WCRP REPORT

World Climate Research Programme

Summary Report

Report of the International workshop on seasonal to decadal prediction

13-16 May 2013, Toulouse, France

September 2013

WCRP Report No. 23/2013

International workshop on seasonal to decadal prediction
Toulouse France 13-16 May 2013

TABLE OF CONTENTS

Introduction
(G.J. Boer and L. Terray)

The International Workshop on Seasonal to Decadal Prediction was held in Toulouse, France, from 13 to 16 May 2013 with support from Météo-France, CNRS, CERFACS, WCRP, NOAA and the foundation BNP-Paribas. The goal of the workshop was to review current abilities to make skilful predictions on seasonal to decadal (s2d) timescales and to motivate new initiatives, approaches and studies in this area of active research. The workshop included 58 oral and 80 poster presentations with 137 registered participants over the four days of the meeting. The workshop website (http://www.meteo.fr/cic/meetings/2013/s2d/) displays the programme and provides links to the abstracts and many of the presentations. The members of the Scientific Programme Committee and the Local Arrangements Committee that helped organize the Workshop are listed in the Annex

Figure 1 is a schematic meant to illustrate the comparative importance of initial conditions and external forcing for climate predictions and projections and the connection between seasonal/inter-annual prediction, multi-annual to decadal prediction, and climate change simulations. The timescales of particular interest to the Workshop are indicated by the box.

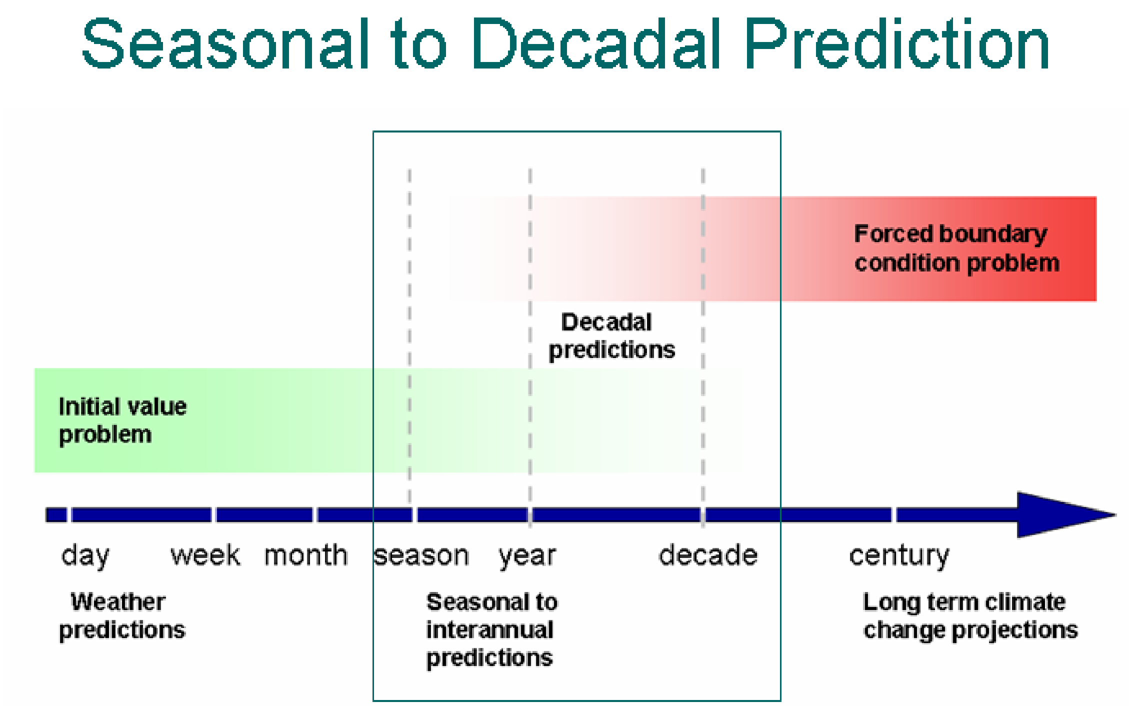

Figure 1. Schematic illustrating the timescales considered and the importance of initial condition and external forcing information for seasonal to decadal prediction.

This is the first workshop, to our knowledge, that attempts to combine the expertise of the seasonal forecasting and climate simulation communities. It follows the WCRP Workshop on Seasonal Prediction held in Barcelona, Spain by some six years and is a response to the continued interest in seasonal prediction and a growing interest in predictions on annual to decadal timescales. The workshop is part of the ongoing melding of climate prediction and climate simulation research spanning timescales from months to centuries. Two World Climate Research Program (WCRP) working groups, namely the Working Groups on Seasonal to Interannual Prediction (WGSIP) and the Working Group on Climate Modelling (WGCM) were involved in the organization of the workshop.

The focus of the Working Group on Climate Modelling is the simulation of climate and climate change using coupled global climate models. The climate is determined by the externally prescribed forcing and climate change is the result of changes to this forcing such as those associated with increasing CO2 and other greenhouse gases, atmospheric aerosols, land use change, solar variation etc. A climate simulation is intended to produce the kinds of dynamic and thermodynamic behaviour that is characteristic of the coupled climate system without attempting to forecast the details of the evolution that is observed. A range of possible evolutions, consistent with the forcing, is produced and the statistics of temperature, precipitation or other climate variables are considered. At the long timescales considered the simulated climate statistics do not depend on initial conditions.

The focus of the Working Group on Seasonal to Interannual Prediction is indicated by its name. By contrast to a climate simulation, a seasonal prediction depends critically on observation-based initialization of the coupled system. The resulting forecasts are intended to give information about the likely evolution of the observed climate system over the next seasons. However, as the forecast range increases the importance of the initial conditions declines and the effects of external forcing begin to become more important as in Figure 1.

The aspects of seasonal to decadal (s2d) prediction that depend on the expertise of both communities include the use of global coupled models of the atmosphere, ocean, land and cryosphere together with the need to properly include the effects of external forcing, to generate an ensemble of forecasts and simulations, to suitably initialize the forecast model including the oceans, to store useful information and to post-process and analyze the results. These common aspects argue strongly for the melding of the expertise of the seasonal prediction and climate simulation communities and this was the impetus for the Workshop.

The areas of interest for the workshop included studies of:

- The "predictability" of the climate system (i.e. idealized studies of the possibility of predicting system evolution) on s2d timescales and the associated physical mechanisms
- Methods of initializing the coupled system and of generating forecast ensembles
- Aspects of coupled models that affect s2d predictive skill and the prospects for improvements
- Methods of bias adjustment, calibration and combination of forecasts
- Forecast quality that is useful and appropriate at s2d forecast ranges
- The application prediction information across the s2d timescales

Six sessions on these topics, including both oral and poster presentations, were held during the workshop and a seventh summary and discussion session followed on the last afternoon.

The first of the World Climate Research Programme's Grand Science Challenges (http://www.wcrp-climate.org/grandcha.shtml) includes the production and use of Regional Climate Information on seasonal to decadal timescales. The Grand Challenges in turn contribute to the Global Framework for Climate Services (http://www.gfcs-climate.org/), the WMO initiative directed toward enabling researchers, producers and users to improve the provision and use of climate information. The considerable interest in the scientific basis for skilful s2d prediction and in the development of sophisticated socio-economically beneficial applications of forecast results will continue to stimulate activity and progress in this important area of climate science. The International Workshop on Seasonal to Decadal Prediction has been a contribution to these endeavours.

Session S1: Sources of Predictability and Skill
(A. Scaife and G. Branstator)

A number of physical mechanisms for climate predictability and skill on seasonal to decadal timescales were identified and examined at the workshop. The main focus is on ocean variability displaying clear evidence of low frequency variability that provides predictability on seasonal timescales (e.g. the Indian Ocean Dipole), inter-annual timescales (e.g. the El Nino southern Oscillation) and multiannual timescales (e.g. the Atlantic Meridional Overturning Circulation or AMOC) in current prediction systems. The AMOC is receiving a lot of attention from the decadal prediction community and appears to have significant predictability to around 5 years or so ahead in current (CMIP5) hindcasts. This is associated with predictability and skill in the temperature of the sub-polar gyre influencing in some systems neighbouring downstream land areas such as northern Europe.

While the AMOC is believed to influence the sub-polar gyre surface temperature there is a parallel debate raging about the role of aerosols in forcing historical decadal variability of the sub-polar gyre. Volcanic experiments show clear influences on the AMOC and sub-polar gyre for a number of years after the eruption. On longer timescales some model simulations show a very strong link between decadal aerosol variability and sub-polar gyre temperatures that appear to correlate well with the observations of sea surface temperature. There are important consequences of this for the reproduction and interpretation of historical hurricane numbers over the 20th century. It remains to be seen whether the aerosol forcing in these models is too strong but correlations between observed and modelled hurricane numbers are high.

The stratosphere is a recurring theme for long range predictions as was highlighted in the 2007 WGSIP meeting and seasonal forecasting conference in Barcelona, along with the land surface and sea ice as potential sources of untapped predictability. The stratospheric quasi-biennial oscillation is now showing signs of multiyear predictive skill in hindcasts while ENSO links to the extratropics via the stratosphere and sudden stratospheric warmings are further key processes on interannual timescales. Both soil moisture and snow cover continue to show promise as areas that might lead to enhanced predictability and skill. Dynamical links between snow cover and winter surface predictability continue to be difficult to reproduce in climate models and there is some evidence that the large positive trend in October snow cover in some datasets may be an artefact of the observational analysis. Similarly, while there have been prominent media reports about possible links between sea ice and winter circulation, model experiments show relatively modest effects, albeit with a similar sign in general.

The workshop highlighted a number of outstanding issues including the fact that increased skill in decadal hindcasts does not necessarily imply that the source of predictability is due to internal variability. Rather, it could be that initialization corrects the forced response. Skill over oceans remains much higher than that over land and rainfall shows much lower skill than temperature. A potentially important result for future analysis is that while the correlation scores in the extratropics in some seasonal forecast systems reach very respectable values, the signal to noise ratio remains small, implying that the noise levels in some models are too high - indeed it turns out that many decadal predictions are under-confident in the CMIP database and so we should be wary of including further stochastic noise in climate prediction systems.

As a final point, the scientists at this workshop were unanimous in their view that understanding the mechanisms for variability and predictability and improving these in models is the difficult but crucially important task for the future. The CMIP database for decadal hindcasts along with the Climate-system Historical Forecast Project (CHFP) database for seasonal forecasts should lead to greater understanding of how we can improve the representation of sources of variability and predictability in current prediction systems.

Session S2: Assimilation, initialization, ensemble generation and bias
(M. Ishii and O. Alves)

This session had eleven oral presentations and more than 20 posters. The participants discussed a wide range of topics associated with climate model initialization for seasonal-to-decadal predictions.

Recent seasonal prediction systems successfully incorporate coupled ensemble data assimilation. By contrast, rather simple initialization strategies for decadal prediction have been adopted primarily because of computational cost and the ease of implementation. Seasonal prediction studies are leaning towards full-field assimilation, while it is not clear whether the anomaly or full-value data assimilation approach is most suitable for decadal prediction. In the EC-Earth climate model (Hazeleger) ensembles with full and anomaly initialisation show similar skill in multiyear global temperature predictions despite the drift having different amplitude, but some regional differences are seen. Full-field assimilation could be promising since this is natural or close to actual climate. Moreover, climate models initialized by full-field assimilation could perform more realistically than those with anomaly assimilation as the model parameters are fitted to unbiased climate. Another issue around the coupled assimilation strategy is fully coupled or weakly coupled.

Coupled ensemble assimilation (Lui) has the potential to yield physically consistent initial conditions of the atmosphere and ocean and flow-dependent ensemble members as initial conditions. The model fields are adjusted towards reality by observations fed through the model dynamics. In the Australian seasonal prediction system, a weak coupling strategy is used in which only ocean data are assimilated, and it is used for the reduction of shock caused by discrepancies between the ocean and atmosphere components (Alves). However, in this case coupled initialisation did not reduce shock due to coupled model error, e.g. spurious vertical velocities along the equator in the Pacific. The results from POAMA system (Fig. 2) show that weakly coupled assimilation had negligible impact on forecasts of El Nino. The impact of coupled initialization may not be fully realized until fully coupled assimilation is realized.

Perturbation generation is required for more reliable predictions of the long-term climate state, particularly in mid-latitudes. Various perturbation strategies have been proposed, but their effectiveness is limited in area and time scale. First of all, the ensembles must cover the range of uncertainty. A simplified singular vector approach for decadal prediction is used in the Max Planck Institute Earth System Model (Marini); it is one of the more sophisticated approaches used in short-term climate prediction. Incompleteness of model physics causes uncertainty in climate prediction as well. There are many approaches for realizing physics perturbation. In this session, we had a few presentations discussing directly perturbed physics ensembles. Meanwhile, climate prediction systems at individual centers have been developed by changing model resolution, model physics and initialization strategies. These studies will also shed some light on the realization of effective model-physics perturbations.

For seasonal-to-decadal predictions, the ocean initialization is critical as presented, for instance, for decadal prediction with the NCAR prediction system (Danabasoglu). However, we expect that there is large uncertainty in the ocean initialization. To overcome this problem, multiple strategies are utilized at several centers: anomaly or full-field, and 3DVAR or coupled ensemble or nudging, etc. As a result, there are large differences in re-analyses among centers and institutes. Huang took a multi-analysis ensemble approach, using a single prediction system (NCEP CFSv2), but using different ocean-reanalyses produced by the other centers. Another point for reduction of the uncertainty, particularly trends, was that we require assimilation strategies that overcome oceanic observation network changes.

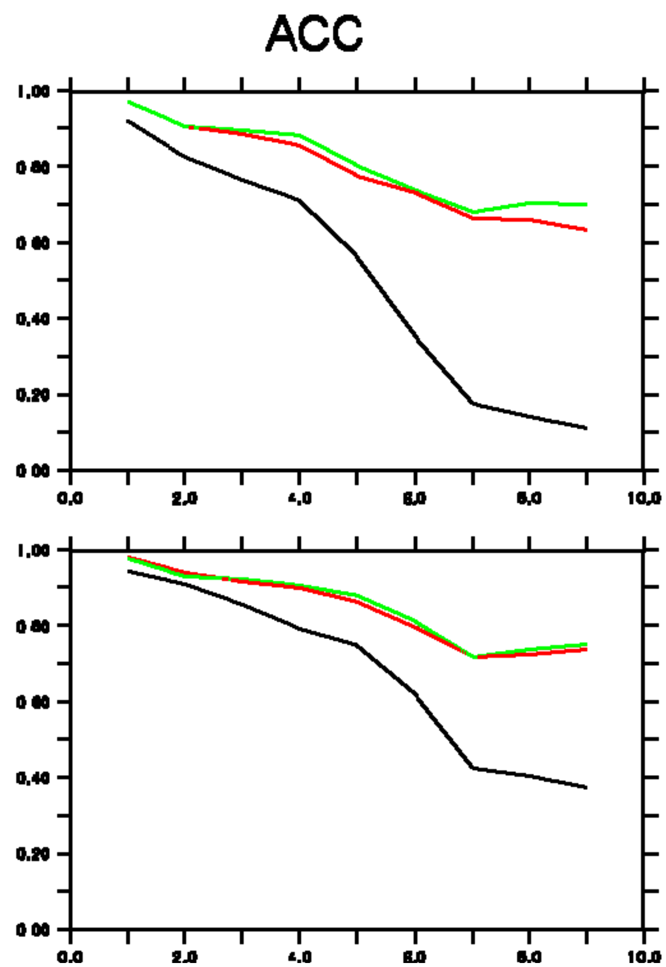

Figure 2: Correlation skill as a function of lead time (in months) for (top) NINO3 and (bottom) NINO4 for forecast from the POAMA seasonal prediction system. Red – with weakly coupled assimilation, Green – with assimilation separately into each component and Black – persistence.

The main sources of uncertainty in climate prediction are due to errors in the radiative forcing, sampling of natural variability, and the models. How to reduce and treat uncertainty in assimilation and prediction is key to making further progress. Prediction experiments have to be designed to understand the uncertainty in decadal prediction and decadal forecasts should be made with initial conditions every year (e.g. the extended CMIP5 protocol). A swap experiment by Corti is a rather new idea. This is an approach that could be used more widely to explore the importance of different components but needs more cases.

Sea ice and land surface are now included in model initialization because they have some signal for climate prediction. Regarding sea ice assimilation, Counillon uses an ensemble Kalman filter with sea ice initialization for seasonal-to-decadal prediction. In addition, many centers are starting to focus on sea-ice initialisation for seasonal-to-decadal prediction. There is a need to improve understanding of the polar climate system as well for achieving consistency with ocean and atmosphere observations. Land surface assimilation may provide additional seasonal-to-decadal skill and some groups are beginning to try to assimilate land surface variables.

Several studies on bias and drift in seasonal-to-decadal predictions were presented. Initialized climate models produce both signals carrying the initial information and climate drift. Teng studied a linkage between the magnitude of the drift and predictive skill by using the CCSM4 decadal prediction experiments. In decal prediction, there are several different sources of forecast bias, including bias due to model errors, insufficient sampling of internal variability, and a lead-time dependent bias caused by using incorrect forcing in the forecast (Hawkins). Gomez analyzed model drifts probably caused by the teleconnections of El Nino and Southern Oscillation. Understanding biases is a key to improving models, reducing model systematic errors, and understanding the quality of predictions. Also, bias correction of predicted variables is helpful for end users.

Overall, we are now taking different approaches for seasonal and decadal predictions. Future studies will hopefully realize initialization techniques that can seamlessly be used to predict climate states on seasonal-to-decadal time scales. At present, we have a lot of technical issues associated with coupled ensemble assimilation, perturbation generation, and model improvement. Counillon showed one example of the possibility of seamless prediction with a twin experiment using an ensemble Kalman filter and the Norwegian Earth System model. With future studies the performance of individual models will improve and as a consequence the performance of multi-model ensemble prediction will improve much more than the current status.

Session S3: Seasonal-interannual forecast systems and results
(M. Déqué and T. Stockdale)

This session was essentially devoted to seasonal to interannual prediction, with a particular stress on operational or quasi-operational forecast systems. The questions of more theoretical predictability and climate services to end-users were addressed in other sessions. There were 12 oral presentations and 12 posters in this session. The presentations (oral and posters) have shown results from 13 forecast systems (in general an atmosphere model coupled to an ocean model), and 2 multi-models. The Seasonal to Interannual Prediction (SIP) community has had meetings of this kind about twice per decade, most recently in Barcelona in 2007. From the results shown here, it is interesting to stress what has been improved in the models since the last workshop:

- Horizontal resolution has increased in the atmosphere (now better than 200 km) and in the ocean where some models are running with 0.25deg gridspacing
- Vertical resolution has increased, with many models fully covering the stratosphere (more than 50 vertical levels)
- The ensemble size is now of the order 20-30 in the hindcasts, instead of 9 in DEMETER
- There are attempts to initialize the soil moisture (summer) and the snowpack (winter)
- The sea-ice is often coupled to the ocean-atmosphere system

Given these upgrades, one observes significantly better scores in the hindcasts:

- Winter NAO can have time correlation up to and above 0.60 in months 2-4
- QBO index forecasts are significantly better than persistence (which is a hard challenge) as shown in Figure 1.
- ENSO scores are still improving (Nino 3.4 SST, and Nino3 SST)
- IOD is significantly predicted
- Sea-ice extent is predictable from May through September in the northern hemisphere. This is not just anomaly persistence.

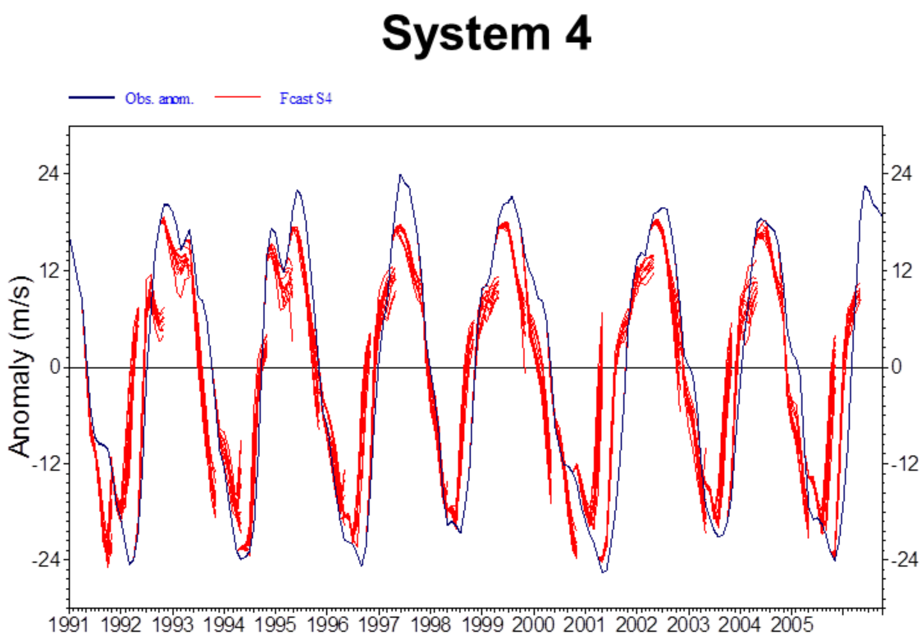

Figure 3: Zonal mean near-equatorial wind anomalies at 30hPa, predicted by ECMWF System 4 (in red, month 1-7) and as analysed (continuous dark blue). The QBO is forecast quite well by the system, although there is room for further improvement.

Seasonal-interannual forecasting (officially referred to as long range forecasting) is now well established as an operational activity by many national meteorological services, and aspects of the operational work are co-ordinated by WMO. Nonetheless, as evidenced by the presentations and posters in this session, there is still substantial experimentation and data analysis by the research community, and one challenge for the future is to make sure that there continues to be good interaction between the research and operational communities.

There are also emerging issues in the progress and evaluation of seasonal-to-interannual prediction:
(i) Signal to noise ratio. In general seasonal prediction models are still overconfident, and in a number of forecast systems stochastic components or other model perturbations are added to increase the spread in the model ensembles. This degrades the signal-to-noise ratio. However, there is also evidence from some forecasting centres that the intrinsic signal/noise ratio in the models might be too weak for the NH wintertime circulation, since the actual skill obtained by some models tends to be higher than would be expected if the real world had the same low level of predictability as the very noisy models. If the models poorly represent the predictability of the real world, it raises interesting questions as to why the models get it wrong, and what the actual true predictability is.

(ii) Choice of verification period and sampling issues. It was noted that scores in the mid-latitudes (e.g. NAO) can be sensitive to the verification period, and sampling considerations suggest that long periods are needed to get stable statistics. On the other hand, we know that for ENSO forecasts, the most recent hindcasts tend to be better, apparently because of improved ocean initial states due to the improvements in observing systems. There are many processes beyond ENSO that may add to seasonal predictability (e.g. land surface and stratosphere), and again, it is in recent periods where better quality data is available to initialize and test seasonal forecast systems. There is thus a dilemma between choosing long hindcast periods (with statistically more stable scores) and shorter hindcast periods (where behaviour and scores may be closer to real time forecasts).

(iii) Better understanding of the performance and characteristics of forecast systems. An important point is the need to understand the mechanisms of what is happening in the models. Equally, there is a need for more widespread use of careful statistical analysis, to properly characterize the uncertainties in skill scores and in comparisons between forecast systems. These two points might come together in the ultimate task of assessing not just whether our models include a certain mechanism or not, but how accurately they represent it.

In summary, seasonal forecast prediction systems are starting to mature nicely, and there is much heartening progress to see. However, it is clear that there are still many unanswered questions and much scope for improving the realism of models. There is also tantalizing evidence that the achievable skill in mid-latitudes may be higher than previously believed. This will need to be further tested and established, but there is no doubt these are interesting and exciting times for seasonal prediction.

Session S4: Decadal Forecast systems and results
(C. Cassou and F. Doblas-Reyes)

A dozen talks and about 10 posters dedicated to decadal forecast systems, their skill, their performance and their limitations, were presented in Session 4. The Toulouse workshop was the first of the traditional quadrennial meetings of the Seasonal to Interannual prediction (SIP) community in which the decadal timescale had a focus and dedicated time. This focus resonated with the CMIP5 agenda and the availability of all the simulations allowing for multi-model analyses and uncertainty assessment of the decadal forecasts. The main conclusions of the workshop are the following:

- There is high skill in the interannual to decadal predictions of temperature, except in some regions like the North Pacific basin. Precipitation shows regions where skill is high, although seldom statistically significant.
- Most of the predictive skill even at short leadtime (from 3-yr onwards) is attributable to external forcings (greenhouse gases –GHG, aerosols and volcanoes). This is especially true over the continents in general and over the Atlantic and Indian Oceans.
- Initializing the forecasts seems to better constrain the model sensitivity to GHG forcings so that temperature trends are better captured compared to observations.
- Added value from ocean initialization (beyond 3-yr or so) with respect to uninitialized historical experiments is found over the North Atlantic Ocean where forecast skill remains statistically significant (p value below 0.1) during the whole forecast time considered (10 years).
- Predictions of the frequency of North Atlantic tropical cyclones are highly skillful, skill that the initialization increases in a non-negligible way. Results from some systems suggest that aerosols play a major role in the multiannual variability of the tropical cyclone frequency.
- There is no temperature skill in the North Pacific beyond 3-yrs or so, not even from the external forcings. Most of the systems even exhibit significant negative correlation skill scores suggestive of major flaws either in the intrinsic representation of the decadal mode of Pacific variability in the models (Interdecadal Pacific Oscillation –IPV, interaction with El Nino Southern Oscillation –ENSO, etc.), in the prediction of the substantial multiannual variability, or in the ocean initialization techniques leading to a misspecification of the internal variability to be predicted. Initialized forecasts are often somehow degraded with respect to uninitialized runs or historical experiments.

Conclusions are conditional on intrinsic model performances in simulating zero-order climate entities and energy balances. Strong biases such as weakness of the Atlantic Multidecadal Overturning Circulation, erroneous sea-ice feedbacks with both the ocean and the atmosphere, cold conditions in the tropical Atlantic affecting the mean position of the Inter Tropical Convergence Zone (ITCZ) and its low-frequency variability, too deep ocean mixed layers, etc. can spoil the information provided by the initial conditions in the decadal forecasts. The lack of appropriate observations to generate robust initial conditions over a sufficiently long time and the lack of experience on how best to initialise decadal forecast systems likely hampered the skill obtained in the CMIP5 exercise. Special attention should also be paid to model responses to external forcing (volcanoes and aerosols) as well as to model sensitivity to GHG concentrations that can alter the added value of ocean initialization. The estimation of the latter is also strongly dependent on sampling. Several studies show that the CMIP5 CORE protocol based on the production of forecasts every 5 years from 1960 onwards may lead to non-robust estimates of skill and might not be adequate for a physical understanding of the sources of forecast skill, although such sampling frequency already offers a first-order estimate of the forecast quality. A clear recommendation emerges for conducting yearly initialized simulations instead of every 5 years despite the considerable demand of computational resources, which enhances the importance of clearly explaining not just the added complexity of decadal prediction over climate projections, but also their cost. The formulation of yearly forecasts also appears to be crucial for the robust estimation of the physical processes at the origin of model drifts (to be removed from raw model outputs) and biases in presence of episodic forcings such as volcanoes. Work on ocean initialization in the tropics also appears as a priority.

The above-mentioned conclusions seem to be quite robust, as they do not depend on the type of initialization adopted in the forecast systems. Outcomes from full-field versus anomaly initialization are generally statistically undistinguishable as are those from ocean initial conditions coming either from reanalyses or ocean hindcasts forced by atmospheric surface fields. Some regional changes are found between systems but are not consistent among models and cannot be interpreted as arising from physical mechanisms, the multi-model (or perturbed-parameter) approach providing the most robust estimates of skill. Tests have also been carried out to evaluate the impact of higher resolution in both ocean and atmospheric components of the models. The added value is not statistically significant at large-scale (typically ocean-basin wide) even if more analyses (regional focus etc.) are necessary to firmly confirm such a finding. The spread of the predictions can be excessive in cases, and for several systems the spread of the initialized predictions is larger than for the uninitialized ones, which is counterintuitive and points at the lack of informed design of the ensemble forecast systems. Careful development of the initialization strategy is clearly important to avoid wasting precious computing resources. All talks and posters clearly highlight the need to go beyond traditional statistics generally applied so far and the need to adopt a process-oriented approach to complete our understanding of the mechanisms of the decadal modes of variability as well as their predictability and their relationships with continental low-frequency fluctuations and extremes.

A special focus was devoted in Session 4 to the Atlantic because of higher skill and the dominance of decadal variability linked to the Atlantic Multidecadal Oscillation (AMO). The highest skill associated with ocean initialization is found in the subpolar gyre even if damped persistence, as a benchmark, is still difficult to beat. Current SST skill is still below what predictability estimates suggests, a gap that is particularly large over the Pacific basin. Results show that the AMO 1995 shift is reproduced in many systems when initialized and is associated with both the initialization of the dynamical component of the ocean through the AMOC and the thermohaline characteristics of the subpolar and subtropical gyres. The importance of salinity has been emphasized and preliminary results would suggest that recently observed anomalous salt content in the subpolar gyre could be a preconditioning sign for a swing of the AMO in the next few years. In the tropical Atlantic, a large portion of the skill comes from the external forcing even if added value from the initialization of the AMO is noticeable. Associated skills for the prediction of the low frequency hurricane activity in the tropical Atlantic basin have been presented even if the latter seem to be related, to a great extent, to the sole existence of the 1995 shift. The role of the subpolar gyre in controlling the hurricane activity has been hypothesized in predictability experiments but needs further analyses to extract the physical mechanisms at work in such a connection.

Some challenges that have been raised on the road to CMIP6 are listed below:

- How can we best exploit the existing CMIP5 extraordinary dataset available to the climate community? Probabilistic assessments and mechanism-oriented approaches to understanding multi-faceted forecast quality are clear targets.
- How to improve the initialization of the forecast systems, bringing in the best observational information for all the modeling components and exploring the best ensemble-generation approaches that lead to reliable predictions? How to initialize the systems in a more balanced way reducing the initial shock and, hence, preserving the initial-condition information for a longer time during the forecast?
- How to reduce the systematic error?
- An appropriate separation between the internal and external sources of skill is needed, but an experimental design to address this issue still does not exist.
- How to liaise with communities that can develop useful empirical prediction methods that can serve both as benchmarks and as additional forecast systems?
- What can we learn/exchange from/with the long-term climate-change community and from/with the extended-range climate prediction (sub-seasonal to seasonal) community as decadal forecast issues appear transverse to both? The model sensitivity constraint by the ocean, sea ice and atmospheric composition initialization and the origin of model drifts (which are at the heart of pervasive systematic errors) along the forecasts are two examples, respectively, for which bridges could be built.
- How to communicate and exchange information with end-users and outside communities (e.g. the climate adaptation and climate services communities)? How to combine the multiple sources of predictions into a single, reliable piece of climate information?
- What should be the main scientific objectives and experimental priorities that we want to address in the next CMIP exercise?
- What types of simulations could be suggested for CMIP6 that address the previous concerns? There was some consensus on the need for dedicated experiments to better understand the physical mechanism link to predictability.
- What will be the role for the real-time decadal forecast exchange, informally led so far by the UK Met Office, and how to operationalize it? How to best specify the natural and anthropogenic forcings in a hypothetical real-time distributed system? What are appropriate links to the GFCS in complement to CMIP6 that can offer information updated with a higher frequency than IPCC reports do?

Session S5: Forecast verification, calibration and combination
(C. Coelho and D. Smith)

This session highlighted a number of relevant aspects regarding seasonal to decadal (s2d) forecast verification (i.e. forecast quality assessment), calibration (i.e. forecast error correction) and combination (i.e. forecasts produced based on multiple forecast systems).

In terms of forecast verification the following relevant aspects for s2d climate prediction quality assessment were raised:
i) Assessing likely forecast skill is difficult. This is especially true on decadal timescales for which the assessment period typically only spans a few decades, and retrospective forecasts are made using far fewer observations for initialization than are available for real-time forecasts. Assessing the physical mechanisms is therefore important for gaining further confidence in forecasts and elucidating the sources of predictive skill: skill measures alone are not enough to fully describe forecast quality.

ii) Climate predictions, both seasonal and decadal, are uncertain. Therefore, skill assessment should not be limited to deterministic measures of the ensemble mean, but should also include probabilistic assessment of the full prediction distribution including skill measures able to summarize other attributes, such as reliability, resolution, etc. For a comprehensive forecast quality assessment several skill measures (both deterministic and probabilistic) are needed.

iii) The most common deterministic skill assessment measures currently used in s2d climate predictions are: mean-squared error (MSE), root-mean-squared-error (RMSE), mean-squared-skill-score (MSSS) and anomaly correlation. For imperfect systems, anomaly correlation is important because it measures potential skill that may be missed by other measures. However, manipulation of the forecast may be required to realise the skill implied by correlations.

iv) Reliability is an important forecast quality that should always be assessed, particularly because it is essential for decision making. Useful predictions must be reliable and skilful. For example, climatological predictions are reliable but not skilful. However predictions that are different from climatology but unreliable would provide misleading information for decision makers.

v) The availability of limited (short) hindcast length implies large sampling uncertainty in forecast quality measures. It is therefore important to try to quantify these uncertainties with confidence intervals (e.g. using approaches such as block bootstrap resampling). The construction of probability density functions (PDFs) for forecast quality measures, and also for differences in these measures when comparing different systems, is recommended.

vi) Beware of trends when assessing s2d prediction quality. High temporal correlation between predicted and observed temperatures does not necessarily indicate ability to predict year to year variability.

vii) For reliable forecasts, RMSE and ensemble spread should be equal. RMSE is typically greater than ensemble spread for seasonal forecasts, but at longer lead times the spread is larger than the RMSE, at least for one forecasting system. This kind of behaviour, if found in other models, suggests a need to better understand the evolution of model variability with timescale.

viii) Assessment of the fraction of variance that is predictable is recommended as this could be different in observations and model predictions. Preliminary results suggest that for some variables and regions the predictable fraction is greater in observations than models. This could arise, for example, if the model atmosphere responds too weakly to ocean forcing, and is consistent with other evidence suggesting this is the case. Each model ensemble member therefore should not necessarily be regarded as a potential realisation of reality, and forecast skill may be maximised by taking the mean of a large ensemble and adjusting its variance accordingly.

In terms of forecast calibration and combination the following relevant aspects for s2d climate prediction were raised during the session:
i) Because of uncertainties in initial conditions, model errors and shocks introduced by imperfect initialization, climate predictions deviate from the real world observed climate. This motivates the need for s2d climate prediction calibration by comparing past forecasts and past observations. Both model bias and drift assessment are therefore needed to quantify and reduce these errors and uncertainties.
ii) For dealing with model biases two approaches are currently being used: full field and anomaly initialisation. Both approaches work, though there is evidence that full field is better on seasonal timescales because teleconnections are more accurately simulated when model biases are small.

iii) Simple bias correction based on retrospective forecasts is a valid approach for climate model prediction calibration. However, assessment of likely forecast skill should take into account that bias correction is imperfect, for example by using uncentred, rather than centred, correlation (in addition to other measures of skill).

iv) It is possible to improve upon simple bias-correction, for example by calibrating spread and trends. However, care should be taken to avoid over-fitting to potentially noisy observations. Careful assessment, for example by detrending the data, is also needed to determine the added value of the forecasts relative to the calibration.

v) Multi-model prediction systems frequently outperform individual models. Further research is needed to investigate improved ways for combining climate predictions produced by the current s2d systems.

vi) The availability of empirical and dynamical model predictions motivates the need for developing and applying novel and robust methods for combining and calibrating all available predictions.

vii) Simple physically motivated empirical prediction systems constructed with past observations can serve as a reference to evaluate more complex physically-based dynamical prediction systems. Good quality empirical prediction systems when combined with state-of-the-art physically-based predictions system can lead to improved quality s2d predictions.

Session S6: Targeted predictions, downscaling and applications
(H. Douville and L. Dubus)

This session on "downstream" activities consisted of 9 talks and 14 posters. They emphasized the increasing demand for skilful long-range climate predictions for both optimization of current systems at sub-seasonal to inter-annual timescales and adaptation to climate change at decadal to multi-decadal timescales, the need for probabilistic forecasts based on multi-model ensembles, the need for downscaling techniques to provide useful predictions at the regional scale, the need for calibration techniques to remove systematic errors in global and regional predictions, the possible synergy between stochastic and dynamical tools, and the emerging hydrological applications of seasonal forecasting based on off-line global or regional hydrological models. Several presentations also emphasized the need for better-suited climate information (i.e. full statistical distribution of well calibrated and relevant climate variables), and for training the end-users to work with such products.

Not surprisingly, most presentations dealt with the seasonal or even sub-seasonal rather than decadal timescale. Seasonal forecasting shows substantial skill in the tropics (including a useful skill beyond a 6-month lead time as far as ENSO is concerned) as well as improved hindcast skill scores in the northern extratropics at least in winter. By contrast, multi-annual to decadal forecasting is a recent development with skill for temperature from initial conditions for the first few forecast years and skill from external forcing at longer forecast periods with comparatively modest skill over land.

Some presentations emphasized the difference between utility and skill and the fact that decadal or seasonal forecasting skill is often discussed on the basis of model outputs that are not necessarily relevant or even useful for decision making: global annual mean temperature (rather than regional and seasonal means), Atlantic Multidecadal Variability (rather than the North Atlantic temperature signal), tropical precipitation (rather than their land-only component), tropical cyclone density (rather than their landfalling-only component), etc… On the other hand, some systems can show little or no skill if direct model outputs are considered, but have positive and even valuable skills when raw model outputs are used to force application models. For instance, seasonal rainfall forecasts over Europe have almost no skill, but soil moisture and river discharge show usable results for water resource management and hydropower applications if snow and/or groundwaters are initialized correctly in an off-line hydrological system.

For most applications, downscaling the global model climate information at a finer scale is a necessary step that can be based on either statistical or dynamical (RCM) techniques several of which were discussed in the session. For example, seasonal averages are often not adequate to force application models which are based on higher frequency (daily, or sub-daily values), and which aggregate the information at the relevant temporal resolution afterwards. Preliminary attempts to downscale some decadal predictions with a RCM have been achieved within the MiKlip project

Two timescales are often distinguished by decision-makers: management (i.e. optimization of existing infrastructures) and adaptation (i.e. developing new infrastructures). Extending the range of seasonal forecasting is clearly useful while the development of methodologies for using annual and multi-annual temperature forecasts have only recently begun, based on the availability of decadal forecasts from CMIP5 for instance.

Finally, the difficulty of decision-making was highlighted and the limited use of climate information is not only due to its limited value or the limited training of potential end users, but also to other constraints or considerations such as the difficulty to anticipate the human component of any management or adaptation strategy.

Annex

Scientific Committee members

Oscar Alves	BOM	Melbourne, Australia
George Boer (Chair)	CCCma	Victoria, Canada
Caio Coelho	CPTEC	Cachoeira Paulista, Brazil
Francisco Doblas-Reyes	IC3	Barcelona, Spain
Masayoshi Ishii	MRI/JMA	Tsukuba, Japan
Ben Kirtman	RSMAS	Miami, United States
Arun Kumar	NOAA	College Park, United States
Jerry Meeh	NCAR	Boulder, United States
Adam Scaife	The Met Office	Exeter, United Kingdom
Doug Smith	The Met Office	Exeter, United Kingdom
Laurent Terray	CERFACS	Toulouse, France

Organizing committee members

Sylviane Balland	Météo-France	Toulouse	CIC secretary
Nathalie Brousset	CERFACS	Toulouse	CERFACS secretary
Philippe Caille	Météo-France	Toulouse	
Christophe Cassou	CERFACS/CNRS	Toulouse	
Michel Déqué	Météo-France	Toulouse	
Hervé Douville	Météo-France	Toulouse	
Isabelle Moity	CERFACS	Toulouse	CERFACS secretary
Laurent Terray (Chair)	CERFACS/CNRS	Toulouse	
Isabelle Varin	Météo-France	Toulouse	CIC secretary