WCRP REPORT
World Climate Research Programme

WMO IOC ICSU
International Council for Science

Position Paper

WCRP Position Paper on Seasonal Prediction

Report from the
First WCRP Seasonal Prediction Workshop
(Barcelona, Spain, 4-7 June 2007)

February 2008

WCRP Informal Report No. 3/2008
ICPO Publication No. 127

WCRP Position Paper on Seasonal Prediction
Report from the First WCRP Seasonal Prediction Workshop, 4-7 June 2007, Barcelona, Spain

Lead Authors:
Ben Kirtman, University of Miami Rosenstiel School of Marine and Atmospheric Science Division of Meteorology and Physical Oceanography and Center for Ocean-Land-Atmosphere Studies (COLA);
Anna Pirani (International CLIVAR Project Office)

This paper summarises the motivation and outcomes of the First WCRP Seasonal Prediction Workshop, which was held June 4-7 in Barcelona Spain, bringing together climate researchers, forecast providers and application experts. To see a full list of talks, posters and to access an electronic version of this paper, please refer to the Workshop website: http://www.clivar.org/organization/wgsip/spw/spw_main.php. The main purposes were to describe the current status and main limitations regarding seasonal forecast skill and applications, and to make recommendations to improve both of these aspects. It is clear that there is substantial scope for improving skill by reducing model biases and including a wider range of climate processes, and improving benefits through better communication of more appropriate information.

Introduction

Our ability to predict the seasonal variations of the Earth's tropical climate dramatically improved from the early 1980s to the late 1990s. This period was bracketed by two of the largest El Niño events on record: the 1982-83 event, whose existence was unrecognized until many months after its onset; and the 1997-98 event, which was well monitored from the earliest stages, and predicted to a moderate degree by a number of models several months in advance. This improvement was due to the convergence of many factors including a concerted international effort to observe, understand and predict tropical climate variability, the application of theoretical understanding of coupled ocean-atmosphere dynamics, and the development and application of models that simulate the observed variability. The international Tropical Ocean Global Atmosphere (TOGA) program successfully demonstrated the potential predictability and societal benefit of seasonal prediction.

After the late 1990s, our ability to predict tropical climate fluctuations reached a plateau with little subsequent improvement in quality. Was this a result of a fundamental change in the predictability of the climate system due to either natural or anthropogenic forcing, or the emergence of a critical failing in the models used to make predictions or merely a sampling effect? Have we accounted for all the critical interactions among all the elements of the climate system (ocean-atmosphere-biosphere-cryosphere)? Are the observations adequately blended with the models to make the best possible forecasts?

About a third of the world's population lives in countries influenced significantly by climate anomalies. Many of these countries are developing countries whose economies are largely dependent upon their agricultural and fishery sectors. The climate forecast successes of the 1980s and 1990s brought great promise for societal benefit in the use and application of seasonal forecast information. However this promise of societal benefit has not been fully realized, in part, because there have not been adequate interactions between the physical scientists involved in seasonal prediction research and production, applications scientists, decision makers and operational seasonal prediction providers. The issues and problems go beyond merely improving forecast quality and making forecasts readily available. The physical scientists need to actively facilitate and understand users' requirements, in order to provide improved climate information, prediction products and services leading to enhanced applications. Users also have to maintain an active dialogue with the physical scientists and forecast providers so that their climate information needs are taken into account.

One of the overarching objectives of the World Climate Research Programme (WCRP) is to facilitate analysis and prediction of Earth system variability and change for use in an increasing range of practical applications of direct relevance, benefit and value to society. In order to, in part, meet this objective the WCRP commissioned in 2005 the Task Force on Seasonal Prediction (TFSP) to assess over a two year period current seasonal prediction capability and skill considering a wide range of practical applications, and to enable the development and implementation of numerical experimentation specifically designed to enhance seasonal prediction skill and the use of seasonal forecast products for societal benefit. The TFSP mandate is now being continued by the CLIVAR Working Group on Seasonal to Interannal Prediction (WGSIP). This is in addition to the ongoing activities within the World Meteorological Organization (WMO), where seasonal to interannual prediction forms the core component of the activities of the Climate Information and Prediction Services (CLIPS) project of the World Climate Applications and Services Programme (WCASP)/World Climate Programme (WCP), which seeks to develop user targeted climate services within the Member countries of the WMO. The WMO Commission for Climatology (CCl), under the Open Programme Area Group on CLIPS, guides this process through its expert teams on research needs, operations, verification and user liaison. WMO has recently established Global Producing Centres (GPCs) of long-range forecasts, providing the National Meteorological and Hydrological Services (NMHSs) worldwide greater access to real-time seasonal forecasts,

As part of the seasonal prediction capability assessment, the TFSP in collaboration with the core projects of the WCRP (CLIVAR, CliC, SPARC and GEWEX) and the WCP (through its WCASP/CLIPS project) organized the First WCRP Seasonal Prediction Workshop, which was held June 4-7 2007 in Barcelona Spain. This report summarizes the key outcomes and recommendations of this workshop. This report is intended to go beyond merely summarizing the workshop presentations; indeed we specifically avoid this sort of summary. The main purpose here is to provide definitive statements regarding current skill in seasonal prediction with emphasis on surface temperature and rainfall and how the forecasts are currently being used for societal benefit. In addition, the report outlines a set of specific recommendations for improving seasonal prediction skill and enhancing use of seasonal prediction information for applications.

The Workshop focused on addressing two basic overarching questions:

(i) What factors are limiting our ability to improve seasonal predictions for societal benefit?

(ii) What factors are limiting the application of our seasonal predictions for societal benefit?

In addition to addressing these questions, the workshop participants developed recommendations spanning both the physical and application sciences for how to overcome these limiting factors. The workshop participants also developed a roadmap for improving skill and setting priorities on the development and application of dynamical models for seasonal prediction recognizing that this process necessarily requires robust interactions between the physical science and applications communities and a delicate balance between scientific feasibility and application requirements. As described below, the workshop participants have also proposed a process by which progress in seasonal prediction can be regularly and comprehensively assessed.

Workshop Report

The workshop brought together the diverse seasonal prediction community. This included researchers of the physical climate system and forecast methodology, operational forecast providers and forecast application experts. There were approximately 180 attendees that represented diverse international interests both in the physical and application fields. Approximately 30 countries or so from the WMO Regions I-IV (Africa, Asia, South America, North and Central America, Southwest Pacific and Europe) were represented. Representatives from all the major operational seasonal prediction centers and the US funding agencies were in attendance.

This report is organized as follows:

(1) We present some critical common language regarding the assessment of seasonal prediction, in particular with regards to the distinction between measures of forecast quality (or skill) and measures of forecast value to the users. Developing a common language for assessing seasonal prediction is critical to successful interaction among forecast providers, forecast users and forecast researchers. A common set of diagnostics for quality assessment are now adhered to by WMO Global Producing Centres (GPCs), following the development of the Standard Verification System for Long-range forecasts (SVSLRF). Much work remains to be done to develop standard methods to assess value.

(2) We enumerate the overarching consensus among the workshop participants regarding the current status and future prospects of seasonal prediction. These consensus statements required considerable discussion among the workshop participants and invited experts, both during and after the Workshop, and carry the full weight of the seasonal prediction community.

(3) For reasons of practical implementation the SVSLRF defines a ‘core’ set of just three diagnostics for assessing quality, The workshop examined other metrics for assessing seasonal prediction quality, which may help provide a key benchmark of evaluating future improvements. Results from one non-core SVSLRF metric, the Brier Skill Score (BSS), are presented here; however, the workshop participants recognized the importance of allowing these metrics to be refined over time. In fact, it is this specific element of this report that makes it a “living” document that will be refined and updated in the future and made available online to the entire community.

(4) While a comprehensive assessment of forecast applications is not possible at this time, we do present some results from the workshop participants. It is the view of the workshop participants that more effort in terms of forecast applications is required. For example, future seasonal prediction projects, like the Climate-system Historical Forecast Project (CHFP, see Appendix A), need to include an applications element, and that communication between the applications and forecast research/provider communities needs to be improved through more events like the Barcelona workshop.

(5) The workshop participants identified some “best practices” in terms of producing, using and assessing seasonal forecasts. These best practices are also presented in this report.

(6) The workshop participants suggested specific areas where additional research and experimentation is required. In particular, the workshop

participants refined the 'total climate system' seasonal forecast experiment developed by the WCRP Task Force on Seasonal Prediction (TFSP). These refinements are described here and the CHFP experiment is outlined in Appendix A.

1.0 A Common Language for Assessing Seasonal Prediction

The need for an authoritative statement on the skill of seasonal, and other extended-range, predictions, both from research and operational perspectives, has long been recognized. Several audiences for such a statement exist, including forecast researchers, forecast producers and distributors, managers, funders, and the wide range of individuals outside the climatological community who either process the forecast information to advise others or who take decisions based on that information and who will be referred to collectively below as 'users'. Each audience has its own specific requirements for such a statement, and the information processing necessary to produce the statement varies accordingly. In order to simplify this current statement the target audience is assumed to be forecast researchers and users, although others will gain benefit.

The term 'skill' covers a complex array of issues; in this statement just two will be covered, *quality* and *value*.

(i) *Quality* refers to the technical measurement of forecast performance; *quality* is of prime concern to scientists and is often queried by users.

(ii) *Value* relates to the practical benefits achieved through decision making based on forecast information, usually in conjunction with other information, and while of fundamental concern to the user should also stimulate scientists.

The adherence of 9 WMO GPCs to the exchange of forecast (quality) verification scores defined by the SVSLRF – and made available on the SVSLRF Lead Centre website (http://www.bom.gov.au/wmo/lrfvs/) makes possible, for the first time, an objectively based authoritative statement on the current status of seasonal prediction quality. Such a statement might be based on the proportion of forecast systems with skill scores above a threshold for a range of regions. However, the production and communication of an authoritative statement of forecast quality remains hindered by a number of factors, including:

- The limited evidence base – unlike daily forecasts seasonal predictions are available only a few times a year, creating difficulties in accumulating the numbers needed to provide stable estimates of *quality* (together with the expense of producing the predictions when ensembles are used)
- Predictability varies between years, with strong evidence linking these variations at least to changes in sea surface temperatures across the tropical Pacific Ocean, thus increasing the sample size needed for stable *quality* estimates
- Frequent updating of models used for ensembles, when compared with the frequency of predictions, restricts sample sizes for *quality* estimates
- Many centres with advanced prediction systems that are engaged in prediction on a research rather than operational basis, do not participate in the exchange of scores defined by the SVSLRF. These and other centres select other metrics based on own needs and experience.
- Most metrics of forecast quality, including those of the SVSLRF, are technical in nature and not easily communicated to audiences outside the seasonal forecasting community. Metrics that allow interpretation of forecast quality across all audiences (particularly in the case of probabilistic forecasts) have yet to be developed.

The above point regarding the difficulty of interpreting quality assessments of probabilistic forecasts is a key barrier to developing an authoritative statement on forecast quality that is understandable to a wide audience. Scientific evidence indicates unequivocally that predictions should be provided only as probabilities, using either ensembles with dynamical models or appropriate approaches with empirical models. Many users, and perhaps some scientists, prefer deterministic approaches, and despite the contra-evidence these are still provided and assessed at some centers. For development scientists deterministic *quality* metrics provide certain information, but for users they are of limited, if any, worth. Metrics associated with probability predictions, and their interpretations, tend to be more complex than those for their deterministic cousins. Certain sectors (insurance, utilities) are nonetheless adept in dealing with probabilistic information.

The need for an authoritative statement of *quality* that covers all aspects of seasonal predictions, including all types of models, is incontestable, and that need is urgent not least given the steadily widening reach and use of the forecasts. While the database of real forecasts continues to grow, there are continuing difficulties in accumulating enough forecasts for quality assessments. The use of hindcasts (or retrospective forecasts) is the typical approach for increasing the sample size. However, hindcasts also have some difficult issues that need to be tackled (e.g., lack of initialization data, non-stationary nature of observing systems, non-stationary nature of the climate system). In order to achieve that statement one or more metrics, each with their own interpretation basis, will need to be agreed through international consensus. Selection of the metric(s) should take into consideration the various audiences to whom the statement will be directed. An international project is required in order to develop the statement with the necessary background.

Despite the above, several broad qualitative statements on *quality* can be made at present:

- Seasonal predictions can be produced in some regions that are more skilful on average than chance, the use of climatology, or persisting recent seasonal anomalies forwards
- Numerical models in general offer similar average prediction *quality* to empirical models in some regions
- *Quality* varies on an inter-annual basis, partly linked to interannual variability of the Pacific Ocean; average *quality* also differs between specific seasons (as illustrated in Table 1)
- Most seasonal forecasts make use of sea surface temperatures. Many require forecasts of sea surface temperatures during the target season; these can be produced typically most successfully over the tropical Pacific Ocean, less so over other tropical basins, and with least general success, sometimes no success, outside tropical waters
- Of the two major climate variables of interest to users, predictions of temperature tend to be of a higher *quality* than those of rainfall
- As a general rule of thumb, but with exceptions, *quality* declines with increasing distance from the Pacific Ocean and from the tropics; at some locations predictions are viable only during ‘windows of opportunity’

Value, the benefit gain through the use of prediction information in decision making, is distinctly more complex to measure, and hence to develop into an authoritative statement, than *quality*. Real-world *value* depends on many factors, the *quality* of the prediction information being but one. Other factors include the manner in which the prediction information is presented and distributed, and the approaches taken to real-world decision making in which climate is typically just a single consideration. A key parameter for estimating the value of a forecast for a given application is the cost function of that application, i.e., the cost of a wrong forecast and the benefit of an accurate forecast. We have demonstrated that in some regions and during some seasons seasonal predictions have quality, but their translation into value for end-users is far from optimal and a concerted effort is required to engage customers and seek their quantitative definition of value, so that the forecasts can be shown to improve decision-making.

The direct link between seasonal forecast quality and value have not been established, therefore, appropriate processes need to be engaged to measure *value* in specific decision making instances independently from the assessment of *quality*. The Relative Operating Characteristic metric (one of the core SVSLRF scores) does link quality and value in a linear way and provides useful information on value in limited circumstances but in general over-simplifies the complexities of real-world decision making. It must be recognized that effective communication of true *quality* does have a bearing on the *value* realized.

No authoritative statement regarding *value* is currently possible, either within specific contexts or generically, in which sufficient cases are available to provide a stable estimate. There are, however, a number of case studies demonstrating beneficial value using individual predictions (e.g., those emanating from forecasts during the 1997/98 El Niño event). It should also be noted that that better knowledge of climate variability (e.g. the chances of various scenarios, regardless of forecasts) could aid applications/planning/management. The application of forecasts requires trust in the quality of the forecasts and knowledge of forecast uncertainty.

2.0 Overarching Consensus Statements

(1) The workshop participants unanimously agreed that the maximum predictability of the climate system has yet to be achieved in operational seasonal forecasting.
   - This position is based on the recognition that: (i) model error continues to limit forecast quality and that (ii) the interactions among the elements of the climate system are not fully taken into account and may lead to improved forecast skill. The fact that model error continues to be problematic is evident from the need for and success of calibration efforts and the use of empirical techniques to improve dynamical model forecasts. Land-atmosphere interactions are, perhaps, the most obvious example of the need to improve the representation of climate system interactions and their potential to improve forecast quality. Essentially there is untapped predictability due to the fact that we currently do not account for all the interactions among the physical elements of the climate system. The maximum achievable predictability is unknown and assessing this limit requires much additional research.

(2) Multi-model methodologies are a useful and practical approach for quantifying forecast uncertainty due to model formulation.
   - There are open questions related to the multi-model approach. For example, the approach is ad-hoc in the sense that choice of models has not been optimized. Nor has the community converged on a best strategy for combining the models. Multi-model calibration activities continue to yield positive results, but much work needs to be done. These issues as well as others require additional research. It is also important to note that the

multi-model approach should not be used to obviate the need to improve models.

(3) A common agreed upon baseline for assessing seasonal prediction skill, as embodied by the SVSLRF, is critical for documenting future improvement. It is recommended that adherence to the SVSLRF is promoted, and that the set of 'core' metrics is frequently reviewed by the wider seasonal forecasting community. This includes best practices in forecasting and appropriate validation/verification techniques; and recognition of the non-stationarity of the climate system. These best practices need to be developed for both global and regional prediction systems.
  - This report and its future evolution in collaboration with WCRP and WCP is an effective process for developing and refining these best practices. The workshop participants argued that there is an immediate need for the international seasonal prediction community to come to a consensus on best practices. Some of these issues are touched upon in this report, but more work is needed.
  - Recognizing the non-stationarity of climate variability is important in terms of assessing quality and enhancing value of season forecasts. As such seasonal forecasts, particularly retrospective forecasts, should be made with observed climate forcing as noted in the Climate-system Historical Forecast Project (CHFP) experimental design proposed by the TFSP (see Appendix 1). Commonality of physical processes and of models is an explicit link across predictive time scales (e.g., seasonal to climate change). This makes seasonal forecasting a vital test bed to assess the reliability of longer-range predictions. This is particularly true when applications, such as those in agriculture and health are considered.

(4) Model errors, particularly in the tropics, continue to hamper seasonal prediction skill.
  - The importance of reducing model error cannot be over stated. There are a number of strategies for improving models including a better representation of the interactions among the elements of the climate system, inclusion of biogeochemical cycles, and substantial increases in spatial resolution. All of these strategies need to be vigorously pursued; better international coordination and commitment would be highly beneficial.

(5) Forecast initialization is an area that requires active research.
  - Ocean data assimilation has improved forecast quality; however, coupled data assimilation is an area of active research that is in need of enhanced support and perhaps international coordination. There is significant evidence that coupled ocean-atmosphere data assimilation is likely to improve forecast quality. Compatible land surface initialization strategies are actively being pursued in GEWEX and continued coordination with the seasonal prediction community is warranted.

(6) Observational requirements for seasonal prediction and the development of applications of seasonal predictions are not being adequately met.
  - While defining the observational requirements for seasonal prediction was beyond the scope of this workshop, the participants agreed that this is an issue that requires attention.

(7) Verification should also be undertaken routinely using simplified but multivariate driven dynamical application models. These models should be

complex enough to capture non-linear interactions, while being simple enough to avoid over tuning through non-constrained parameters.

- The relationship between forecast quality in applications models and meteorological models is often highly non-linear. Quality in the prediction of seasonal mean rainfall may not translate into quality in the prediction of crop yield, for example. Thus application models can provide additional metrics of forecast quality. Furthermore, these metrics usually have a specific user group in mind.
- There is a clear need to provide information at local scales. For many applications, forecast information is required at local space (particular locations that may be sensitive to e.g. terrain) and time (e.g. monsoon onset) scales. Further work is required to provide and improve such information, e.g. through statistical and/or dynamical downscaling. Work is also required to improve the consistency and continuity of medium- and long-range information (‘seamless prediction’). It is, however, important to note that dynamical downscaling does not on the whole improve an already lacking forecast at the global model resolution.

(8) Web based tools need to be developed to allow users of the prediction information to tailor the underlying climate information more easily to their needs (e.g. climate range/thresholds, spatial scale(s)).

- Progress on this front is critical to improving the value of seasonal forecast.

(9) Although there are many examples of seasonal forecast application (e.g., health, agriculture, water management), there is potential to do much more.

- More progress needs to be made in bringing seasonal prediction providers and seasonal prediction users together. More work is required to develop the production and understanding of probabilistic forecasts. Understanding of what is predictable and what is not predictable need to be enhanced. The importance of predicting 'extremes' (even top and bottom quintile categories are extreme for seasonal prediction, and probability forecasts are increasingly presented in these terms) was also noted. Communication of forecasts and warnings continues to be a problem even as our forecasting capability and infrastructure improves. Stronger links between operational forecasting centers and the WMO/CLIPS network need to be encouraged.

(10) Seasonal predictability research needs to be encouraged.

- While it is not possible to say unambiguously what is and what is not predictable in general, it is possible to provide specific examples of predictability beyond weather; however, these examples have not been adequately quantified. Indeed, this document, as it evolves over time should include, where possible, unambiguous statements regarding what is predictable and what is not predictable. Focusing of assessing the state-of-the-art in prediction quality is one approach for assessing predictability. Collaborations and interactions with the climate change community needs to be encouraged and has the potential for significant benefits. Indeed, seasonal prediction must be addressed in the context of a changing climate.

3.0 Assessing Seasonal Prediction Quality

The workshop participants agreed that the feasibility of seasonal prediction rests on the existence of slow, and predictable, variations in the Earth's boundary conditions. Within the paradigm of atmospheric predictability due to external forcing, the potential for skillful forecasts depends on a ratio of the externally forced signal relative to the atmospheric generated internal noise. The majority of external variance is known to originate from sea surface temperature variations, and less is known about the seasonal signals due to other external forcings such as soil moisture, land use, sea ice, atmospheric chemical composition and aerosols. Additional skill due to atmospheric initial conditions is also expected for certain slow modes of the atmosphere (for instance, annular modes), but there is little evidence that atmospheric initial conditions contribute to skill for lead time forecasts beyond a few weeks.

The workshop participants made presentations on validating and assessing the state-of-the-art and quality in seasonal forecasts by bringing together retrospective forecast data issued from international research projects (e.g., SMIP2/HFP, DEMETER, ENSEMBLES, and APCC) as well as data available from operational centers. Assessments were made in terms of scientific quality and factors limiting improvement. The presentations highlighted issues important for interfacing seasonal forecasts with applications including calibration, downscaling and validation, and determining whether there is an emerging consensus on approach and methodology. The workshop participants addressed seasonal prediction from a wide-ranging multi-disciplinary perspective looking at the role of cryospheric processes, stratospheric processes and air-land interactions on seasonal prediction, as well as the role of ocean initialization, aiming to explore additional sources of potential seasonal predictability. A number of the presentations emphasized the quality of seasonal prediction in the monsoon regions of Africa, Asia and South America.

Based on these presentations the workshop participants converged on a metric (the Brier Skill Score) that might be used in addition to the core SVSLRF to help develop an overarching assessment of the quality of seasonal prediction. It was clearly acknowledged that these metrics are not necessarily sufficient for assessing the value of applications, which require more detailed information beyond the scope of this report. Here we provide examples – they are not intended to be comprehensive. These metrics; however, do provide a simple benchmark from which progress can be measured. It was also acknowledged that future refinements and enhancement may be required and the workshop participants urged that this assessment be viewed as an evolving or "living" document that will be periodically updated and reviewed. Some additional detail of current seasonal forecast quality based on presentations from the workshop participants is provided in Appendix B.

3.1 Multi-model Bier Skill Score (BSS)

This first metric is a multi-model Brier Skill Score (BSS) for seasonal mean (DJF and JJA) 2m temperature and rainfall over 21 standard land regions (Giorgi and Francisco, 2000, *Clim. Dyn.*, **16**, 169-182). These regions, seasons and lead times are not necessarily the optimum for all users and forecast providers. Nevertheless, they do provide a reasonable overall measure of state-of-the-art quality. The one month lead seasonal mean multi-model (based on DEMETER data; see Palmer et al. 2004, *BAMS*, **85**, 853-872) BSS is summarized in Table 1. The BSS is calculated over the period 1980-2000. Here the BSS is calculated for binary events (i.e., precipitation exceeds the upper tercile, $E^{+}_{P}(x)$; precipitation exceeds the lower tercile, $E^{-}_{P}(x)$ and similarly for temperature: $E^{+}_{T}(x)$, $E^{-}_{T}(x)$). Positive values indicate forecasts with better Brier Scores than climatological 'forecasts'. Underlined values indicate greater than 90% confidence in the BSS. Negative underlined values indicate that the multi-model ensemble fails to predict the occurrence of the event. Whether the negative underlined values provide useful information is the subject of debate and research.

Overall it is clear that 2m temperature is more reliably predicted than precipitation regardless of season. Tropical regions generally show more temperature reliability (e.g., Central America, Amazon Basin, Western Africa), although there are sub-tropical regions of considerable forecast quality (e.g., Tibet). While some regions can be reliably predicted in both JJA and DJF, there is significant seasonality in 2m temperature forecast quality. We note that some of the forecast quality in the 2m temperature is due to the warming trend over the verification period 1980-2000.

In contrast to 2m temperature, the models have significant difficulty capturing the rainfall variability over these land regions. There is notable forecast reliability in the local summer seasons over the Amazon Basin and Southeast Asia. Elsewhere the precipitation forecast reliability is desultory.

In defining this metric, the workshop participants identified two points that highlight the importance of improving models: (i) calibration can improve the reliability and (ii) exploiting known dynamical and physical relationships (i.e., teleconnections) can also be used to improve the quality. The fact that forecast quality can be improved using these techniques indicates that models and predictions can and should be improved.

It is important to note that all skill scores have disadvantages. For example, one disadvantage of the BSS relates to the use of climatology as the reference ‘forecast’. Reliability is a component of the Brier Score. Since climatological forecasts have perfect (but useless, in a forecast sense) reliability, the BSS is prone to providing rather pessimistic measures of quality.

3.2 Sea Surface Temperature

Numerical prediction systems are used to predict the future evolution of sea surface temperature variations associated with El Nino. The available evidence suggests that the accuracy and reliability of real-time forecasts from a general circulation model (GCM) are just as good as the accuracy of hindcasts with the same model - there is no sign of "artificial skill" in the testing period. (There are hints in some cases that more recent forecasts are better than is possible for earlier dates, presumably because of improvements in the observing system).

Comparing the skill of different forecasting systems on common sets of forecasts shows that there has been slow but steady progress over the last 10 years. Improvement in ENSO SST forecasts is expected to continue in the years ahead - the errors in today's forecasting systems are still substantially above what we believe to be the predictability limit.

Consideration of a set of individual forecasts shows that today's models give mostly moderately good guidance as to the future evolution of SST, but failures can still occur. Even though some of the failures in the past might be related to an inadequately observed initial state, we are not yet at the stage that today's model forecasts are foolproof. The use of multi-model ensembles can give a definite boost to the skill compared to that obtained by a single model, and multi-model approaches to ENSO prediction should be strongly encouraged. Nonetheless, today's multi-model ensembles are still quite some way from the predictability limit, and improvement of the individual models is strongly needed to improve the quality of future forecasts (single or multi-model).

Typically, models are used to produce ensemble forecasts in order to quantify uncertainty and estimate higher moments. Attempts to verify higher moments of the distribution suggest little skill. Forecast spread does vary according to season and ENSO phase in the models, but at the moment forecast errors are still dominated by model error rather than predictability error, and the relationship of forecast error to model spread is weak. For real applications, any model forecast must be post-processed in some way. Probabilistic verification of dressed forecasts is to be encouraged, but at the moment the information content of the forecasts is thought to be very largely dominated by the first moment, i.e. the ensemble mean.

The Mean Square Skill Score (MSSS) against climatology is a good way of assessing the deterministic skill of ENSO forecasts. In particular, it gives a measure of forecast error (variance) scaled by the signal (variance), which is forcing the atmosphere (i.e. it shows us the error relative to the signal strength). State-of-the-art MSSS for Nino 3.4 SST at 5 months lead-time is about 0.7 for the best single models, and 0.75 for multi-model combinations. See Figure 1 which shows the MSSS for 1981-2001.

3.3 Climate Indices

In addition to surface air temperature, precipitation and SST, well-known climate indices representing major climatic features such as North Atlantic Oscillation (NAO), ENSO, etc. are also predictable, which can be used to estimate the associated regional impacts.

4.0 Assessing Seasonal Prediction Value

As noted in the introduction, the workshop participants did not attempt to provide an overarching assessment of the value of seasonal predictions. Despite this there were several presentations describing the application and value of seasonal forecasts. Some examples of how seasonal forecasts have been used and lessons learned from these activities are briefly summarized here. In fact, some of the best practices noted Section 5.0 are based on the lessons learned noted here.

Seasonal Forecast Use

- During 2006-07, agriculture and disaster management sectors gained considerable benefit from the use of seasonal forecast information in Venezuela.
- Météo-France uses dynamical long range forecast information for the Senagal Manatali dam via a water management model.
- The UK Met Office for several years has provided seasonal forecasts specifically for the Volta River water management project.
- The International Research Center of El Nino, Ecuador (CIFFEN) is a regional organization that links climate and long range forecast providers and users. Widespread use of such information by government departments has been developed – a bulletin is sent to around eight thousand influential users.

Lessons Learned

- Decision makers should not omit seasonal forecast information, however we must be aware that our products are not the only factor they consider in the decision process.
- Successfully communicating uncertainty and the limitations of seasonal forecasts is critical to the process of making seasonal forecasts valuable.

- It is very difficult for many users to make explicit use of climate forecasts, especially when they are offered only maps rather than data. The categorical formats issued by most forecasting centers are not consistent with the needs of many decision makers. In some occasions our best forecast may simply be climatology – this can be useful information. Availability of and access to hindcast data is also essential to assist users in assessing the model performance and the potential benefit of the forecasts.
- Often, if there is some quantitative information, users will try to transform that (usually through an assumption of an underlying analytical distribution) into terms that more closely meet their needs. Similar statements can be made regarding the spatial scale of forecast information – the resolution from global models is considered too coarse for many decision makers.
- Seriously misleading situations can occur when users take information appropriate at the large scale and apply it to local scales without considering the additional uncertainty associated with such action.
- Applications models can be used with seasonal forecasts to produce a metric that combines quality and value. For example, the quality of a seasonal forecast of crop yield is both a formal measure of skill and measure of potential value.

5.0 Best Practices in Seasonal Forecasting

The following is a list of best practices in terms of producing, using and assessing seasonal forecasts:

- Address forecast error by appropriately quantifying dynamical model uncertainty (using either several models, stochastic physics approaches or perturbed parametertizations of a single model). A useful resource in this respect is output from the WMO GPCs, which is being made available in a standard format to NMSs via a GPC lead website. Accompanying verification information for the GPC systems is available on the SVSLRF lead website (http://www.bom.gov.au/wmo/lrfvs/)
- Recalibrate model output based on historical model performance;
- Issue probabilistic forecast information;
- Provide description of forecast process (including post-processing methodologies);
- In retrospective forecast mode no information about the future should be used. The procedure should mimic real-time forecasting in this respect.
- Provide forecast quality information including several metrics of quality
    - The WMO-SVSLRF provides a set of procedures and algorithms for verification (http://www.bom.gov.au/wmo/lrfvs/scores.shtml)
    - Tailor some of these metrics for specific requirements of application users – e.g., in monsoon systems late or early onset of rains, frequency of break cycle lengths.
    - The WMO Climate Watches make recommendations on issuing, monitoring, and verifying forecasts for particular events. These should be used.
- Regional climate service providers need to work with both the forecasting and application communities to develop tailored downscaled products and a range of downscaling approaches including dynamical, statistical and hybrid methods need to be assessed as applicable to the region of interest.

- Must encourage users not to use only the ensemble mean or a sub-sample of ensemble members. All the ensemble members should be used.
- Web based tools need to be developed to allow users of the prediction information to tailor the underlying climate information more easily to their needs (e.g. climate range/thresholds, spatial scale(s)).
- Use regional mechanisms like Regional Climate Outlook Forums (RCOFs) to develop consensus based regional climate outlooks based on a scientific assessment of multiple prediction outcomes
- Actively promote user liaison to understand their climate information needs in decision making and also raise their awareness of the uncertainty aspects of seasonal forecasting
- Promote regional/national ownership of seasonal forecasts through effective and sustained capacity building and infrastructural support

6.0 Total Physical Climate System Seasonal Prediction

One of the key issues addressed in the workshop was how the interactions among the physical components of the climate system might potentially contribute to improved forecast quality and how to tap into this potential. Here we summarize the discussion and recommendations.

6.1 Cryosphere

Sea ice is an active component of the climate system and is highly coupled with the atmosphere and ocean at time scales ranging from synoptic to decadal. When large anomalies are established in sea ice, they tend to persist due to positive feedback in the atmosphere-ocean-sea ice system. These characteristics provide predictability of sea ice at seasonal time scales. In the Southern Hemisphere, sea ice concentration anomalies can be predicted statistically by a linear Markov model on seasonal time scales with reasonable quality. The best cross-validated skill is at the large climate action centers in the southeast Pacific and Weddell Sea, reaching 0.5 correlation with observed estimates even at 12-month lead time, which is comparable to or even better than for ENSO prediction.

Generally, seasonal prediction models use climatological ice or initialize the ice model with climatology. Despite the potential for predictability, sea-ice effects are poorly included in seasonal prediction models. There is a clear need to identify the remote effects (and causes) of sea ice anomalies and understand associated processes and influence on predictability.

Land ice/snow in the Northern Hemisphere is a highly variable surface condition in both space and time making it a viable candidate for amplifying atmospheric anomalies. For example, it has been demonstrated that the time series for fall Eurasian snow cover is significantly correlated with the winter Arctic Oscillation. There is also evidence that spring snow cover can significantly influence climate variability. Models typically have some initialization and representation of snow cover effects. The quality and impact of these effects are unknown.

Recommendations: Assess the impact of Antarctic sea-ice anomalies on the atmosphere and the impact of specifying observed ice cover in coupled forecast runs (feasible for models with/without active ice models). Examine the quality of real-time snow initial conditions. There is scope for conducting snow initial condition sensitivity experiments. Could tentative steps be taken toward initializing existing ice models?

6.2 Land Surface

A wealth of numerical model analyses and some complementary observational studies have shown that soil moisture anomalies can induce anomalies in precipitation and air temperature. This is not true everywhere; in very dry regions, evaporation is too low for its variation to influence an atmosphere already disinclined to generate rainfall, and in very wet regions, evaporation is controlled mostly by atmospheric demand and thus does not respond strongly to soil moisture variations. In the transition zones between wet and dry regions, however, soil moisture variations do seem to lead to precipitation and air temperature variations, as revealed by the GEWEX/CLIVAR-sponsored GLACE modeling experiment. This finding, coupled with the fact that soil moisture anomalies (both in models and in nature) persist for weeks to months, suggests that the initialization of land moisture states in a seasonal forecast system may lead to improved forecasts, at least in some areas.

Recommendations: The above potential will be addressed in GLACE-2, an ambitious follow-on to GLACE. In GLACE-2, modeling groups will perform the same two series of 2-month forecasts, one in which land moisture states are initialized realistically (through an offline exercise utilizing realistic meteorological forcing) and one in which the land state initialization is essentially random. Evaluation of precipitation rates and air temperatures produced in both sets of forecasts against observations will allow us to isolate any increase in skill stemming from the realistic land state initialization. GLACE-2 ostensibly focuses on soil moisture, but because of the way all the land states are initialized together in the experiment, GLACE-2 also addresses, at least peripherally, two other potential land-based sources of predictive skill: snow cover and subsurface heat reservoirs. Future studies should address the snow component more directly (e.g., with spring transition forecasts). A fourth potential source of land surface memory, vegetation health (leafiness), is also worth considering in future studies.

6.3 Stratospheric Processes

In many ways the stratosphere acts as a boundary condition for the troposphere. The stratospheric circulation can be highly variable, with a time scale much longer than that of the troposphere. The variability of the stratospheric circulation can be characterized mainly by the strength of the polar vortex, or equivalently the high latitude westerly winds. Stratospheric variability peaks during Northern winter and Southern late spring. When the flow just above the tropopause is anomalous, the tropospheric flow tends to be disturbed in the same manner, with the anomalous tropospheric flow lasting up to about two months. The surface pressure signature looks very much like the North Atlantic Oscillation or Northern Annular Mode. Surface temperature signals are also similar to those from the NAO and SAM and there are associated effects on extremes. In particularly sensitive areas such as Europe in winter, experiments suggest that the influence of stratospheric variability on land surface temperatures can exceed the local effect of sea surface temperature.

The stratospheric aspects of seasonal prediction can only be captured by models that properly simulate stratospheric variability. Thus far, the stratosphere´s potential to improve seasonal forecasts is largely untapped. It is essential that seasonal forecast models simulate the intense, rapid shifts in the stratospheric circulation, as well as the downward propagation of circulation anomalies through the stratosphere. In addition, models must be able to simulate the poorly understood connections between the lower stratosphere and the tropospheric circulation.

To maximize predictability from the stratosphere, forecasting systems also need to predict stratospheric warmings and other variability at as long a lead time as possible. Recent experiments suggest this is typically 1 to 2 weeks but it can be longer in some cases.

Recommendations: 1) To exploit predictability from stratospheric processes, seasonal forecast models must have accurate representations of stratospheric processes. Our current understanding suggests that a model would be required to have a model top significantly above the stratospause (or the order of 0.01hPa) and to have a high vertical resolution (of the order of 30 levels in the stratosphere) to have a good simulation of the stratosphere. 2) It will also be necessary to diagnose stratosphere-troposphere coupling in seasonal forecast models. This can be done by producing diagnostics based on multi-level annular mode indices. To do this, daily, zonal-mean geopotential is required at all model (or pressure) levels. If these zonal-mean geopotential data are available, then diagnostics such as variance of the annular modes, and timescale of the annular modes can be examined and compared to observations. Such analyses are necessary to know if the model's representation of stratosphere-troposphere coupling is realistic.

Summary Comments:

This report represents the consensus of a large number of active seasonal prediction scientists. It is clear that much work needs to be done to improve seasonal prediction quality and to maximize the use of seasonal prediction information for societal benefit. While there are notable successes in seasonal prediction (i.e., high quality predictions of tropical Pacific sea surface temperature), there are daunting challenges in terms of predicting land surface temperature and rainfall. Model fidelity and forecast initialization remain continue to limit forecast quality. The seasonal prediction community is in the early stages of interacting with the climate change community – this interaction needs to be fostered and encouraged. Indeed, the use of forecast information for societal benefit will ultimately know no boundaries between seasonal prediction and climate change.

The use of seasonal prediction information for societal benefit is in its early stages, yet the initial promise still remains elusive. The use of seasonal prediction information is hampered by forecast quality and successfully communicating uncertainty and the limitations of seasonal forecasts. We are, however, clear that increased interactions between climate scientist and climate information users will ultimately increase the use of seasonal prediction information for societal benefit.

While the assessment of forecast quality presented here is by no mean comprehensive, it is establishing a process for the seasonal prediction community to regualarly evaluate progress both in forecast quality and value. This document is the starting point, and when warrented, will be updated and made available both in electronic copy (by means of the CLIVAR WGSIP webpage: www.clivar.org/organization/wgsip/wgsip.php) and in hard copy.

Appendix A

The TFSP Climate-system Historical Forecast Project (CHFP)

One of the overarching goals of WCRP is to determine the predictability of the complete climate system on time scales of weeks to decades. The prediction experiments discussed here are a complement to this overarching predictability goal and can be used to make predictability assessments. Here we focus on seasonal time scales. By complete climate system, we mean contributions from the atmosphere, oceans, land surface, cryosphere and atmospheric composition in producing regional and seasonal climate anomalies. Advances in climate research during the past decade has lead to the understanding that modeling and predicting a given seasonal climate anomaly over any region is incomplete without a proper treatment of the effects of SST, sea ice, snow, soil wetness, snow cover, vegetation, stratospheric processes, and chemical composition (carbon dioxide, ozone, etc.,). The observed current climate changes are a combination of anthropogenic influences and natural variability. In addition to possible anthropogenic influence on climate due to changing the atmospheric composition, it is quite likely that land use in the tropics will undergo extensive changes, which will lead to significant changes in the biophysical properties of the land surface, which in turn will impact atmospheric variability on seasonal time scales. It is therefore essential that the past research by two somewhat non-interacting communities (i.e., climate change and seasonal prediction) be merged into a focused effort to understand the predictability of the complete climate system.

This problem of prediction and predictability of seasonal climate variability is necessarily multi-model and multi-institutional. We argue that the multi-model approach is necessary because there is compelling evidence that, with imperfect models, perturbing the physics of the models is superior to perturbing initial conditions of one model in terms of resolving the probability density function or quantifying the uncertainty. A multi-model approach is essentially a simple and consistent way of perturbing the physics. Moreover, by testing our hypotheses with multiple models it is possible to determine which results are model independent, and hence likely to be robust. This problem is also necessarily multi-institutional simply because the level of effort and computational resources required is just too large for any one institution.

Total Climate System Seasonal Prediction Experiment Proposal

The TFSP proposed a comprehensive seasonal prediction experiment that is designed to test the following hypothesis:

> There is currently untapped coupled predictability due to interactions and memory associated with all the elements of the climate system (Atmosphere-Ocean-Land-Ice).

The core experiment is an ‘interactive Atmosphere-Ocean-Land-Ice Prediction Experiment’ emphasizing the use of comprehensive coupled general circulation models, which includes realistic interactions among the component models. The experiment is to perform seven-month lead ensemble (10-members) predictions of the total climate system. If possible longer leads and larger ensembles are encouraged. The initialization strategy is to use the best available observations of all the components of the climate system.

While the emphasis in on comprehensive coupled general circulation models, uncoupled component, intermediate, simplified and statistical models are encouraged to participate where appropriate. The fundamental experimental design is to mimic real prediction in the sense that no "future" information can be used after the forecast is initialized. For example, the PROVOST or DSP experiments would be excluded because they use observed SST as the simulation evolves, whereas the SMIP/HFP experiment and DEMETER could be included as subset since no future information is used as the forecast evolves[1].

The component models should be interactive, but this is left open to accomplish a wider participation, e.g. for groups without sea-ice or vegetation model. The only requirement is that no "future" information is used once the prediction is initialized. This requirement necessarily includes any tuning or training either the component models or the development of statistical prediction schemes.

Thus, the component models are:

- Ocean – Open but interactive (e.g., slab mixed layer or GCM)
- Atmosphere – Open but interactive (most likely a GCM)
- Land – Open but interactive (e.g. SSiB, Mosaic, BATS, CLM, Bucket)
- Ice – Open but interactive (e.g., thermodynamic or dynamic)

The results of these experiments provide a framework for future experiments, specifically these prediction results will:

(i) Provide a baseline assessment of our seasonal prediction capabilities using the best available models of the climate system and data for initialization.
(ii) Provide a framework for assessing current and planned observing systems, and a test bed for integrating process studies and field campaigns into model improvements
(iii) Provide an experimental framework for focused research on how various components of the climate system interact and affect one another
(iv) Provide a test bed for evaluating IPCC class models in seasonal prediction mode

The TFSP recognizes that certain elements of the proposed experiment are already part of various WCRP activities. The intent here is to leverage these ongoing activities and to coordinate and synthesize these activities into a focused seasonal prediction experiment that incorporates all elements of the climate system. These experiments are the first necessary steps in developing seamless weekly-to-decadal prediction of the complete climate system.

The parameters of the experiment are as follows:

(i) Coupled models and resolution are left to the individual participants, but it is desirable that the models have a realistic simulation of the atmosphere, ocean, land and ice and the interactions among these components. Simplified component models (e.g., slab mixed layer or statistically

[1] The SMIP/HFP experiment is viewed as a subset of the experiments proposed here since they do not necessarily include feedbacks from land surface or sea ice processes or the initialization of these components of the climate system.

predicted ice) are acceptable as long as the no future information is used in developing the simplified model.

(ii) Atmospheric initial states to be taken from NCEP (or ECMWF) reanalysis each February, May, August, and November of each year from 1979-present. Forecasts should be initialized on 00Z and 12Z on the last five days of each preceding month forming a 10-member ensemble. Other strategies for generating the ensemble members are acceptable as long as the basic principle of no future information as the forecast evolves is not violated. Each ensemble member should be run for at least six months. Additional ensemble members and longer leads are encouraged (out to decades).
   a. Additional retrospective forecasts using each month of each year from 1979-present are encouraged.
   b. Additional retrospective forecast using initial conditions from each February, May, August, and November 1960-1978 are encouraged.

(iii) Oceanic initial states: (if appropriate) to be taken from most appropriate ocean data assimilation system.

(iv) Sea Ice initial states: (if appropriate) to be taken from best available observational data.

(v) Land initial states: (if appropriate) to be taken from most appropriate land data assimilation system or consistent offline analyses driven by observed meteorology (i.e., GSWP; GLACE2).

(vi) Atmospheric output:
   a. Every 24 hours at 00 GMT-
      i. Pressure levels (instantaneous): Geopotential Height, Temperature, Velocity and specific humidity for 850, 500, 200, (if available 100, 50, 10; these higher pressure levels are used for interactions with SPARC) hPa.
      ii. Surface (instantaneous): 2m Tmax – daily, 2m Tmin – daily, Total soil moisture, Snow depth, Sea surface temperature and surface radiative temperature over land (if available), Mean sea level pressure, soil heat flux over land (if available).
      iii. Surface (accumulated): Total precipitation, Downward surface solar radiation, Downward surface longwave radiation, Surface net solar radiation, Surface net longwave radiation, Top net solar radiation, Top net longwave radiation, Surface momentum flux, latent and sensible heat flux.
   b. Every 6 hours at 00, 06, 12, 18 GMT-
      i. Surface (instantaneous): Total cloud cover, 10m wind, 2m Temperature, 2m Dew Point, 2 m specific humidity.

(vii) Oceanic output (where appropriate)
   a. Every Month-
      i. Accumulated temperature, salinity and currents in the (at least) upper 400 m, surface fluxes of heat, momentum and fresh water, sea level height, mixed layer depth.
   b. Every 24 hours at 00 GMT-
      i. Temperature, salinity and currents in the (at least upper 400 m) at the equator, 2N and 2S (5N and 5S optional).
   c. Every 6 hours at 00, 06, 12 18 GMT-
      i. Surface fluxes of heat, momentum, and freshwater. Sea surface temperature and mixed layer depth

(viii) Sea Ice output (where appropriate)
   a. Every 24 hours at GMT –
      i. Surface fluxes of heat and momentum. Snow cover, Sea ice concentration, thickness and temperature.

(ix) Soil wetness and vegetation predicted.
(x) Snow cover and depth predicted.
(xi) Chemical Composition (carbon dioxide, ozone …) prescribed and varying. This explicitly includes the transient changes in the chemical composition from 1979-present.
(xii) Finally, it is noted that some forecast provides may not be able to provide the complete list noted here. Participation is strongly encouraged even if all the data requirements cannot be met.

Examples of Potential Diagnostic Sub-Projects

In order to maximize collaboration and duplication of effort, the proposed experiment will include a diagnostic sub-project approval process. The following is an abbreviated list of potential sub-projects. It is anticipated that a large number of additional sub-projects will be implemented as the experimental results become available.

- Limit of Predictability Estimates: One potential estimate for the limit of predictability is to determine when a particular forecast probability density function (pdf) is indistinguishable from the climatological pdf of the forecasts.
- ENSO mechanism diagnostic: Recharge oscillator versus delayed oscillator, role of stochastic forcing, westerly wind events.
- Impact of the AO on seasonal predictability
- Regional predictability
  - Local land surface predictability
  - Extreme events
  - Monsoon predictability
  - Diurnal cycle in ocean
  - Diurnal cycle in the atmosphere
- Coupled Feedbacks
  - Intra-seasonal oscillations
- The diagnostic sub-projects will also include extensive interactions with the applications community and the regional panels within CLIVAR, GEWEX, SPARC and CliC. These interactions and collaborations are viewed as critical elements of the implementation plan and are strongly encouraged.
- Opportunity to carry out supplementary experiments – e.g., case studies, process studies.

2.3 CHFP Time-Line

- *Experiment announcement from JSC March 2006.*
- *Experiment to be completed end of 2008*.
- *First WCRP Seasonal Prediction Workshop Mid-*2007
- Second *WCRP Workshop 2009*.

Appendix B

Some Examples Seasonal Prediction Forecast Quality for Various Regions

This appendix provides some examples of seasonal prediction quality in various regions around the globe. This is intended to grow over time and the discussion included here is merely a brief overview.

North America

Using retrospective forecast data provided by WCRP SMIP-2, DEMETER, CFS, and IRI, the **US CLIVAR Prediction, Predictability and Applications Interface Panel (PPAI; Goddard et al)** have produced a skill assessment for seasonal mean 2m air temperature and total precipitation for June-July-August and December-January-February for the period 1981-2000 over North America. The model skills are assessed individually as well as for a simple multi-model combination. PPAI examine both deterministic skill of the ensemble means, as judged by mean squared error (MSE), and its decomposition (correlation, bias, and variance ratio), and probabilistic skill, as measured by reliability and relative operating characteristic areas. In addition, PPAI examined the ability of these dynamical prediction tools to capture aspects of the climate variability of particular interest to society, such as regional temperature trends and multi-year regional drought.

Although models vary in their performance metrics, there is no clear preference between coupled models and atmosphere-only models. Further, as found in many previous studies, a simple combination of many reasonable quality models improves most performance metrics relative to any individual model. The one aspect of the forecast degraded by the multi-model combination is sharpness of the probabilities, most likely owing to our simple and un-calibrated combination of the models. A surprising finding is that although the skill metrics would suggest temperature is better predicted than precipitation, the models predicted the persistent drought conditions of the southwest North America much better than the upward temperature trend, suggesting the potential importance of missing sources of climate variability in these seasonal prediction models.

Southern Africa

The **South African Weather Service (SAWS)** issues seasonal rainfall forecasts for South Africa, Lesotho, Swaziland, Namibia and Botswana. Unfortunately, verification of these forecasts for South Africa’s neighbors cannot be performed by the SAWS owing to a lack of observed data. The forecasts are based on the output from statistical and dynamical models. Seasonal rainfall forecast skill of the models is primarily found during summer (DJF) and during autumn (MAM) at lead-times of a few months, but very little skill is found during spring (SON) and winter (JJA). Forecast skill is also heavily dependent on ENSO. Recent work on the use of coupled models (DEMETER) suggests that very little forecast information is to be derived during ENSO-neutral years, but multi-model forecasting systems (tested with simulation data) show general improvement in forecast skill. Consensus probabilistic forecasts (subjectively compiled from model output) are skilful during the major summer rainfall months, but show very low skill during winter. The Department of Agriculture incorporates the forecasts in advisories for farmers. Commercial companies also use these forecasts, but it has not been determined yet if they gain financially from using them. Such companies include soft drinks and beer and commercial agricultural companies.

South America
From the middle of 2006, outputs of relevant Seasonal Forecast models, from ECMWF, NOAA-CFS, CPTEC, IRI and CIIFEN were analyzed for regional skill and to identify critical areas.

The analysis showed low to medium skill for precipitation forecast and good skill for SST forecast during the last half of 2006, when the latest ENSO event started its evolution, however, low skill was evidenced between October and December 2006, in most of South America. The unusual evolution of the final ENSO phases was not well simulated by all the analyzed models.

Some comparison between statistical and dynamical models in the region suggest a stronger than normal influence of the physical processes from the Atlantic in South American precipitation patterns, with more emphasis on Bolivia, Peru and Ecuador. A non typical upper atmosphere circulation pattern was persistent between the Caribbean and the northeast of South America. This atmospheric feature blocked the convection due to the ENSO warming in NIÑO 1+2 region most of the time and could be one element that contributed to its rapid weakening. For a second consecutive year, SST over NIÑO 1+2, showed a different evolution than the central equatorial Pacific, however, a significant improvement of SST forecast was evidenced after the abrupt switch from warm to cold temperatures in February 2007.

The documented experience of this year suggests the importance of the processes close to the Eastern Pacific coast, and the influence of the Atlantic and the Caribbean ocean-atmosphere circulation over South America.

The use of the seasonal forecast was considerable in the South America region. Despite of the limitations of the model, available information facilitated decision making that in some cases were followed by adequate preventive actions; however, extreme events and existing vulnerability caused considerable damage and social and economical impacts in several countries.

It is concluded that, combining global seasonal prediction with regional statistical and dynamical models could improve predictability if the regional data is available for assimilation by global models. CIIFEN is working in close cooperation with National Institutions of the region to work in this direction and strengthen the regional capacities of statistical and dynamical seasonal forecasts.

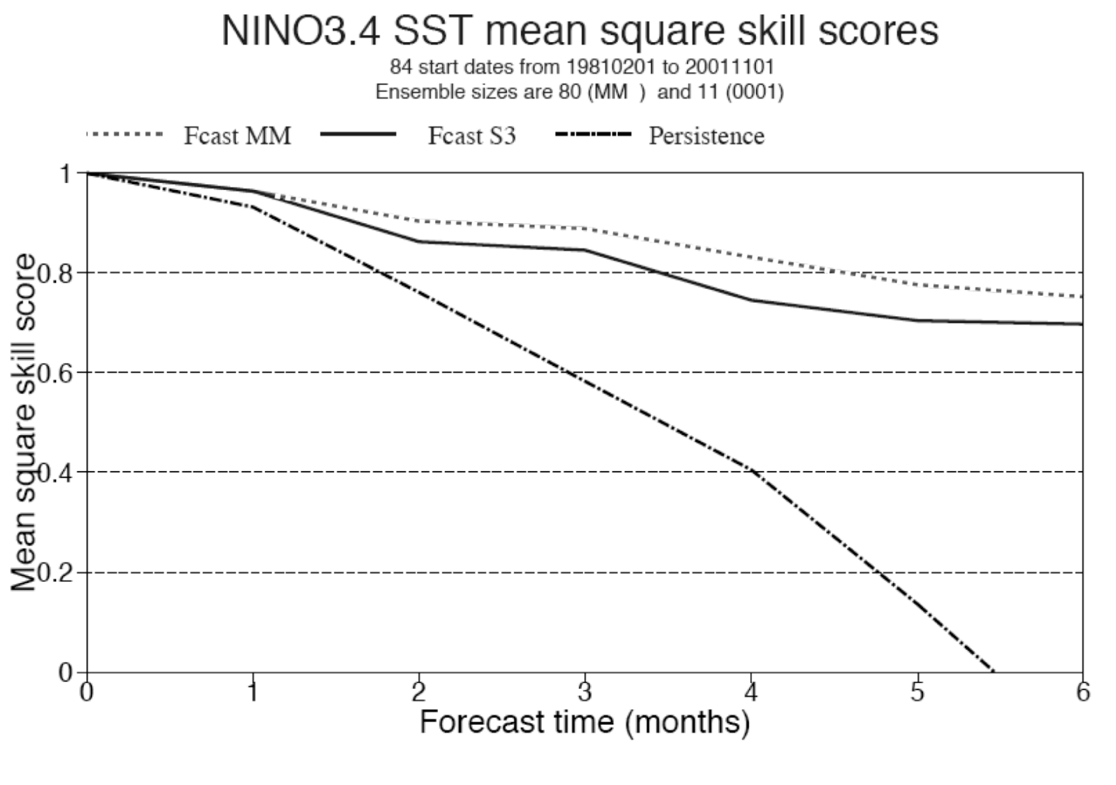

Figure 1: Nino3.4 forecast quality based mean squared skill scores from an individual model (black curve) and from the multi-model ensemble (grey dotted curve).

Region	2m Temperature				Precipitation			
	JJA		DJF		JJA		DJF	
	$E_T^-(x)$	$E_T^+(x)$	$E_T^-(x)$	$E_T^+(x)$	$E_p^-(x)$	$E_p^+(x)$	$E_p^-(x)$	$E_p^+(x)$
Australia	**10.7**	**10.1**	1.3	-0.4	-1.3	-2.5	-3.1	-3.6
Amazon Basin	**14.4**	9.1	**23.4**	**25.7**	2.2	2.1	**9.5**	**8.9**
Southern South America	**8.5**	**8.2**	-1.2	1.8	**7.8**	5.0	-0.7	-2.8
Central America	**12.1**	**9.9**	**14.8**	6.3	2.6	-0.7	8.7	8.5
Western North America	**6.5**	**7.7**	3.9	2.3	3.2	**5.5**	-0.6	0.0
Central North America	-4.1	-3.6	**-7.5**	0.3	-1.8	**-7.0**	3.7	5.3
Eastern North America	0.6	5.7	4.1	9.5	**-4.5**	**-8.3**	**9.2**	6.0
Alaska	3.0	2.1	0.0	-0.7	-0.1	0.3	2.4	4.9
Greenland	3.6	4.2	**8.0**	5.8	**-1.4**	-0.5	-2.1	-2.0
Mediterranean Basin	**7.6**	**10.7**	3.2	3.2	-0.5	0.1	1.6	-0.9
Northern Europe	-4.4	-4.2	4.8	2.9	-1.0	1.9	-1.1	-0.9
Western Africa	**10.4**	**11.8**	**18.1**	**17.2**	-1.6	-2.0	**-4.9**	**-3.5**
Eastern Africa	**12.6**	5.8	**13.3**	**10.3**	0.1	-0.3	1.2	0.6
Southern Africa	5.6	-1.1	**15.9**	**15.7**	0.7	-1.2	5.4	3.6
Sahara	**7.6**	**7.4**	6.9	3.9	**-2.6**	**-4.8**	**-2.7**	**-2.7**
Southeast Asia	10.7	5.9	8.7	**18.1**	**14.7**	**10.3**	3.4	2.5
East Asia	**4.7**	**7.9**	**10.8**	**10.0**	0.6	-1.0	-1.6	-0.9
South Asia	4.9	**13.1**	**7.6**	**8.6**	-1.6	**-3.0**	2.0	0.5
Central Asia	0.8	3.8	1.3	-0.4	0.5	0.1	-3.1	-3.6
Tibet	**10.7**	**10.1**	**23.4**	**25.7**	-1.1	0.0	**9.5**	**8.9**
North Asia	**14.4**	9.1	-1.2	1.8	-1.3	-2.5	-0.7	-2.8

Table 1:
Forecast quality of the DEMETER multi-model seasonal re-forecasts in terms of Brier Skill Scores (BSS) for near-surface temperature and precipitation upper and lower tercile categories in JJA and DJF for 21 standard land regions (multiplied by 100). The scores for $E^{\pm}_{T,P}(x)$ have been computed over the re-forecast period 1980-2001 using seasonal means from 1-month lead ensembles started on the 1st of May/November. Bold underlined numbers indicate scores with a probability p≥0.9 that a random sample based on a 10,000 bootstrap re-sampling procedure would yield BSS<0 (significantly negative) or BSS>0 (significantly positive).

Co-sponsors of the workshop: