Coordinated experimentation to study multi-decadal prediction and near-term climate change

Version 2.0 December 20, 2007

Aims: This proposal describes a framework for coordinating intended experimentation covering two different but related objectives:

O1: Short-term prediction of climate for the next 30 years [to 2030 or 2035]
O2: Developing the science of multi-decadal prediction in the context of a changing climate

Experimentation addressing Objective 1 was called for at the Aspen meeting of 2006, is currently being planned by various groups, in many cases using high resolution models, and is expected by the wider international community. However, many questions remain about how best to initialize such forecasts, and how to assess the uncertainties in the resulting predictions. Experiments to assess and develop the science of multi-decadal prediction (Objective 2) are thus needed, and individual groups are already starting to work in this area.

By requesting a coordinated set of low resolution experiments, and a common framework in which individual groups explore ideas and sensitivities, and by linking the framework of the scientific development (often at low resolution) with the requested "best guess" predictions to 2035 (often at high resolution), scientific return is maximized and comparisons between the various experiments are facilitated.

The rest of this document outlines (i) the requested mandatory and optional runs for coordinated experimentation for each of the two objectives; (ii) the framework concept, whereby the principles of how experimentation is supposed to fit together is explained, and additional proposals for coordinated work are requested; and (iii) a discussion of the reasoning behind this proposal, and a response to some commonly asked questions.

Objective 1 experiments

Aims:

> Provide model integrations to allow estimation of the evolution of expected climate for the period 2005-2035, relative to the climate of recent decades.
>
> Encourage use of higher resolution climate models, with the hope of better resolving synoptic processes associated with extremes, and assessing the benefits of higher resolution in general.

The final analysis of expected climate in 2005-2035 should:

- Aim to give guidance on the changing risk of extremes
- Aim to give guidance on the possibility of changes in the monsoons
- Assess likely errors in the prediction, based on the spread of results from different models and different initialization techniques, and the errors seen in the ability of

initialized climate forecasts to reproduce the observed changes of the last few decades - results from Objective 2 experimentation will be needed for this.

- Be a step towards probabilistic forecasts of near-term future climate, using a variety of methods and sources, including multi-decadal model predictions, empirical predictions based on attribution of observed changes to date, and results of other modeling studies.

Requested model runs:

30 year integrations with initial dates 1st November 1960, 1980 and 2005.
Each start date to be run with a 3 member ensemble, optionally to be increased to O(10)
Ocean initial conditions must represent in some measure the observed anomalies for the start date.

Model run time: 270 years (optionally, an additional 630 years)

Details of model runs:

- Actual integration length should be 30 years and 2 months, to complete the calendar year. We expect the first two months to be discarded in the analysis.
- Choice of initial conditions is up to each group, subject to the principle that they should represent the observed anomalies for the start date. Possible initialization strategies are discussed below.
- All forcings should be included as observed values for past dates, with prescribed concentrations of well-mixed GHGs. The details should be the same as used in the CMIP5 historical (20th C) runs, with the same flexibility on the treatment of ozone and aerosol and the same specified observational datasets.
- For future dates, a single GHG concentration scenario should be used. We propose to use the CMIP5 “medium stabilization” scenario, which prescribes concentrations leading to an eventual 4.5 W/m2 radiative forcing. Again, specification of reactive species and aerosols will exactly follow those used in the CMIP5 “medium stabilization” future concentration scenario. For the purposes of Objective 1, a single scenario for anthropogenic aerosol is preferred.
- Volcanic aerosol has a specific treatment: observed values should be used for past dates, as per CMIP5, but values to be used for the future should be specified based on the assumption of no further volcanic eruptions. The model runs are thus configured to predict what will happen to climate if no major eruptions take place, which is a possible outcome for a thirty year period. Sensitivity of predicted climate to hypothetical volcanic eruptions should be explored under Objective 2, with lower resolution models.
- Any deviations from the above specifications should be properly documented.

Note that there is no requirement for the initialization procedure to guarantee that the coupled model forecasts start at a given offset to a fully spun-up equilibrium for pre-industrial climate, and no requirement for a “control” run of the coupled model with constant radiative forcing to assess climate drift. The aim of the experiment is to predict the climate of the next few decades relative to the climate of the last few decades, not relative to a hypothetical “no emissions” world. This is why at least one “past” start date is required. Two past start dates (as requested here) are better, since they can be averaged to give a more stable reference, and the difference between them can be checked against reality and against corresponding lower resolution runs. The interpretation of the Objective 1 runs to give a prediction of the likely evolution of climate will be greatly strengthened by relating the results to those obtained with the Objective 2 runs.

The optional extension of the ensemble size from 3 to O(10) might be expensive for high resolution models. The reasons for wanting a larger ensemble are that it will give better sampled resolution of modest shifts in climate (which will aid in interpretation of results, for example the difference between low and high resolution runs); and that it will allow a better look at “extreme events”, in the sense of infrequent but high impact events. The latter point can also be investigated with large ensembles of low resolution models - how similar the results might be is not known. The decision on whether to run the additional ensemble members will be up to each group (and may depend on the performance of the first three ensemble members).

Objective 2 experiments

Aims:

A multi-model study of the sensitivity of multi-decadal forecasts to initialization method and model

Characterize the errors and uncertainties in multi-decadal predictions

Characterize the impact of uncertainties in aerosol forcing (volcanic and anthropogenic)

Estimate changes in the chemistry of the atmosphere over the next 30 years

In particular:

- the role of initial conditions, and the methods by which they can be specified
- the value of higher resolution models
- sensitivity to choice of model, and effectiveness of multi-model composites
- testing actual level of error in decadal predictions against that expected from ensemble spread and estimated initial condition uncertainty
- uncertainty arising from unknown future forcing (tropospheric aerosols, volcanism)
- comparison of predictions based on different techniques

Requested model runs:

2.1 30 year integrations with initial dates 1st November 1960, 1980 and 2005.
Each start date to use a 3 member ensemble, optionally to be increased to O(10)

Ocean initial conditions must represent in some measure the observed anomalies for the start date.

Model run time: 270 years (optionally, an additional 630 years)

These runs exactly match the runs in Objective 1. If the Objective 1 runs are carried out with a high resolution model, then they should be repeated with a low resolution version of the same model and same initialization method for the purposes of Objective 2. In this case, the runs act as a first comparator for the low-resolution experimentation described below, can be used to translate between low and high resolution results, and allow a direct estimate of the sensitivity of results to resolution for a given model. If the Objective 1 runs are not carried out with a high resolution model, then a single set of runs with a low resolution model will be used for the purposes of both Objective 1 and Objective 2.

2.2 10 year integrations with initial dates of 1st November 1965, 1970, 1975, 1985, 1990, 1995 and 2000.
Ensemble size of 3, optionally to be increased to O(10)
Ocean initial conditions represent the observed anomalies

Model run time: 210 years (optionally, an additional 490 years)

These runs extend those of 2.1 to a wider range of start dates, but are only 10 years long. They allow a better estimate of the ability of the model and initialization technique to capture the impact of the initial conditions on the first decade of the forecast, and will also give an estimate of the predictability of climate over a decade, as estimated by the model. An ensemble size of O(10) will be particularly helpful in this regard. The runs will also give information on how typical the chosen main historical start dates (1960 and 1980) are in terms of the ability of the model to make decadal forecasts, and will thus give some help in refining the calibration of the actual 30 year forecast from 2005. Some groups may want to extend the length of the runs to 30 years to help with this latter point. Note that these runs (to 10 years) are already taking place in Europe as part of the ENSEMBLES project.

2.3 Extended ensemble of 20th Century/future climate runs
Additional ensemble members for CMIP5 runs
Start date 1960 or earlier
End date 2035, using same forcing as integrations in 2.1
Augment CMIP5 ensemble size to 3, optionally to O(10).

Model run time: historical runs: 0 years (optionally, typically 375 years)
future runs: 60 years (optionally an additional 210 years)

These runs form a “control” against which the value of initilizing short-term climate and decadal forecasts can be measured. They provide decadal and climate predictions based only on a starting point of pre-industrial climate and a specfied evolution of GHG concentrations and other forcings. The period extending to 2035 gives a 30 year “forecast” made using the traditional method of scenario integrations. The expanded ensemble size of historical and near-term forecast runs may also be of interest to the detection and attribution community.

The runs needed here are only to augment whatever ensemble is generated for CMIP5. At the moment, CMIP5 calls for a single mandatory 150 year historical run, with a recommendation that this be extended to 5 ensemble members. Here we request that the historical runs are extended to O(10) ensemble members, for the period 1960 onwards. Although the most robust method of ensemble generation might be to run the full ensemble from pre-industrial climate, other cheaper approaches might also be considered. Given the emphasis on decadal/multi-decadal timescales, starting the full ensemble 10-30 years before 1960 would enable substantial spread to develop. Also, the possibility of using a pre-existing small ensemble to construct initial conditions for a larger one using linear combination of the ensemble members could be investigated (is non-linearity in the ocean equation of state a significant issue?). Assuming CMIP5 has a 5 member ensemble, an additional 5 members for 1930-2005 would require 375 years of model integration.

CMIP5 only calls for a single ensemble member for each future scenario run. Although the GHG forcings will be identical in these runs and he single CMIP5 “medium stabililization” scenario run, it is possible that the treatment of volcanic aerosols might be different (given the different emphases of the experiments). In this case, we might need to run all the ensemble members specifically for the “short-term climate prediction” scenario - this would only require another 30 years of model integration.

2.4 Intercomparison of initialization strategies (optional - a subset of groups only)

Repeat of runs (2.1) and (2.2) using an alternate initialization strategy.

Model run time: 480-1600 years (all optional)

Assuming that runs (2.1) and (2.2) are made by a number of different models, each with its own initialization scheme, we will learn about the aggregate behaviour and spread of initialized climate forecasts. However, to understand some of the reasons for the differences in forecasts with different systems, and to assess whether some initialization methods tend to work better than others, it will be valuable to have some controlled experimentation where the impact of different initialization schemes on a given model (and the use of different models for a given initialization scheme) can be compared. This can most effectively be done by collaboration between several interested groups. Groups interested in developing this idea should get in touch with the WGCM/WGSIP working group responsible for this proposal.

2.5 Aerosol sensitivity runs

Repeat of the 2.1 2005 forecast with a high and/or low anthropogenic aerosol scenario
Repeat of the 2.1 2005 forecast with an imposed “Pinatubo” eruption in 2010
Ensemble size 3, optionally O(10)

Model run time: 180-270 years (optionally an additional 420-630 years)

The forecasts in 2.1 assume a particular evolution of anthropogenic aerosol, and a lack of major volcanic eruptions, as well as specified GHG concentrations. Although the GHG concentrations are unlikely to vary in aggregate much from the specified values within the 30 year period, anthropogenic and volcanic aerosol could each turn out to differ significantly from the specified values. These experiments aim to quantify the impact of such variation. In the case of volcanic aerosol, the proposal is to specify a "repeat" of the Pinatubo eruption, to take place in 2010.

Experiments to consider the impact of changes in anthropogenic aerosol could also be run. Details of these need to be defined.

2.6 Chemistry runs

To be defined.
As a starting point, simply run the 2.1 integrations with active chemistry.
Ensemble size 3? Or 1?

Model run time: 90-270 years

Groups with chemistry models are invited to propose what they would like to do here. Is simply running the models for the relevant start dates sufficient? Are there sensitivities that they would like to explore?

<u>Discussion of the Framework concept and its implementation</u>

The experiments outlined above fit into a common framework. Each set of experiments makes sense on its own terms, and will deliver useful information. But by placing all of the experimentation in a common framework, we increase our ability to extract results from the total set of experiments, and reduce the need for running of similar-but-slightly-different controls or of having to make unnecessary approximations/assumptions when comparing results. The above experiments do not exhaust the issues relevant to decadal/short-term climate prediction that might be looked at. For example, studies of the impact of vertical resolution, stratospheric processes, sea-ice initialization and prediction might all be looked at. Groups that are planning such studies are encouraged, where appropriate, to fit them into the framework outlined here, to facilitate comparison with the wider set of results.

It is also worth noting that we have in essence two parallel parts of the framework. One is the strategy of initialized short-term climate prediction, as described in 2.1. This does not require long model spin-ups, and is the primary framework used in this proposal. However, the work in 2.5 and 2.6 could also be made using the runs of 2.3 as a reference, ie using traditional transient climate change integrations as the comparator. In terms of the organization of this proposal, the chemistry work could legitimately be placed in a separate "Objective 3" section.

Initialization methods

A variety of initialization methods have already been tried by different groups, and other possibilities are also available. Forecasts from the DEPRESYS system at the Met Office (Smith et al, 2007) are initialized with a purpose-made ocean analysis, which assimilates observed anomalies into an ocean mean state taken from a long run of the coupled model which is then used to make forecasts. This is an anomaly initialization method which depends on a long control run of the coupled model. Keenlyside et al (2007) use the minimalist approach of directly initializing the coupled system with observed SST anomalies. Troccoli and Palmer (2007) take the direct approach of initializing the ocean with the best possible analysis of the actual state, as is often done for seasonal prediction. This latter method is immediately accessible to anyone using an ocean model for which such analyses exist - no long coupled runs are needed to define a model mean state. Pohlmann et al (2007) nudge the ocean component of the coupled system towards T and S anomalies defined from an ECCO analysis made with the same ocean model.

With the exception of Troccoli and Palmer, who make *a posteriori* corrections to the forecasts to account for drift, the above methods try to initialize the system on the coupled model attractor. Some results suggest that for a warming world, predictions of global mean temperature have little sensitivity to initial conditions beyond the first few years. However, predictions of ocean heat content, sea level, regional anomalies and regional climate may well be more sensitive.

Discussion amongst the groups who are planning high resolution coupled runs might be helpful in clarifying the best methodology to use, as might early experimentation with low resolution models. One potentially useful technology that seems not yet to be established is the ability to transfer an ocean analysis (perhaps in the form of anomalies) from one ocean model to another. It is also worth noting that a wider set of results from the ENSEMBLES decadal prediction studies should become available over the next year or so.

Other issues

- The ENSEMBLES project is running decadal prediction experiments with start dates 1960, 1965, ... 2005, but with volcanic aerosol from eruptions after the start of the forecast excluded, and using GHG concentrations from an older scenario. Despite these slight differences, it is expected that the ENSEMBLES runs will form a core base for the 2.2 runs described above. Although most runs will be 10 years long, some models will run 30 years.
- Given the observed behaviour of Arctic ice in the last few years, initialization and forecast of sea-ice is an issue which groups will need to address.
- The output of the model integrations needs to be defined. We start with the assumption that the output will be the same as that requested for the main CMIP5 runs. But this may need additions or subtractions to meet particular needs, eg in characterizing extremes, and/or to be feasible for the high resolution runs.

- Data handling needs to be defined. It is hoped, but needs to be confirmed, that data from the central parts of this proposal can be archived as part of the general AR5 archive, to allow access by the appropriate part of the scientific community. Some of the additional experimentation may need to be handled separately by the groups involved.
- The protocol calls for 30 year long runs, since we have been told that that is what the community expects. If the last integration starts in November 2005, this means that the forecasts extend to the end of 2035. Is this what is wanted? What is the reason that 30 year forecasts are expected?

Summary of runs

Section	Experiment	minimum # years	optional # years
1	30 year runs (hi res)	270	630
2.1	30 year runs (lo res)	270	630
2.2	10 year runs	210	490
2.3	Additional CMIP5 members	60	585
2.4	Initialization strategies	0	Varies
2.5	Aerosol sensitivity	180	420
Totals	(excluding 2.4)	270 hi-res 720 lo-res	630 hi-res 2125 lo-res

For groups participating in all sections apart from 2.4, the minimum requirement is for 720 years of low resolution integration, increasing to 2845 years if everything is done with a large ensemble size. Runs under 2.4 would be additional to this.

Reasoning behind the proposal

This proposal is intended as a step towards the development of a mature and comprehensive climate prediction capability. It is based on the combined experience of the ACC modelling community and the seasonal prediction community (who have long experience of making coupled ocean-atmosphere forecasts dependent on initial conditions), together with the fledgling community of decadal prediction modellers. There is no doubt that in the long run, we want our forecasts of future climate to be based on the well-observed present, rather than considered as long term trajectories from a distant and sparsely observed "pre-industrial" climate. There is also no doubt that our capabilities to do this are still at a very early stage, which is why the proposed experimentation is less prescriptive than for eg CMIP5, and is designed to allow groups to choose which sections they feel ready to undertake or explore. It should be stressed that groups should not feel obliged to undertake any or all of the work packages described here. What we do ask is that groups should strive to work within this framework once it has been approved (and, critically, provide feedback on what might need changing before such approval is given).

One of the lessons of seasonal prediction is that when we confront models with data in the prediction problem, model error is found to be quite large, even after uncertainties in the initial conditions are carefully accounted for. It may well be that we find the same in the decadal / short-term climate prediction problem, particularly for forecasts initialized in recent, better observed times. Testing initialized models against actual data is a powerful tool to help drive model improvement. The ability to do this across multiple timescales (NWP to seasonal to decadal and beyond) should eventually lead to a maturing of our modelling and prediction capabilities. This document does not discuss the testing of climate change models in seasonal prediction mode, but the Coupled Historical Forecast Project (CHFP) organized by WGSIP/WCRP provides a protocol for such testing, and modelling groups may wish to participate in this.

Comments on this proposal are welcome.

Proposal elaborated by a WGCM/WGSIP/CLIVAR/WCRP sub-group:

Tim Stockdale, Gabi Hegerl , Jerry Meehl, James Murphy, Ron Stouffer, Marco Giorgetta, Masihide Kimoto, Tim Palmer, Wilco Hazeleger, Detlef Stammer, Ben Kirtman and George Boer.

Replies to specific comments already received, and other FAQs:

Q1. Why not have a proper study of resolution dependence?
A1. A full study of resolution dependence of the results would require both initialized forecast and C20th control runs to be made at both high and low resolution. However, this would be very expensive, and is judged not to be appropriate for coordinated experimentation at this stage. By running control and forecast runs at low resolution, we can assess the value of initialization on (low resolution) climate forecasts; by running a limited number of forecast experiments at high resolution we can learn something about the benefits of high resolution versus low resolution in the context where it matters most (ie prediction). A comprehensive test of the benefits of higher resolution might best be made in the future, when we are more confident of our initialization methods, and are more able to afford to run the large number of high resolution cases needed to bolster statistical significance of the results.

Q2. Why start on the 1st November?
A2. When looking at just the multi-decadal prediction problem, the time of year to start the forecasts is of no physical importance. However, for seasonal prediction, standard practice is to start runs at quarterly intervals (1st Feb, 1st May, 1st Aug, 1st Nov),the reason being that this allows a one month initial period to lose deterministic predictability of the synoptic scales, immediately followed by a three month target season (such as DJF). This allows the best fair assessment of the skill of the forecast systems. The CHFP organized by WCRP/WGSIP specifies these start dates. In the European ENSEMBLES project, which combines seasonal, decadal and climate change prediction into a single framework, all of the decadal runs start on the 1st November. For annual forecasts (ie of discrete calendar years), a November start has also been found convenient. Given the desirability of a common forecasting framework across multiple timescales, and perhaps more importantly given the very substantial existing base of 1st November forecasts by the Europeans, it seems sensible to adopt the 1st November convention.

Q3. Is the idea to have a common specified initilization method (or specified initial conditions) for the main experiments, including the forecasts to 2030?
A3. No. The best way to initialize such a forecast is not known, and inserting a common analysis into a number of different coupled models might cause problems. Rather, the idea is for the community to try a range of methods that are representative of the state-of-the-art. The uncertainty in how best to initialize the forecasts is an important part of the uncertainty in the forecast, and we will be pleased if this is (partially) represented by a range of processes.

Q4. Why not do the high resolution runs uncoupled?
A4. Karl Taylor has put forward the option of running the high resolution runs as uncoupled “time-slice” experiments. This is certainly possible, and is one way of dealing with the spin up problems that might otherwise be prohibitively expensive. The proposal put forward here is an alternative way to get affordable high resolution forecasts whose performance can reasonably be characterised. There are several issues to consider when weighing the two approaches:

(i) Prescribed SST runs are physically less realistic, particularly in terms of SST/convection relationships in parts of the West Pacific and elsewhere.
(ii) What do we want from these experiments? Changes in extremes of precipitation events, extremes of windstorms, other changes in atmospheric circulation which might be driven by "global warming" and dependent on model resolution, can all be studied using high resolution runs where the exact specification of SST does not matter (so long as it is broadly consistent with the GHG induced warming).For these studies, the initialization of the coupled forecasts will not be critical, and prescribed SST is also an option.
(iii) Other aspects of short term climate change are changes in ocean heat content (perhaps linked more to changes in wind circulation than diabatically forced) and global mean temperature (again affected by circulation patterns, and perhaps of political importance). Here, initialization does matter. Regional climate change may also depend on the state of the ocean. Low resolution experiments will be vital for getting a handle on some of the uncertainties associated with the initialization problem, and will be our main tool for scientific investigation. Some aspects of regional climate change might be addressable with high resolution atmosphere-only runs forced by SST from intialized low-resolution runs. However, in the worst case, a proper treatment may require both a good initialization and a high resolution coupled model.
(iv) Perhaps one of the motivations for high resolution coupled runs is as a showcase and (more importantly) a technology driver. If we are comfortable that the runs can be initialized and run sensibly, and we provide the right scientific framework of low resolution runs to aid interpretation and progress the science, then the high resolution runs can be welcomed.
(v) It is worth noting that seasonal prediction still works with both 2-tier (uncoupled runs with prescribed SST) and 1-tier (fully coupled) systems, although the 1-tier systems are slowly taking over. A co-existence of high-resolution "time-slice" runs with prescribed SST with coupled short-term climate runs is quite possible.

Q5. Why not completely separate the high resolution runs from the low-resolution "decadal prediction" runs?
A5. This is also a possible approach. Certainly, running low-resolution decadal prediction runs (eg as in the ENSEMBLES project, or as done with DEPRESYS) makes sense as a stand-alone scientific project. However, in reality decadal prediction for the future is more about climate change than natural variablity, so we need to treat decadal prediction in the proper context. Fundamentally, we need to learn how to initialize short term climate predictions in order to maximize society's ability to adapt to the changing climate. This gives strong justification for working on the multi-decadal problem, not just 10 year forecasts.

If we want to do the scientific work at low resolution, and ask the high resolution people to act independently, we have the following problems: (i) the high resolution people need to do something for initialization - and as Karl mentions, the traditional approach involving 500 year control runs etc does not look attractive. (ii) if some method is arbitrarily chosen, we will have little idea as to its robustness or the expected "skill" of the forecasts. By running a limited number of high resolution experiments in the same framework as a larger set of low resolution experiments, we can get a better idea as to what the results actually mean.

Q6. How well can we initialize the ocean?
A6. We should learn more about this from the results we get in the Objective 2 experiments. It is clear from the results from GSOP that at present there are many aspects of the ocean initial state that we know poorly (quite apart from the problem of optimizing the observed initial state for making forecasts), and it is clear that the technology of initialized forecasts needs a lot of development. Nonetheless, a pertinent question is whether we know and can specify the ocean state more accurately than we would obtain from a transient C20 run. Key advantages that an analysis has over a transient run are some knowledge of changes in the wind field over recent decades (one important driver of ocean circulation change), a fairly good knowledge of SST (perhaps especially useful in winter, when it reflects heat content to some depth), and some (limited) information on large scale changes in water masses. It may be that the optimum analysis of the evolution of “anomalies” in the ocean state over the twentieth century requires a different approach to optimizing the description of the mean state, and that relatively “light” use of the data can help, but it is hard to believe that we cannot do better than a transient run with pre-industrial initial conditions. Results from existing decadal prediction work suggests that we can indeed do better. Nonetheless, we will clearly benefit from input by GSOP and others in helping define the best practical strategies, both in the short and longer term.