, anomaly assimilation for soil moisture	wind stress perturbations to generate ensemble of ocean reanalyses; SST perturbations at initial time
MF	ARPEGE4.6; T63	OPA8.2; 2˚/L31	ERA-40/oper. analysis	wind stress, SST and water flux perturbations to generate ensemble of ocean reanalyses
IFM-GEOMAR	ECHAM5; T63/L31	MPI-OM1; 1.5º/L40	initial condition permutations of three coupled climate simulations from 1950 to 2005 with SSTs restored to observations	
CMCC-INGV	ECHAM5; T63/L19	OPA8.2 ; 2˚/L31	AMIP-type simulations with forced SSTs	wind stress perturbations to generate ensemble of ocean reanalyses, SST perturbations at initial time

Table 3.1. *Overview of the forecast systems contributing to the second stream of s2d hindcasts.*

The **initialisation strategy** was achieved through a substantial effort on the ocean initialisation for s2d climate prediction, building on work done in the previous European projects DEMETER and ENACT. The work can be summarised as follows:

1. Improvement of existing assimilation systems, particularly those developed previously in ENACT, in order to facilitate production of multi-decadal re-analyses (see Deliverable D1.3). Several partners then pursued further developments of their data assimilation systems in order to

prepare the ENSEMBLES Stream 2 experiments. This involved, for example, better calibration of the systems, introduction of new datasets, better covariance models for the representation of the remote effects of available observations.
2. Improvement of the common EN3 database of observed temperature and salinity profiles: Recent data were included, historical records were updated with data from the recent World Ocean Database and the quality control was improved (Ingleby and Huddleston, 2007).
3. Definition of an approach to deal with uncertainties in estimates of the observed state of the ocean: The approach chosen by most of the groups was to produce sets of surface forcing perturbations and apply them to ocean models and assimilation systems in order to generate ensembles of ocean reanalyses and/or initial conditions. In particular, a set of perturbations for sea surface temperature (SST), wind stress and fresh water flux were produced and made available by ECMWF. Usage of these perturbations differed somewhat from one system to another, but they constitute a coordinated method of dealing with ocean uncertainty.
4. Final production of ensembles of multi-decadal ocean reanalyses, for both ocean state estimation and coupled seasonal to decadal hindcast initialisation.

A project report (Deliverable D2A.1, also appearing as Weisheimer et al, 2007) summarised the choices made. The ocean reanalyses provided the ocean initial conditions for the stream 2 multi-model ensemble hindcasts, and were based on improved assimilation schemes and observational datasets compared to the stream 1 hindcasts. Four (ECMWF, UKMO, CMCC-INGV and MF) of the five groups contributing to the multi-model ensemble sampled initial state uncertainties by perturbing the ocean analyses consistent with observational uncertainties in wind stresses and sea surface temperatures (SSTs). The fifth group (IFM-GEOMAR) generated initial conditions from three coupled simulations of 20th century climate change in which the model SSTs were restored to observations, obtaining initial conditions for the nine hindcast ensemble members by taking different initial condition permutations from these three runs. In the perturbed parameter hindcasts (DePreSys_PP), an anomaly initialisation approach was used. The ocean was initialised by running each model variant in assimilation mode from December 1958 to November 2005, relaxing the ocean towards off-line analyses of temperature and salinity anomalies previously created off-line, and added to model climatological values in an attempt to minimise climate drift during hindcasts. Members of the multi-model ensemble were started from analyses of absolute observed values, seeking to initialise hindcasts as realistically as possible at the cost of accepting larger drifts during hindcasts (although IFM-GEOMAR employed an anomaly initialisation approach for their decadal hindcasts). The strengths and weaknesses of these alternative approaches remains a question for future research. Most groups also used atmospheric data from ERA-40 and operational ECMWF analyses to initialise the atmosphere (see Table 3.1). Further details of the experimental set-up are provided in Deliverable D1.18.

3.2 Seasonal-annual hindcast results

Results from a large set of seasonal hindcasts show that significant progress has been made in reducing systematic model errors compared with previous generations of models. The forecast quality of the different approaches to model uncertainty on the seasonal-to-annual time-scale has been carefully analysed and assessed in terms of different aspects of deterministic and probabilistic forecast performance.

For tropical Pacific SSTs, the multi-model ensemble was shown to outperform any of the perturbed initial condition ensembles from the participating individual models, demonstrated by reduced RMS errors and enhanced ensemble dispersion at all lead-times (Weisheimer et al., 2009). A considerable reduction of systematic error compared with a previous multi-model ensemble from the DEMETER project was also found. Probabilistic forecast skill scores indicated that the new ENSEMBLES multi-model ensemble is, on average, more skilful than DEMETER in the 4-6 month forecast range. The degree of these improvements depends on the region, season and event of interest. The combination of ENSEMBLES and DEMETER into a grand multi-model ensemble did not improve the seasonal forecast skill further.

Preliminary simulations with the new stochastic physics scheme showed that this approach is capable of reducing systematic errors in the system and improving forecast scores over the control model version (Palmer et al., 2009a). The new scheme increases the ensemble spread substantially without adversely affecting the magnitude of forecast errors very much, thus successfully reducing the overconfidence in seasonal predictions. This is reflected in a better match between the ensemble spread and the root-mean-square (RMSE) error in the ensemble-mean, and also in improved reliability when hindcasts are expressed in probabilistic form.

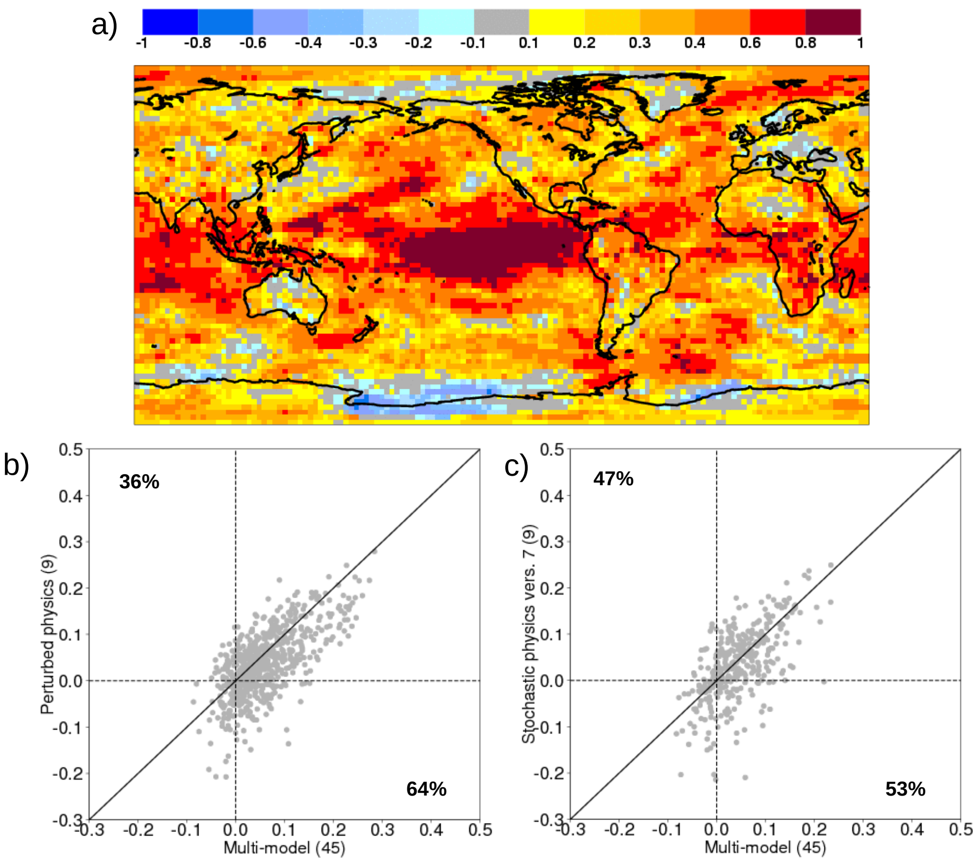

Figure 3.1. *a) Probabilistic forecast skill score (relative operating characteristics, ROC) of the multi-model ensemble for DJF warm temperature seasonal anomalies from hindcasts started on the 1st of November once a year over the period 1960-2005. b) Scatter plot of Brier Skill scores (BSS) for the perturbed parameter and multi-model forecasting systems for the standard land-regions for temperature and precipitation. The inset numbers indicate the percentage of wins for each sy stem. c) As in b), but for the stochastic physics hindcasts versus the multi-model.*

The multi-model ensemble has, in general, a high standard of forecast skill (Figure 3.1a). Encouragingly, the two new schemes to represent model uncertainty also provide forecasts competitive with the multi-model approach, for climate forecasts on seasonal time scales. The multi-model ensemble shows a higher anomaly correlation than any of its constituent individual models. When considering ensembles of equal size (nine members), the perturbed parameter ensemble gives similar skill to the multi-model ensemble, and in general the two methods give similar estimates of the spread of possible outcomes. The skill of the multi-model ensemble improves somewhat when all 45 of its members are pooled to provide a better (combined) sample of initial state and model uncertainties.

The relative performance for the three methodologies for forecasting temperature and precipitation over a set of standard land regions around the globe up to 7 months ahead is summarised in Figures 3.1b and 3.1c. This is based on a probabilistic measure of skill which credits both the ability to discriminate between different events, and to forecast probabilities which are reliable, in the sense that an event predicted x% of the time should occur in practice x% of the time. Here, the multi-model ensemble gives slightly better results than the perturbed parameter ensemble on average, although the relative performance varies according to the region, variable and lead time considered. Figure 3.1c shows that the new stochastic physics results provide a level of skill comparable to that of the multi-model ensemble on average, based on a similar comparison.

Overall, we find that the multi-model ensemble gives seasonal predictions competitive with (and in some respects better than) those obtained in previous projects, while the stochastic and perturbed parameter techniques provide promising indications that a similar level of performance can potentially be achieved through the application of systematic techniques for the sampling of uncertainties in a single-model system. The complementary benefits of the different approaches provide future potential to address model uncertainty more comprehensively in climate predictions across seasonal to decadal and longer time scales.

A preliminary study (ENSEMBLES deliverable D1.17) was performed to assess the potential achieve improved skill by combining results from the different approaches. Table 3.2 shows some of the results, for seasonal hindcasts of upper and lower tercile near-surface temperature and precipitation events over a set of 21 standard land regions (Giorgi and Francisco, 2000). The skill of forecast probabilities is measured using a version of the Brier Skill Score adjusted to remove approximately the effects of differing ensemble sizes. For temperature, a single system produces the best results in 75% of cases (consisting of 32%, 22% and 21% of cases for the multi-model, stochastic physics and perturbed parameter ensembles respectively), whereas for precipitation a single system gives the best scores in 73% of cases (noting that the stochastic physics and perturbed parameter ensembles both contribute more “wins” than the multi-model ensemble in the case of precipitation). The preponderance of wins for an individual system indicates that the benefits of combining information are modest in this example, although we note that no attempt was made to weight the contribution from different types of ensemble in this early study. Nevertheless, we do find 25% of cases for temperature (23% for precipitation) for which hindcasts achieved by pooling results from different ensembles do provide the best results, illustrating the potential noted above for achieving improved skill by combining the different methods to take advantage of their complementary nature. For precipitation, the option of combining the stochastic physics and perturbed parameter results is more often effective at improving skill than that of combining all three methods, however for temperature hindcasts both types of combination are effective in more than 10% of cases. We note, however, the statistical robustness of the results remains to be assessed.

More generally, further research is needed to investigate more fully the extent to which the different forecast systems assessed in ENSEMBLES can be combined to improve forecast skill, and associated estimates of error and reliability, recognising that the scope for generating improved information will depend on the variable, location and lead time of interest, and which metrics of a)

	temperature									
	JJA			DJF						
	cold	warm		cold	warm			best		
Australia								MM	26	32%
Amazon Basin								PP	17	21%
Southern South America								SP	18	22%
Central America								PP+SP	9	11%
Western North America								MM+PP+SP	12	15%
Central North America									82	
Eastern North America										
Alaska										
Greenland										
Mediterranean										
Northern Europe										
Western Africa										
Eastern Africa										
Southern Africa										
Sahel										
South East Asia										
East Asia										
South Asia										
Central Asia										
Tibet										
North Asia										

b)

	precipitation									
	JJA			DJF						
	dry	wet		dry	wet			best		
Australia								MM	12	14%
Amazon Basin								PP	22	27%
Southern South America								SP	30	36%
Central America								PP+SP	14	17%
Western North America								MM+PP+SP	5	6%
Central North America									83	
Eastern North America										
Alaska										
Greenland										
Mediterranean										
Northern Europe										
Western Africa										
Eastern Africa										
Southern Africa										
Sahel										
South East Asia										
East Asia										
South Asia										
Central Asia										
Tibet										
North Asia										

Table 3.2. *Comparison of the BSS for a) temperature and b) precipitation events between the multi-model (MM), the perturbed physics (PP), the stochastic physics (SP), the combination of PP+SP and the combination of MM+PP+SP ensembles, for seasonal hindcasts of December-February and June-August, 2-4 months ahead. For each of the 21 standard land regions and events (anomalies in the lower or upper tercile), the system with the largest score is displayed, see colour code at the right of the table. The percentages indicate the number of cases for which each system (or combination of systems) has the largest score.*

forecast performance are being considered. More sophisticated postprocessing techniques will be required to assess this issue comprehensively. For example, in D1.17 we also assessed the ensemble skill and dispersion characteristics of seasonal hindcasts of sea surface temperature hindcasts over the tropical Pacific, using root-mean-square and standard deviation metrics which credit the improved sampling achieved by increasing the number of available ensemble members (in contrast to the metric of Table 3.2). In this case, we found that it is difficult to improve significantly on the performance of the full 45 member multi-model ensemble hindcast by including information from the nine member perturbed parameter and stochastic physics results. This is likely to be partially a simple consequence of the relative sizes of the three types of ensemble, and partially due to intrinsic variations in their skill and spread properties independent of ensemble size. In principle a weighting scheme could be used to account for these issues and optimise the performance of a combined system, however this has yet to be attempted.

In ENSEMBLES for the first time, coordinated multi-model hindcasts on the annual time range have been performed. Four of the participating modeling groups (Table 1) extended the length of the stream 2 hindcasts for the November start dates beyond seasonal time scales up to 14 months. An approximately linear growth of tropical Pacific SST RMSE and ensemble spread for the multi-model was found (Weisheimer et al., 2009). The SST anomaly correlation reached a level of approximately 0.5 around forecast month 9 and stayed relatively constant thereafter.

3.3 Decadal hindcast results

The basis for investigating decadal prediction rests on evidence from observed low frequency climate variations around the world, results from idealised modelling studies and evidence that forced climate change can also provide skill (see Meehl et al. (2009) for a recent review). The stream 2 decadal hindcasts provided a first opportunity to assess the benefits of combining projections from different models in a coordinated experiment, following initial studies carried out with individual climate models (Smith et al., 2007; Keenlyside et al., 2008; Pohlmann et al., 2009).

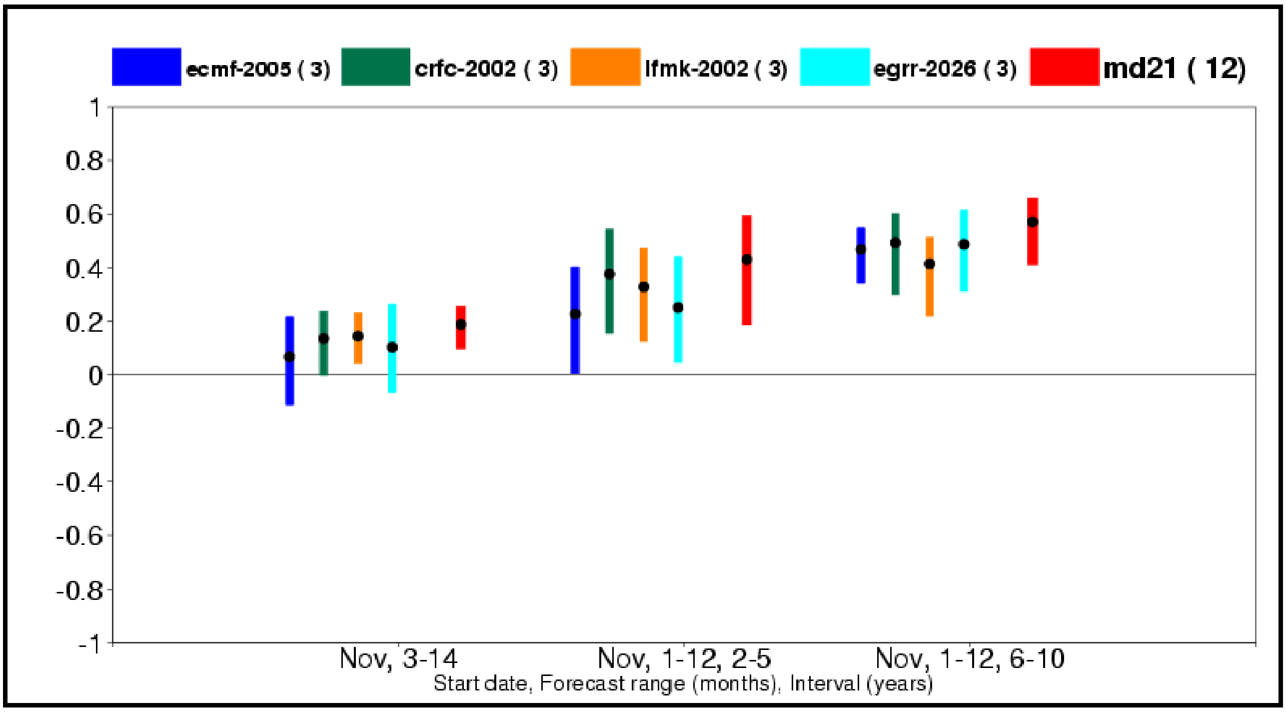

Figure 3.2. *Anomaly correlation coefficient for near-surface temperature over the NH extratropics from the ENSEMBLES decadal hindcasts over the period 1960-1995. The three groups of bars stand for lead times of 3-14 months, i.e. the first forecast year (left), for lead times 2-5 years (middle) and lead times of 6-10 years (right). The vertical bars indicate the uncertainty range based on a bootstrap resampling. Colour code: blue - ECMWF, green - CERFACS, orange - IfM Kiel, cyan - Met Office HadGEM2 and red - multi-model ensemble*

The inevitable existence of simulation biases in the models used for the decadal hindcasts necessitates (as in seasonal prediction) the use of strategies to account for these systematic errors when comparing forecasts against observations. This was achieved by expressing each hindcast as anomalies relative to either a long term model climatology (if available), or to the average of other hindcasts. Some groups initialised their hindcasts using observed anomalies added to a model climatology in order to reduce model drift, whereas others initialised using full observed fields in order to provide starting conditions as close as possible to the real climate system (see section 3.1).

Figure 3.2 shows that each of the models contributing to the multi-model ensemble achieves modest skill in projections of surface temperature anomalies averaged over the northern hemisphere extratropics. The skill increases for longer lead times, being larger for 6-10 years ahead than for 3-14 months or 2-5 years ahead. This is because the forced climate change signal, the sign of which is highly predictable, is greater at longer lead times. Encouragingly, the multi-model ensemble mean, which consists of the average of 12 individual projections, gives somewhat higher scores than any of the individual models, whose projections are derived from three members with perturbed initial conditions.

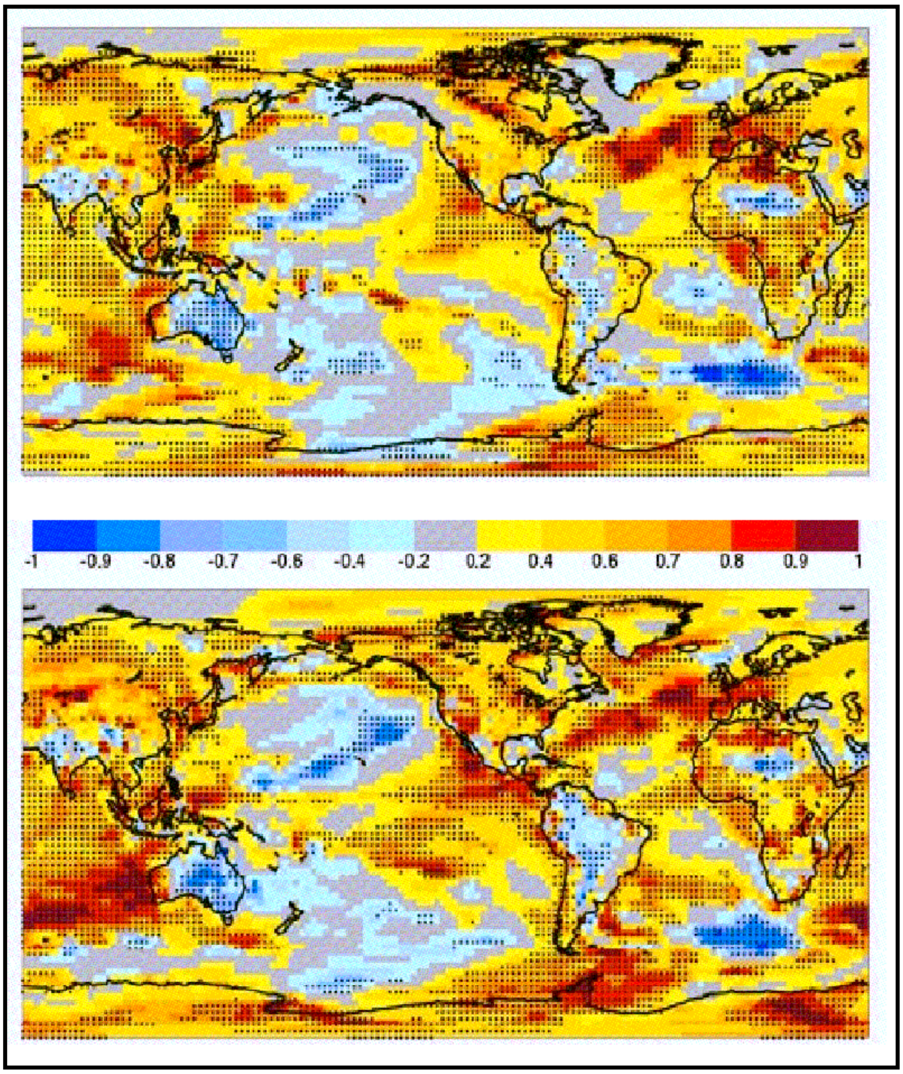

Figure 3.3. *Correlations between hindcast anomalies of surface air temperature for 2-5 years ahead, against ERA40/ERAInterim reanalyses of observations. Top panel shows skill for the multi-model ensemble-mean; bottom panel shows skill for the ensemble mean of the DePreSys_PP perturbed parameter hindcasts. Scores were computed over hindccasts initialised once every five years from 1960-1995.*

Figure 3.3 shows maps of the correlation against observations for ensemble-mean hindcasts from the multi-model and perturbed parameter ensembles, for hindcasts of near-surface temperature 2-5 years ahead. The two systems show statistically significant skill over large regions, especially over the tropics and the North Atlantic, but also over large parts of the continents and even Europe. The spatial variations in skill are similar between the systems, with the largest differences appearing over the tropics.

The perturbed parameter hindcasts also show improved skill when results from the individual model variants are averaged to form an ensemble mean (in this case there is a single hindcast from each variant, so the ensemble mean is made from nine members). Figure 3.4 shows an example, plotting a time series of global pattern correlations for nine year average hindcasts of surface temperature throughout the stream 2 period. While individual ensemble members sometimes give better results than the ensemble-mean (not shown), the average skill of individual members is consistently smaller (compare dashed and solid red curves). The results also show that the skill increases for more recent hindcasts. In order to diagnose sources of skill, the blue curve of Figure 3.4 shows ensemble mean results from the parallel ensemble of "No_Assim" hindcasts containing the same external forcing from greenhouse gases, sulphate aerosols, volcanoes and solar variations, but initialised from randomly-selected model states rather than analyses of observations. The No_Assim results replicate the general trend in skill found in the initialised hindcasts from 1980, showing that the trend arises mainly from the strengthening influence of external forcing, particularly that due to man-made greenhouse gases. However, the average correlation skill is slightly smaller in the No-Assim hindcasts (0.25 cf 0.30), indicating that initialisation provides a modest increase in average skill.

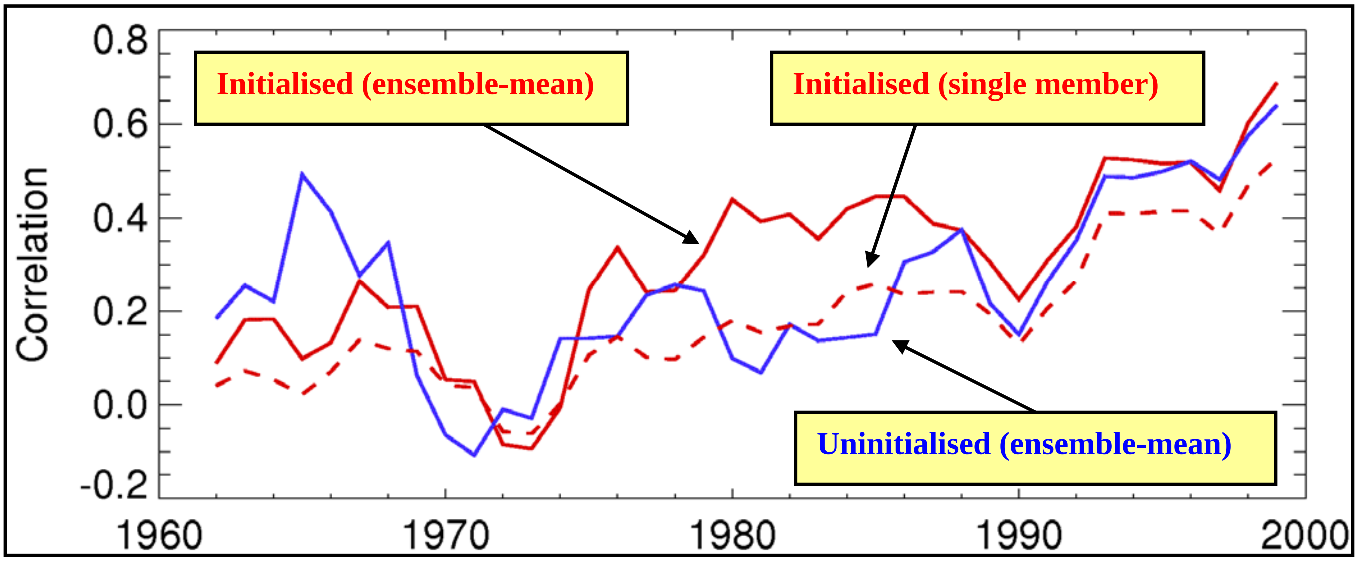

Figure 3.4. *Time series of correlation between hindcast and observed global patterns of near-surface temperature anomalies, for hindcasts of nine-year means during the stream 2 period. Red dashed curve shows average scores for individual variants of HadCM3 included in the DePreSys perturbed parameter ensemble, and solid red line shows scores for the ensemble mean of the nine constituent variants. Blue curve shows scores for the ensemble mean of a corresponding "No_Assim" perturbed parameter ensemble in which hindcasts are driven by the same time-dependent specification of external radiative forcing anomalies, but lacking the initialisation from analyses of atmosphere and ocean observations used in the DePreSys hindcasts. Time-dependent forcing anomalies arise from man-made greenhouse gases and sulphate aerosols from the SRES A1B scenario, and projected natural forcing from volcanoes and solar variations, assuming no prior knowledge of volcanic eruptions after the initialisation date.*

4. Centennial predictions

An overarching aim if the ENSEMBLES project has been to produce projections of climate in which sources of uncertainty due to internal variability and model formulation are expressed in terms of probability distribution functions (pdfs) representing the spread of possible outcomes in response to specific future emissions scenarios, consistent with current understanding of known climate feedback processes.

The principal tool chosen to quantify both internal variability and modelling uncertainties in our long term climate projections was the “perturbed parameter” approach, whereby uncertainties in global climate model parameters which determine the magnitude of climate feedbacks associated with physical, chemical and biological processes are systematically explored. The methodology was based on a coordinated set of perturbed parameter experiments using the HadCM3 climate model, and run at the Met Office Hadley Centre (MOHC). Further studies of the properties of perturbed parameter ensembles were carried out at the University of Oxford and the Freie Universität Berlin, and results from these were used to assess the basis for the approach, and to inform aspects of experimental design. A brief survey of results from some of these ensembles is given in section 4.1, while the method used to convert our perturbed parameter ensemble results into probabilistic climate change projections is described in section 4.2. In these projections, uncertainties due to internal variability were included via the use of long model simulations spun up from initial conditions statistically independent of recent observed conditions (as is currently typical in long term climate projections). A topic for future work is to assess the prospects for constraining some aspects of internal variability in projections beyond a decade ahead, by initialising the model with observations as in section 3.

4.1 Sampling uncertainties in future climate change using perturbed parameter ensembles

Early studies had shown that many surface and atmospheric model parameters were capable of influencing projections of global climate change in responses to changes in greenhouse gas forcing (Murphy et al., 2004; Stainforth et al., 2005). It was therefore necessary to sample possible outcomes using relatively large ensembles of projections exploring the effects of multiple combinations of plausible parameter values. The need for such an approach is illustrated by Figure 4.1, which shows projections of climate sensitivity (the global mean equilibrium surface warming in response to a doubling of CO_2) obtained using several thousand perturbed variants of HadSM3, a configuration of HadCM3 in which the atmosphere is coupled to a simple mixed layer (slab) ocean. This result, obtained from HadSM3 simulations run on personal computers by Oxford University under the climate*prediction*.net initiative, shows the relative likelihood of the present-day climate (based on a simple goodness-of-fit to observed surface pressure, temperature and rainfall) plotted against climate sensitivity, where points are coloured by the values assigned to one of the parameters of the model. This “entrainment coefficient” parameter controls the mixing between ascending plumes and the surrounding environment in the HadCM3 parameterisation of convection, and was found to have the largest impact on sensitivity. The highest sensitivities, in excess of 10^{o}C, are associated with low values of this parameter (red symbols) which consistently give unrealistic control climates. However, increasing the entrainment coefficient from the standard values (green symbols) to a high value (blue symbols) displaces the distribution of sensitivities in models with more plausible control climates towards higher values. This illustrates the importance of non-linear interactions between parameters in perturbed parameter ensembles, since if the entrainment coefficient is increased on its own sensitivity is reduced (Murphy et al, 2004).

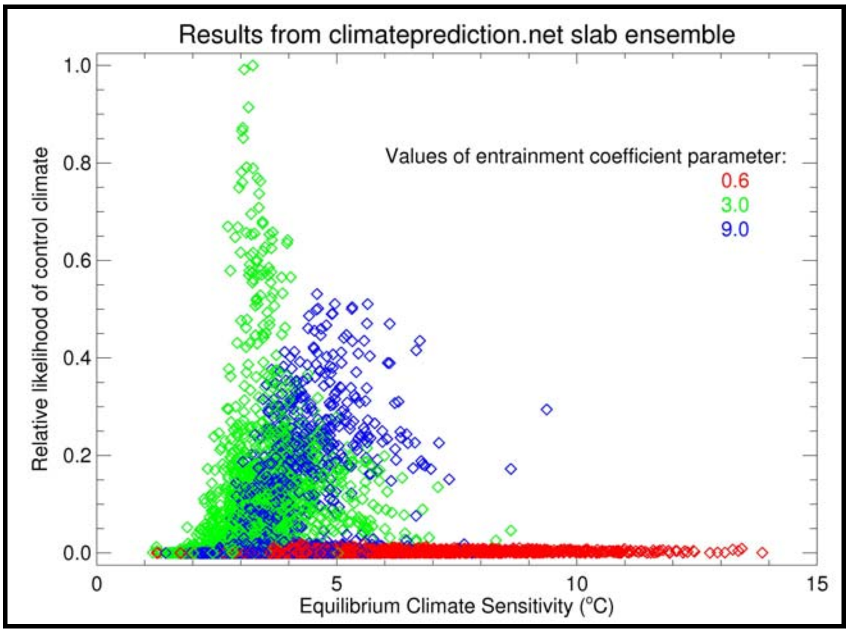

Figure 4.1. *Climate sensitivity versus an estimate of the relative likelihood of different model variants, selected from a multi-thousand member perturbed parameter ensemble of the atmosphere-mixed layer ("slab") ocean configuration of the HadCM3 climate model. Colours denote different values for the convective entrainment parameter in the different model variants.*

For the probabilistic prediction system, an ensemble of 280 variants of HadSM3 was therefore run, in order to sample the effects of process interactions such as those of Figure 4.1. This ensemble was run on the Met Office supercomputer, in order to allow the archival of a detailed set of regional climate diagnostics to provide a basis for the probabilistic projections described in section 4.2. Most of these were available as part of version 1 of the ENSEMBLES prediction system, with further simulations to improve sampling of multiple parameter combinations run since. A further suite of perturbed parameter experiments were necessary, however, in order to provide projections of transient climate change in combination with the simulations of equilibrium changes described above. Figure 4.2 illustrates the transient projections, performed at MOHC using the configuration of HadCM3 including a full dynamical ocean component (climate*prediction*.net also carried out ensembles of this type). In this set of experiments, parameters in one of the model components (atmosphere, ocean, sulphur-cycle and terrestrial carbon cycle) were varied while parameters in the other components were held fixed. The ensemble sampling the effects of surface and atmosphere parameters was carried out as part of version 1 of the ensemble prediction system, subsequently augmented by further ensembles varying parameters in the ocean, sulphur cycle and terrestrial carbon cycle modules of HadCM3 (more details in Collins et al., 2009). For global mean surface temperature, uncertainties in atmosphere parameters have a relatively important impact on the ensemble spread via differences in the surface and atmospheric physical feedbacks (black lines in Fig. 4.2). Feedbacks associated with clouds are the dominant source of uncertainty but other feedback processes also contribute to the spread of projected outcomes (Webb et al., 2006; Yokohata et al., 2009; Collins et al. 2009). Uncertainties in parameters in the terrestrial carbon cycle of the model have a similarly important effect on the global mean (red lines in Fig. 4.2), via differences in feedbacks associated with the changing balance of carbon sinks and sources (Booth et al., 2009). Ocean-component and sulphur-cycle component parameter uncertainties have a smaller effect on the global mean temperature response (Collins et al. 2007; Brierley et al. 2008; Brierley et al. 2009).

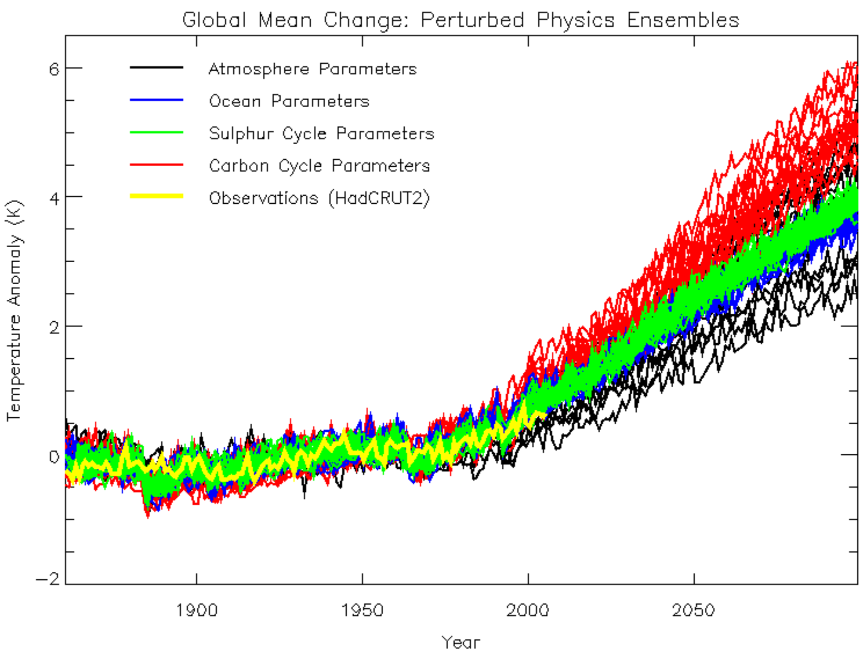

Figure 4.2. *Global-mean temperature anomalies from 1860-2100 in perturbed parameter HadCM3 experiments forced by historically observed changes in anthropogenic and natural forcing agents and future greenhouse gas and sulphate aerosol emissions under the SRES A1B scenario, compared with observations to 2000. The different colours indicate ensembles with perturbations to parameters in different model components (as indicated in the legend) while keeping parameters in the other components fixed.*

While the system for probabilistic projections uses perturbed parameter ensembles based on HadCM3, the effects of some of the parameters were compared with corresponding perturbations applied in another model (the Freie Universität Berlin EGMAM model). Figure 4.3 shows a comparison of the effects of perturbing three key cloud and convection parameters in the two models. The magnitudes of the changes are different due to different experimental setups; The results show equilibrium responses to doubled CO_2 in simulations employing the slab ocean configuration of HadCM3, whereas the EGMAM simulations show smaller changes, because a full dynamical ocean component was used to simulate the early stages of the transient response to an instantaneous change in CO_2. Nevertheless, the results show that the global mean temperature response of the two sets of experiments is highly correlated (see Niehörster (2009) for further analysis of the contributing feedback processes). This comparison gives further evidence that perturbing the parameters within a climate model is a useful way of systematically exploring uncertainty and thus provides a suitable technique for producing probabilistic projections of climate change.

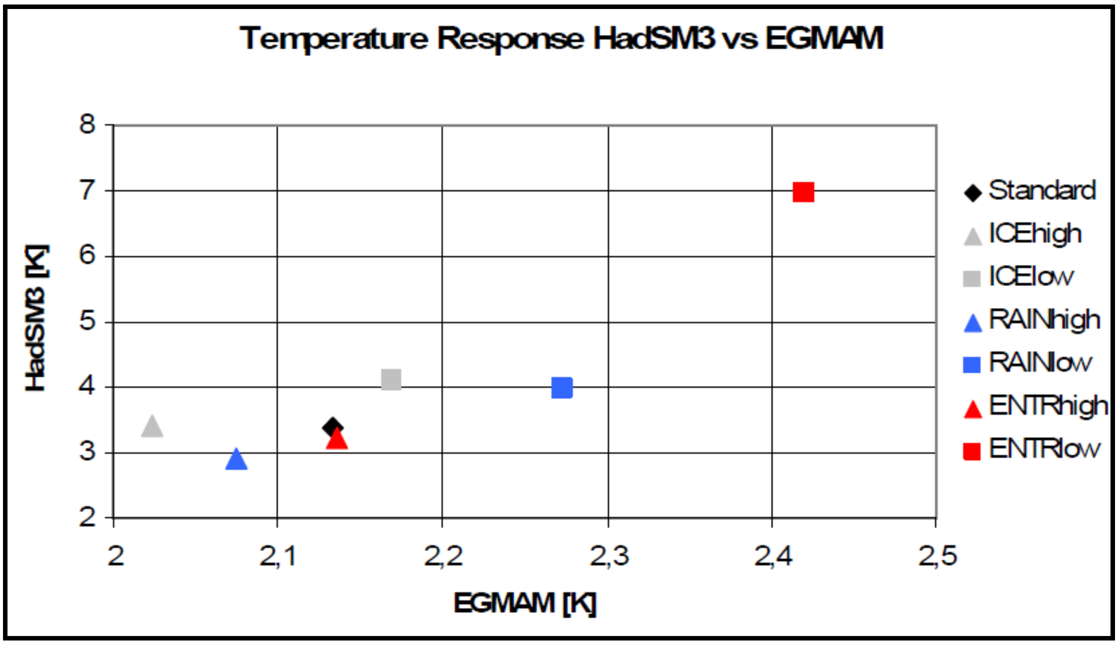

Figure 4.3. *A comparison of perturbed parameter experiments performed with different ENSEMBLES models. The figure shows global mean temperature responses of perturbed variants of HadSM3 and EGMAM models for low (squares) and high (triangles) values for the convective entrainment rate (red), the rainout efficiency of cloud droplets (blue) and the cloud ice fall speed (grey) different parameters. The standard model variants are shown in black. Note the differences in the scale of the axes which is due to the different experimental setup for each ensemble (see text).*

4.2 Construction of probabilistic climate change projections

Given the available set of perturbed parameter ensembles, there are further steps required to produce probabilistic projections expressed in terms of pdfs. An overview of the approach is given by Murphy et al. (2007), the detailed implementation of which follows that used in probabilistic projections recently issued for the UK (Murphy et al., 2009), except that the RT1 projections (Harris et al., 2009) do not include downscaling beyond the scale of global climate model grid boxes. In ENSEMBLES, finer scale projections are provided by RT2B, using an ensemble of regional climate model simulations designed in RT3. The steps involved in producing the RT1 probabilistic projections were:

1. Production of an ensemble of 280 equilibrium 1xCO_2 and 2xCO_2 simulations using HadSM3, as mentioned above. Each pair of simulations was carried out using a variant of the model distinguished by different perturbations to a set of 31 parameters controlling surface, atmospheric and sea-ice processes. Some variants sampled perturbations to a single parameter (Murphy et al. 2004), whereas many sampled multiple perturbations (Webb et al., 2006, Collins et al. 2009).
2. Construction of a statistical emulator allowing the historical climate and the response to doubled CO_2 of HadSM3 to be rapidly estimated for any combination of input parameter values. This allows large numbers of estimated results to be produced, making it possible to sample the entire parameter space of the model defined from expert-specified prior distributions.
3. Production of smaller ensembles using the atmosphere component of HadCM3 coupled to a full dynamical ocean component. These ensembles (described above and shown in Figure 4.2) simulate transient climate change in response to historical and future changes in forcing, and allow us to sample uncertainties in additional earth system processes .

4. Implementation of a time-scaling approach which maps equilibrium changes in climate variables at a regional level to transient climate changes under specified emissions scenarios. This allows transient changes to be estimated for any point in the model parameter space, providing a basis for the generation of probabilistic estimates of regional, time-dependent climate change. The time-scaling combines a method for inferring transient patterns of change from equilibrium patterns (calibrated using equilibrium and transient model simulations with corresponding parameter settings) with projections of global mean temperature obtained from a simple climate model. The simple model uses input parameters fitted to the HadCM3 ensemble output in order to sample uncertainties due to the main global-scale feedbacks accounted for in Figure 4.2. The method is based on Harris et al. (2006), with updates summarised by Murphy et al. (2009).
5. Estimated climate changes are converted into probabilistic projections using a general Bayesian statistical framework designed to support inference of real world information from complex but imperfect models (Goldstein and Rougier, 2004; Rougier, 2007). Probabilities are obtained by integrating changes sampled at different points in the model parameter space, accounting for relationships between different variables and weighting each point in parameter space (i.e. each possible variant of HadCM3) according to the likelihood of each variant. Likelihood estimates are based on the ability to reproduce observed spatial patterns of seasonal-mean climate for sea surface temperature, land surface air temperature, precipitation, pressure at mean sea level, shortwave and longwave radiation at the top of the atmosphere, shortwave and longwave cloud radiative forcing, total cloud amount, surface fluxes of sensible and latent heat, and latitude-height distributions of zonally averaged atmospheric relative humidity. The ability of the scaled transient projections to reproduce historical trends in large-scale temperature variables also contributes to the likelihood weights. The likelihood is calculated in a reduced dimension state space and takes into account covariances between different variables.
6. The calculation includes an estimate of the additional effects of structural modelling uncertainties, necessary because some simulation errors in HadCM3 (as in any model) arise from basic choices made when building the model in the first place (see also section 2), and cannot be resolved by varying model parameters. The effects of structural errors in surface and atmospheric processes are estimated by using the emulator to find points in the parameter space of HadCM3 which best represent the behaviour of alternative coupled atmosphere-mixed layer ocean models in the CMIP3 archive used by the IPCC AR4 (Meehl et al., 2007). Due to structural differences between HadCM3 and the other models, this calculation fails to replicate perfectly the projections of the latter (beyond the level of mismatch consistent with internal variability), and the results are used to inflate the variance and adjust the mean of the future pdfs. This calculation assume that structural differences between models are reasonable proxy estimates for the effects of structural errors in HadCM3 relative to the real world, and cannot account for the effects ob biases common to all models. The methodology does not support treatment of structural errors in ocean transport and carbon cycle processes to the same degree. However, a simpler *ad hoc* allowance is made for these, by including results from multimodel ensembles (Friedlingstein et al., 2006; Meehl et al., 2007) along side those of Figure 4.2 when sampling possible settings for global mean feedback values in step (4) above.

There are various ways of presenting the spatio-temporal information contained within the pdfs. Figure 4.4 shows European maps of the 20-year 10^{th}, 50^{th} and 90^{th} percentiles of surface air temperature and precipitation changes under the SRES A1B scenario at the end of the 21^{st} century expressed as anomalies with respect to a 1961-90 baseline. Median temperature changes vary substantially with location, and are largest in the Mediterranean region in summer and in north east

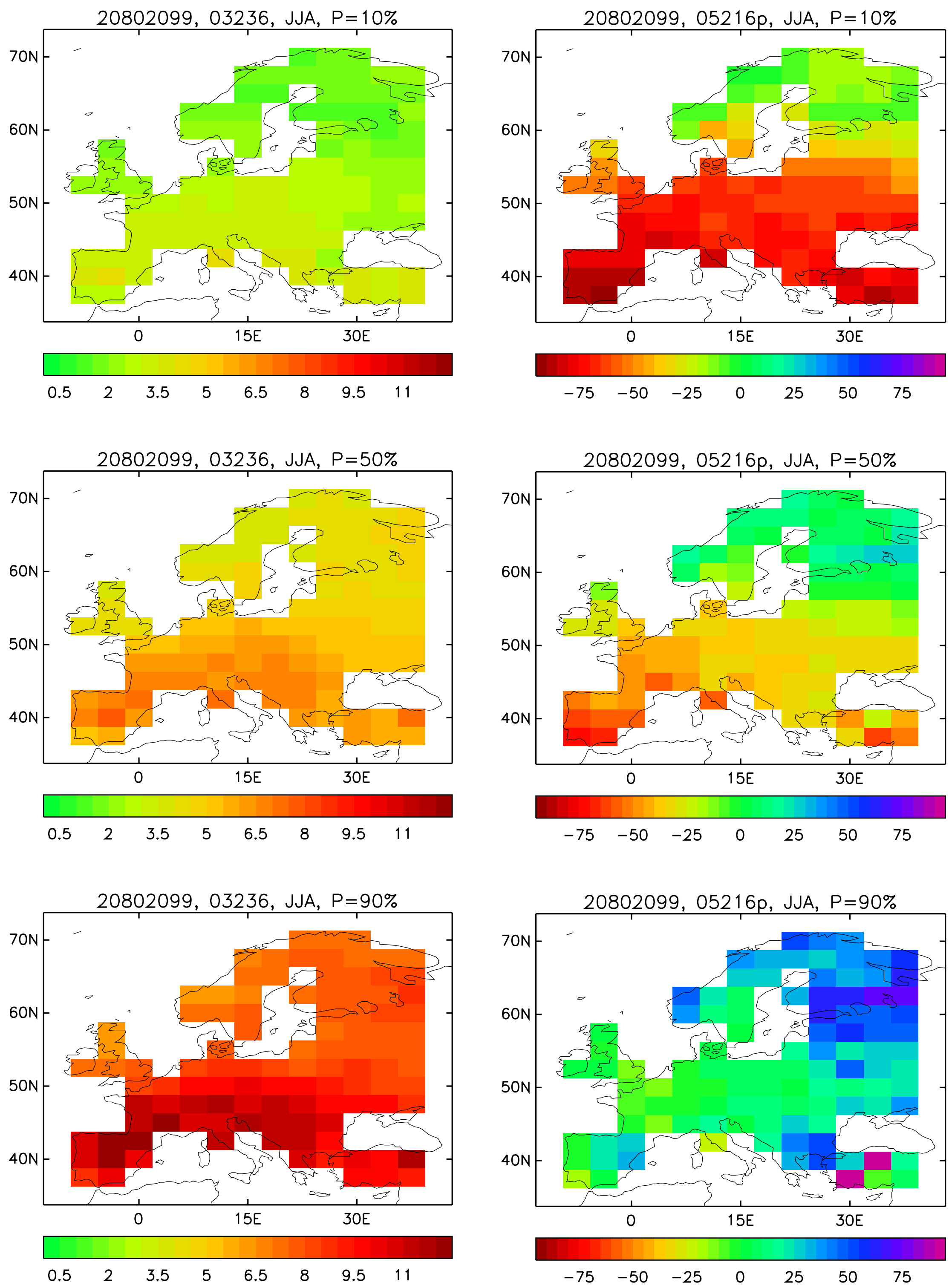

Figure 4.4. *The ENSEMBLES probabilistic projections for Europe under the A1B emission scenario. The maps show the 10%, 50% (median) and 90% percentiles of (a) European surface temperature change and (b) European percentage precipitation change, for the summer season for the period 2080-2099 relative to the 1961-1990 baseline period.*

Europe in winter. The uncertainty, as measured by the 10-90% range, is large for this time period - as much as 10ºC in some locations. This is due to a combination of factors; parameter uncertainty in HadCM3, structural uncertainty from the CMIP3 ensemble, carbon cycle feedback uncertainty, internal variability and time-scaling uncertainty. No single source of uncertainty dominates. For the projections of changes in precipitation, the canonical signals of summer Mediterranean drying and winter Northern Europe wetting are evident, but again the uncertainty range is wide. For many grid boxes there are significant probabilities of both drier and wetter future climates and this may be important for impacts studies.

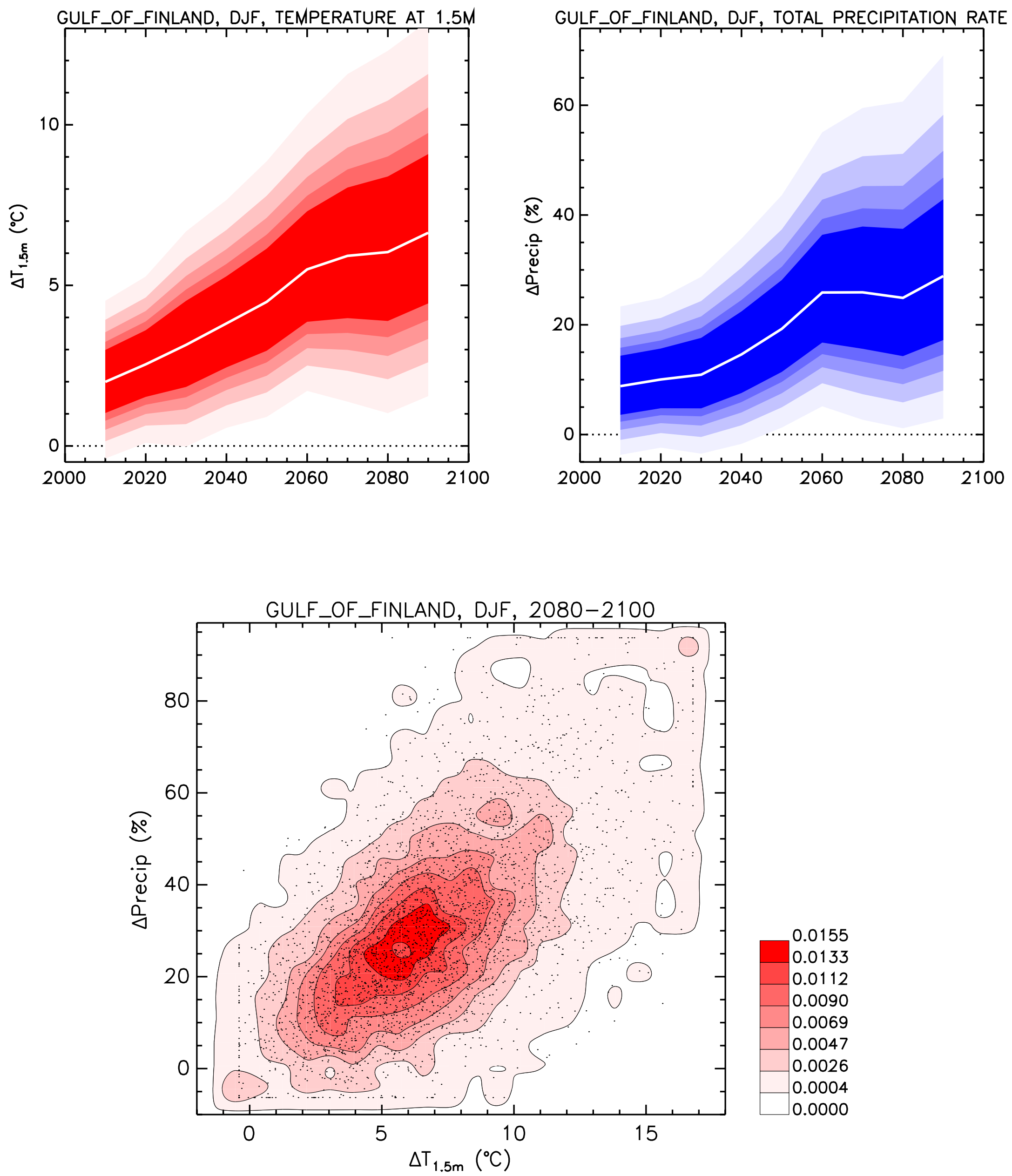

Figure 4.5. *Evolution of the median (white curve) and the 50, 60, 70, 80 and 90% confidence ranges for: (a) 20 year mean winter surface temperature change for the Gulf of Finland grid point; (b) percentage change in 20 year mean winter precipitation for the Gulf of Finland; (c) contours of the Winsorised sampled joint probability distribution function for surface temperature change and percentage precipitation change for the winter season for the Gulf of Finland, for the period 2080-2099 relative to the 1961-1990 baseline period.*

Figure 4.5 shows alternative formats in which pdfs can be presented; in this case the distributions are for the Gulf of Finland grid box. Plumes show changes in temperature and precipitation through the 21st century under the A1B scenario. The data underlying these figures are supplied in numerical form in terms of 10,000 distribution sample points per grid box and can be presented as a contour plot in order to highlight potential relationships between variables.

5. Summary and outlook

.

During the course of ENSEMBLES, new systems for seasonal to decadal forecasting and multidecadal projections of climate change have been developed in Research Theme 1. Substantial effort has resulted in the supply of improved climate models, an improved database of ocean observations, and better methods of using these and other observations to initialise the models for near-term climate forecasts. In addition, there has been a major focus on methods for improved quantification of the inevitable uncertainties arising from model imperfections, as well as from internal climate variability.

As a result, ENSEMBLES has delivered three approaches for seasonal to decadal prediction, consisting of an updated method based on a multi-model ensemble of models drawn from different European modelling centres, plus two novel methodologies applying either stochastic or sustained perturbations to the outputs of physical parameterisation schemes in the atmospheric component of a single model. Major Milestone MM1.2 reported progress in building and assessing these new systems based on a first stream of hindcast experiments, and the present report documents an updated set of systems based on improved models and initialisation strategies, assessed using an expanded set of hindcasts initialised during the period 1960-2005. For seasonal to annual forecasts, the multi-model ensemble results demonstrate performance competitive with, and in some aspects superior to, the previous European system from DEMETER (e.g. Weisheimer et al., 2009). The multi-model results also provide levels of hindcast skill and estimates of the associated uncertainties comparable to or (in some aspects) slightly better than those of the new methodologies. However, our demonstration that the stochastic physics and perturbed parameter methods can provide results of similar utility to multi-model hindcasts is itself a significant result, as there is potential to optimise the design of such systems to improve performance in future, something that has not been attempted during ENSEMBLES.

The decadal hindcasts build on pioneering studies with individual models by several ENSEMBLES partners (Smith et al., 2007; Keenlyside et al., 2008; Pohlmann et al., 2009), and constitute the first coordinated international experiment in this area. Skill is found in projections of surface temperature at large regional scales, for multiyear averages out to a decade ahead. The skill is found to be greater for hindcasts started from more recent dates (1990s and later), compared to the 1960s-1980s, due mainly to the emerging climate change signal. Combining hindcasts from different model versions, in either multi-model or perturbed parameter ensembles, is found to increase skill, as is also the case in the seasonal to annual hindcasts. The added value of initialising hindcasts from observations was also assessed, finding evidence of a modest but potentially useful enhancement to the skill arising from forced climate change. The ENSEMBLES effort places European groups in the vanguard of emerging worldwide efforts to provide better information on climate variability and change for the decadal time scale (Meehl et al., 2009), and this new arm of climate research is likely to develop significantly in future.

For decadal to centennial projections, a series of perturbed parameter ensembles were designed and run to sample uncertainties in key processes, sampling carbon cycle feedbacks in addition to uncertainties arising from surface and atmospheric feedbacks, ocean transport, and man-made forcing from sulphate aerosols. The model simulations were processed using an advanced statistical framework designed to support the inference of probabilistic projections of real world systems from complex but imperfect models of those systems (Rougier, 2007). The methodology allows us to produce thousands of plausible outcomes for 20-year average temperature and precipitation changes during the 21st century for Europe. The effects of structural model errors in atmospheric processes are accounted for by using the perturbed parameter ensembles to "predict" the results of an ensemble of alternative international climate models from the CMIP3 archive used in the IPCC AR4 assessment (Meehl et al., 2007), and the projections are also constrained by a multivariate set of observations of historical climate, consisting of time-averaged seasonal patterns of several key climate variables, plus historical changes in large scale surface temperature patterns during the 20th century. This allows us to present changes at spatial scales resolved by global climate models, noting that the same methodology with an additional downscaling procedure has recently been used to provide probabilistic projections at 25km resolution for the UK (Murphy et al., 2009). The projections should be interpreted as an attempt to quantify the relative probability of different future outcomes, consistent with climate modelling technology, physical understanding and observational evidence currently available. Inevitably they cannot account for errors common to all current climate models, or feedback processes yet to be discovered or not yet included in models (e.g. methane cycle processes). The results also depend on the chosen methodology, which includes expert choices such as the assumed prior distributions of plausible values for uncertain model parameters.

While the work summarised in this report represents a substantial development in capabilities compared to the state of the art at the beginning of the project, there remains considerable scope to develop, improve and combine the techniques considered during the project.

1. The stochastic parameterisation and perturbed parameter techniques were each applied to a different single modelling system in the s2d hindcast work. This meant that comparisons of performance between different systems were complicated by the use of different models and different initialisation techniques. In future, a cleaner and more comprehensive approach would be to apply the techniques to reach of several models, using a common approach to initialisation and the treatment of external forcing agents.

2. There is potential to optimise the performance of the stochastic parameterisation and perturbed parameter approaches in s2d prediction, by selecting values for variables controlling the stochastic or sustained perturbations (respectively) to maximise some chosen set of metrics of s2d hindcast skill. This was beyond the scope of the ENSEMBLES work.

3. There was little opportunity in ENSEMBLES to consider the potential for combining the methods of sampling modelling uncertainties in the s2d work. A preliminary study (D1.17) was performed, assessing the potential to improve skill when hindcasts from the three methods are simply combined to form a larger ensemble. In the work on multidecadal climate projections, a more sophisticated approach was taken to combine results from perturbed parameter and multi-model approaches, using multi-model results to add the effects of structural model errors not sampled in the perturbed parameter projections. However, no attempt was made to include stochastic parameterisation approaches in these projections. In future, there is potential to develop generalised approaches to sampling modelling uncertainties, recognising that the three approaches considered during ENSEMBLES are essentially complementary: the multi-model

approach samples structural variations in model formulation, but does not systematically explore parameter uncertainties for a given set of structural choices, whereas the perturbed parameter approach does the reverse. The stochastic physics approach recognises the uncertainty inherent in inferring the effects of parameterised processes from grid box average variables that cannot account for unresolved sub-grid scale organisations in the modelled flow, whereas the other methods do not. A future system in which both stochastic physics and perturbed parameter techniques were applied to each of several models built using alternative but plausible structural choices (in their grids, choice of advection scheme, basic assumptions in their physical parameterisation schemes, etc) would likely provide a more comprehensive basis for production of probabilistic projections.

4. In ENSEMBLES we developed separate systems for s2d and multidecadal climate projections, recognising that the development of a single "seamless" system (Palmer et al., 2008) to address seasonal to centennial time scales was beyond the scope and resources of the project. However, seamless prediction offers significant potential advantages, including a possible ability to constrain some aspects of long term climate projections through initialisation of multidecadal aspects of internal climate variability such as the Atlantic Multidecadal Oscillation. The fact that ensemble climate forecasts on seasonal to decadal time scales can be verified against observations or analyses make them potentially a powerful tool for reducing uncertainty in longer term projections, by adding to the set of model performance metrics available for use as observational constraints. More effective constraints could be achieved by combining metrics which to date have only been considered in isolation, including those related to the simulation of long term historical climate and measures related to more detailed process-based evaluation of models, as well as measures related to s2d hindcast verification. The prospects for achieving improved projections and uncertainty estimates from a seamless approach depends on the extent to which common physical processes account for forecast anomalies and errors on different time scales (e.g. Scaife et al., 2009; Palmer et al., 2009b). A preliminary study during ENSEMBLES (Deliverable D1.16) found modest relationships between s2d errors and multidecadal climate change projections for regional temperature and precipitation anomalies in a perturbed parameter ensemble. Much more work is needed to assess the potential of seamless approaches, and this seems likely to be an emerging focus for climate research in coming years (e.g. Hurrell et al., 2009).
5. In this context, the initialisation of model projections is also a key future theme. The importance of ocean initialisation is already recognised (e.g. Meehl et al., 2009), noting the importance of maintaining recent improvements to the coverage of ocean observations achieved through the Argo programme, and of further improving methods for analysing and assimilating these observations into climate models (beyond the progress achieved in ENSEMBLES and elsewhere). Ultimately, fully-coupled data assimilation schemes, that take advantage of covariances between ocean and atmosphere variables to generate an optimal estimate of the climate system, are likely to offer the most comprehensive approach to initialisation (e.g. Sugiura et al, 2008). Improving the initialisation of other slowly-varying components of the climate system (sea-ice, snow cover and soil moisture) may offer potential for improved s2d projections. To some extent this may be achievable through the indirect impact of the initialisation of primary atmospheric and ocean variables (as implemented, for example, by ENSEMBLES s2d partners), however, direct assimilation of relevant observations is also likely to be important, including those likely to become available from forthcoming satellite missions (Meehl et al., 2009). Realistic representation of the uncertainties associated

with initialisation will also be a key issue. The use of an ensemble of alternative ocean reanalyses in ENSEMBLES (with associated perturbation strategies) has undoubtedly been a significant step forward in this respect, however the task of fully and efficiently representing initial state uncertainties in ensemble climate projections remains a research challenge.

6. Efforts were made by the participating groups to include external forcing agents in the s2d projections, however there is scope to improve the consistency and comprehensiveness of the approach. Ideally, all groups should include the anthropogenic forcing due to greenhouse gases and aerosols, and the natural forcing arising from solar variability and major volcanic eruptions, given that these are all potentially important sources of predictability in decadal forecasts (e.g. Hawkins and Sutton, 2009; Meehl et al., 2008; Shiogama et al., 2009).
7. The prediction systems developed in RT1 involved ensembles of global climate models, but did not consider the issue of downscaling to scales below that of a typical global model grid box (typical spatial scale ~100x100km^2). While dynamical regional climate models and statistical methods were extensively used elsewhere in ENSEMBLES (RTs 2B and 3) to provide downscaled regional projections, this separation meant that it was not possible to present probabilistic projections at local scales which combined the effects of uncertainties in large-scale and regional processes in an end-to-end system. The UKCP09 probabilistic climate change projections (Murphy et al., 2009) provide one example of how the techniques developed in RT1 could be extended to provide such a system, however there are many open research questions remaining to be addressed in how best to characterise uncertainties at the fine spatial scales required for many impact studies. These include: To what extent global high resolution models (see below) are needed to provide credible regional projections (cf cheaper options involving global model ensembles of lower resolution augmented by downscaling strategies); how best to identify and combine observational constraints relevant to large scale climate processes with constraints more specific to regional processes; how best to sample different types of model uncertainty in both global and regional climate model ensembles (if large ensembles at globally high resolution are not feasible), and whether the concept of end-to-end prediction systems should be extended to include impacts models.

The points above identify some priorities for the development of improved ensemble prediction systems in future, but the development of better climate models with higher horizontal and vertical resolution is also widely recognised as a priority (e.g. Hurrell et al., 2009). Better models are needed, for example, to improve the simulation of detailed processes associated with clouds and convection, tropical intraseasonal variability, blocking, troposphere-stratosphere interactions and mechanisms of coupled ocean-atmosphere variability on decadal time scales. The combined impact of these future requirements to support better climate predictions will create a need for a major improvement in the level of computational resources available to modelling groups worldwide (Shukla et al., 2009).

References

Berner, J., F.J. Doblas-Reyes, T.N. Palmer, G. Shutts and A. Weisheimer (2008): Impact of a quasi-stochastic cellular automaton backscatter scheme on the systematic error and seasonal prediction skill of a global climate model. *Phil. Trans. R. Soc. A*, **366**, 2561-2579, DOI:10.1098/rsta.2008.0033.

Booth, B.B.B., C.D. Jones, M. Collins, I.J. Totterdell, P.M. Cox, S. Sitch, C. Huntingford and R. Betts (2009): Global warming uncertainties due to carbon cycle feedbacks exceed those due to CO_2 emissions. Submitted to *Nature Geoscience*.

Brierley, C., A.J. Thorpe and M. Collins (2009): An example of the dependency of the transient climate response to the warmth of the modelled climate state. *Atmospheric Science Letters*, **10,** 23-28, DOI: 10.1002/asl.205.

Brierley, C., M. Collins and A. J. Thorpe (2009): The impact of perturbations to ocean-model parameters on climate and climate change in a coupled model. *Clim. Dyn.*, in press. DOI 10.1007/s00382-008-0486.

Collins, M., C.M. Brierley, M. MacVean, B.B.B. Booth and G.R. Harris (2007): The sensitivity of the rate of transient climate change to ocean physics perturbations. *J. Climate*, **20**, 2315–2320.

Collins, M., B.B.B. Booth, B. Bhaskaran, G.R. Harris, J.M. Murphy, D.M.H. Sexton and M.J. Webb (2009): Climate model errors, feedbacks and forcings. A comparison of perturbed physics and multi-model ensembles. Submitted to *Clim. Dyn*.

Doblas-Reyes, F.J., A. Weisheimer, M. Déqué, N. Keenlyside, M. McVean, J.M. Murphy, P. Rogel, D. Smith and T.N. Palmer (2009): Addressing model uncertainty in seasonal and annual dynamical seasonal forecasts. *Quart. J. R. Meteorol. Soc.*, **135**, 1538-1559. DOI: 10.1002/qj.464.

Friedlingstein, P., et al. (2006): Climate-carbon cycle feedback analysis: Results from the C^4MIP model intercomparison. *J. Clim.*, **19**, 3337-3353.

Giorgi, F. and R. Francisco (2000): Uncertainties in regional climate change prediction: a regional analysis of ensemble simulations with the HADCM2 coupled AOGCM, *Clim. Dyn*., **16**, 169-182.

Goldstein, M. and J.C. Rougier (2004): Probabilistic formulations for transferring inferences from mathematical models to physical systems. *SIAM J. Sci. Comp.*, **26**, 467-487.

Harris, G.R., D.M.H. Sexton, B.B.B. Booth, M. Collins, J.M. Murphy and M.J. Webb (2006): Frequency Distributions of Transient Regional Climate Change from Perturbed Physics Ensembles of General Circulation Model Simulations. *Clim. Dyn*., **27**, 357-375.

Harris, G.R., M. Collins, D.M.H. Sexton, J.M. Murphy and B.B.B. Booth (2009): Probabilistic Predictions for 21st Century European Climate. *Natural Hazards and Earth System Sciences*, submitted.

Hawkins, E. and R. Sutton (2009): The potential to narrow uncertainty in regional climate predictions. *Bull. Amer. Met. Soc*., doi: 10.1175/2009BAMS2607.1.

Hurrell, J., G. Meehl, D. Bader, T. Delworth, B. Kirtman and B. Wielicki (2009): A unified modeling approach to climate system prediction. *Bull. Amer. Met. Soc*., In Press.

Ingleby B. and M. Huddleston (2007): Quality control of ocean temperature and salinity profiles - historical and real-time data. *J. Mar. Sys.*, **65**,148–175.

Keenlyside, N. S., M. Latif, J. Jungclaus, L. Kornblueh, and E. Roeckner, 2008: Advancing decadal-scale climate prediction in the North Atlantic Sector. *Nature*, **453,** 84–88.

Meehl, G. A., T. F. Stocker, W.D. Collins, P. Friedlingstein, A.T. Gaye, J.M. Gregory, A. Kitoh, R. Knutti, J.M. Murphy, A. Noda, S.C.B. Raper, I.G. Watterson, A.J. Weaver and Z. –C. Zhao (2007): Global Climate Projections. In: *Climate Change 2007: The Physical Science Basis*. *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 747-845.

Meehl, G.A., J.M. Arblaster, G. Branstator, and H. Van Loon, 2008: A coupled air-sea response mechanism to solar forcing in the Pacific region. *J. Clim.*, **21**, 2883-2897.

Meehl, G. A., L. Goddard, J. Murphy, R. J. Stouffer, G. Boer, G. Danabasoglu, K. Dixon, M. A. Giorgetta, A. Greene, E. Hawkins, G. Hegerl, D. Karoly, N. Keenlyside, M. Kimoto, B. Kirtman, A. Navarra, R. Pulwarty, D. Smith, D. Stammer, and T. Stockdale (2009): Decadal Prediction: Can it be skillful? *Bull. Am. Meteorol. Soc.*, **90**, 1467-1485. doi 10.1175/2009BAMS2778.1

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.

Murphy, J.M., B.B.B. Booth, M. Collins, G.R. Harris, D.M.H. Sexton and M.J. Webb (2007): A methodology for probabilistic predictions of regional climate change from perturbed physics ensembles. *Phil. Trans. R. Soc. A*, **365**, 1993-2028.

Murphy, J.M., D.M.H. Sexton, G.J. Jenkins, P.M. Boorman, B.B.B. Booth, C.C. Brown, R.T. Clark, M. Collins, G.R. Harris, E.J. Kendon, R.A. Betts, S.J. Brown, T.P. Howard, K.A. Humphrey, M.P. McCarthy, R.E. McDonald, A. Stephens, C. Wallace, R. Warren, R. Wilby, R.A. Wood (2009). UK Climate Projections Science Report: Climate change projections. Met Office Hadley Centre, Exeter, U.K.

Niehörster, F. (2009): Uncertainty in Feedback Mechanisms in Climate Change Projections, PhD Thesis, Freie Universität Berlin.

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

Palmer, T.N. and co-authors (2004): Development of a European multi-model ensemble system for seasonal to inter-annual prediction (DEMETER). *BAMS*, **85**, 853-872.

Palmer, T.N., F. Doblas-Reyes, A. Weisheimer, and M. Rodwell (2008): Reliability of climate change projections of precipitation: Towards "seamless" climate predictions. *Bull. Amer. Meteor. Soc.*, **89**, 459-470, DOI: 10.1175/BAMS-89-4-459.

Palmer, T.N., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G.J. Shutts, M. Steinheimer and A. Weisheimer (2009a). Stochastic parametrization and model uncertainty. *ECMWF Tech. Memo*. **598**, 42pp.

Palmer, T.N., F. Doblas-Reyes, A. Weisheimer, and M. Rodwell (2009b): REPLY to: COMMENTS: Toward Seamless Prediction, Calibration of Climate Change Projections Using Seasonal Forecasts by Scaife, A.A., C. Buontempo, M. Ringer, M. Sanderson, C. Gordon and J. Mitchell, *Bull. Amer. Met. Soc*., **90**, 1551-1554.doi: 10.1175/2009BAMS2916.1

Pohlmann, H., J. H. Jungclaus, A. Köhl, D. Stammer, and J. Marotzke, 2009: Initializing Decadal Climate Predictions with the GECCO Oceanic Synthesis: Effects on the North Atlantic. *J. Clim.*, **preprint,** 0000-0000.

Rougier, J.C. (2007): Probabilistic inference for future climate using an ensemble of climate model evaluations. *Clim. Change,* **81**, 247-264.

Scaife, A. A., C. Buontempo, M. Ringer, M. Sanderson, C. Gordon and J. Mitchell (2009): COMMENTS: Toward Seamless Prediction, Calibration of Climate Change Projections Using Seasonal Forecasts. *Bull. Amer. Met. Soc*., **90**, 1549-1551.doi: 10.1175/2009BAMS2753.1

Shiogama, H., S. Emori, T. Yokohata, M. Ishii, T. Nozawa, and M. Kimoto (2009): Can decadal predictability of the natural variability survive against disturbances due to a large volcanic activity in near-term climate predictions? *Geophys. Res. Lett.*, submitted.

Shukla, J., R. Hagedorn, B. Hoskins, J. Kinter, J. Marotzke, M. Miller, T.N. Palmer and J. Slingo (2009): Revolution in climate prediction is both necessary and possible: A declaration at the World modelling Summit for Climate Prediction. *Bull. Amer. Met. Soc.*, **90**, 175-178. doi: 10.1175/2008BAMS2759.1

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

Smith, D. M., S. Cusack, A. W. Colman, C. K. Folland, G. R. Harris, et al., 2007: Improved surface temperature prediction for the coming decade from a global climate model. *Science*, **317,** 796-799, doi:10.1126/science.1139540.

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.

Sugiura N., T. Awaji, S. Masuda, T. Mochizuki, T. Toyoda, T. Miyama, H. Igarashi, Y. Ishikawa (2008): Development of a four-dimensional variational coupled data assimilation system for enhanced analysis and prediction of seasonal to interannual climate variations. *J. Geophys. Res*., **113**, C10017, doi:10.1029/2008JC004741

Webb, M. J., C. A. Senior, D. M. H. Sexton, W. J. Ingram, K. D. Williams, M. A. Ringer, B. J. McAvaney, R. Colman, B. J. Soden, R. Gudgel, T. Knutson, S. Emori, T. Ogura, Y. Tsushima, N. Andronova, B. Li, I. Musat, S. Bony and K. E. Taylor (2006): On the contribution of local feedback mechanisms to the range of climate sensitivity in two GCM ensembles. *Clim. Dyn*., **27**, 17–38.

Weisheimer, A., J. Murphy and M. Collins (2006): Provision of a “first generation” ensemble prediction system (Version 1) for use in RT2. ENSEMBLES Major Milestone MM1.2.

Weisheimer, A., F. Doblas-Reyes, P. Rogel, E. Da Costa, N. Keenlyside, M. Balmaseda, J. Murphy, D. Smith, M. Collins, B. Bhaskaran, and T. Palmer (2007): Initialisation strategies for decadal hindcasts for the 1960-2005 period within the ENSEMBLES project. *ECMWF Technical Memorandum*, **521.**

Weisheimer, A., F.J. Doblas-Reyes, T.N. Palmer, A. Alessandri, A. Arribas, M. Deque, N. Keenlyside, M. MacVean, A. Navarra and P. Rogel (2009): ENSEMBLES - a new multi-model ensemble for seasonal-to-annual predictions: Skill and progress beyond DEMETER in forecasting tropical Pacific SSTs. *Geophys. Res. Lett*., **36**, L21711, DOI: 10.1029/2009GL040896..

Yokohata, T., M.J. Webb, M. Collins, K.D. Williams, M. Yoshimori, J.C. Hargreaves and J.D. Annan (2009): Structural similarities and differences in climate responses to CO_2 increase between two perturbed physics ensembles. *J. Climate*., submitted.