INTERNATIONAL COUNCIL FOR SCIENCE

INTERGOVERNMENTAL OCEANOGRAPHIC COMMISSION

WORLD METEOROLOGICAL ORGANIZATION

WORLD CLIMATE RESEARCH PROGRAMME

# Report of the 11th Session of Working Group on Seasonal to Interannual Prediction (WGSIP)

**7-8 June 2007, Barcelona, Spain**

November 2007

*ICPO Publication Series No.124*

**WCRP Informal Report No.16/2007**

CLIVAR is a component of the World Climate Research Programme (WCRP). WCRP is sponsored by the World Meteorological Organisation, the International Council for Science and the Intergovernmental Oceanographic Commission of UNESCO. The scientific planning and development of CLIVAR is under the guidance of the JSC Scientific Steering Group for CLIVAR assisted by the CLIVAR International Project Office. The Joint Scientific Committee (JSC) is the main body of WMO-ICSU-IOC formulating overall WCRP scientific concepts.

**Bibliographic Citation**

INTERNATIONAL CLIVAR PROJECT OFFICE, **2007: Report of the 11th Meeting of the Working Group on Seasonal to Interannual Prediction (WGSIP). November. International CLIVAR Project Office, CLIVAR Publication Series No. 124. (not peer reviewed).**

**CONTENTS**

Action Items

1. WGSIP should consider how to contribute to additional WCRP crosscutting activities, particularly on the monsoons, extremes, and the IPY (*all*).

2. WGSIP proposes to co-lead WCRP crosscutting activity on decadal prediction activity with WGCM, bringing together their seasonal prediction and climate (IPCC) activities. Report back to JSC WGSIP decision to co-lead WCRP decadal activity (*T. Stockdale to attend 11th Session of WGCM, Hamburg 3-5 Sept. 2007*).

3. Discuss general ideas on decadal prediction experiment design (*initially G. Boer, T. Stockdale and B. Kirtman*).

4. Discuss initialisation of decadal prediction experiments with GSOP.

5. WGSIP to retain TFSP mission (*agreed by all*).

6. Circulate WCRP Workshop on Seasonal Prediction position paper for further review (*B. Kirtman*).

7. WGSIP to request an active but modest role in the IPCC process, especially in terms of influencing who is selected to represent seasonal prediction in any future AR5. Communicate the fact that WGSIP is willing to make someone available for the AR5 scoping meeting if required. (*B. Kirtman, T. Stockdale, and A. Henderson Sellers*)

8. WGSIP requests a one page write up to digest proposed simplified protocol for studying model error (*E. Guilyardi*).

9. WGSIP to discuss role in defining protocols in coordinated experimentation in CHFP sensitivity experiments.

    a. SPARC: The SPARC community needs to be challenged to demonstrate predictive skill, and encouraged to look at the CHFP data (*G. Boer*).

    b. GEWEX: Support continued investigation through the GLACE project into the quality of real-time soil moisture initialisation (*R. Koster*).

    c. CliC: Need more on how the role of cryospheric processes should be assessed (*B. Kirtman*).

10. CHFP Data Handling Strategy: recommend that software be adapted to keep up with CHFP output format. Recommend that data be served on a 2.5x2.5 grid and, if possible, to keep the option of retrieving data on the original grid. Document needs to be circulated (*T. Stockdale*).

11. Need to develop a webpage to give directions to CHFP data. (*T. Stockdale, A. Pirani*).

12. WGSIP should take the proposal for a standard hindcast verification system further by comparing some plots of ACC, MSSS and RMS error for temperature and precipitation at the next meeting. This should ideally be done together with plots from applications models, perhaps for a limited area (*all*).

13. Need to put information on the status of SMIP onto the webpage (*G. Boer, A. Pirani*).

14. A white paper is in preparation this summer on seamless range and approaches for user applications (*A. Morse*).

15. WGSIP recommends that the Pacific Implementation Panel coordinates ocean model error analysis experiments and sensitivity of models to resolution, bulk formulae etc, to include CORE forcings and to report back to WGSIP (*M. Balmaseda, representing PIP*).

16. Encourage AAMP to make use of SMIP and CHFP data, and to do additional tier-1 (uncoupled) and tier-2 (coupled) experiments in monsoon region as required.

17. Encourage VACS to use data from CHFP Experiment. SAWS could provide output from multi model output that is planned at SAWS to DMC in exchange for rainfall data. Propose that this becomes a VACS project that can be endorsed by WGSIP (*W. Landman*).

18. Membership actions to be discussed by email (*T. Stockdale and B. Kirtman*).

19. Consider organising the next meeting possibly in Salvador, Brazil, in early 2009 (*T. Stockdale, B. Kirtman*).

## 1. Introduction

The 11th session of WGSIP took place on 7-8 June 2007 immediately after the WCRP Workshop on Seasonal Prediction in Barcelona. This had several advantages, notably the presence of several distinguished visitors on the first afternoon to provide us with perspectives from WCRP, the JSC and the CLIVAR SSG, and the necessity to deal with matters concisely, so as to fit the meeting within the scheduled 1.5 days. We would like to acknowledge our hosts from the Catalan Met Service, who provided facilities for the WGSIP meeting in addition to their support for the Seasonal Prediction Workshop itself.

The report is arranged by topic and is broadly chronological, but a few of the later items were actually dealt with on the first afternoon due to the limited availability of some of our invited guests.

## 2. Overview on WCRP strategic framework and its implementation

The World Climate Research Programme (WCRP, http://wcrp.wmo.int/) published its Strategic Framework: 2005-2015: Coordinated Observation and Prediction of the Earth System in August 2005. The document reiterates the WCRP objectives to determine the predictability of climate and the effect of human activities on climate. The strategic framework seeks 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. The key issues are:

- Move from physics-only to Earth-System models (with IGBP, http://www.igbp.net/)
- Prediction across all timescales: “seamless predictions”
- Develop sustained climate observing system with GCOS, GEOSS etc…
- Integration of models and data: 1) use of data assimilation to initialize models over widest range of climate prediction timescales possible; 2) synthesis through reanalysis (atmosphere, ocean, coupled)
- Link to applications through existing mechanisms (e.g. START http://www.igbp.net/, World Climate Application Programme) and new ones

The 28th session of the WCRP Joint Scientific Committee was held in Zanzibar, Tanzania, from 26 to 30 March 2007. The meeting aimed to determine the future of the WCRP and review progress in implementing its Strategic Framework 2005-2015. All the documents prepared for the JSC-28 are available on the WCRP website: http://wcrp.wmo.int/TableDocs.html.

### 2.1. WCRP crosscutting activities

Particularly relevant to CLIVAR is the outcome of the JSC Task Team on Implementing the WCRP Strategic Framework, which recommends WCRP to focus on crosscutting issues, aimed at meeting society’s and stakeholder needs. The crosscutting activities identified by JSC are:

- Anthropogenic Climate Change
- Atmospheric Chemistry and Climate
- Monsoons and the Year of Tropical Convection
- Decadal Prediction
- Extreme Climate Events
- International Polar Year
- Sea-level Rise
- Seasonal Prediction

WGSIP is taking the lead on seasonal prediction, and will be significantly involved in decadal prediction (see later).

> **ACTION** WGSIP should consider how to contribute to additional WCRP cross-cutting activities, particularly on the monsoons, extremes, and the IPY (all).

### 2.2. Funding issues

The total budget allocated to core CLIVAR activities will be reduced considerably for the period 2008-9 as a result of introducing cross-cutting activities within WCRP, zero nominal growth in WCRP budgets and the ever weakening state of the US dollar. Funding being allocated to the WCRP cross-cutting activities will be relevant for the CLIVAR Panels and Working Groups directly involved, as is the case for WGSIP in terms of the decadal prediction activity.

How best to handle the lack of resources is still not clear. It is hoped that the CLIVAR Panels and Working Groups will continue to aim to meet every 18 months, generally in conjunction with a scientific workshop or conference. Since CLIVAR funding will be insufficient to fully support all groups, meetings may need to actively seek additional sources of funding. WGSIP noted the situation.

## 3. WCRP cross-cutting theme: Decadal Prediction

The decadal prediction cross-cutting activity was presented by T. Palmer at the 28$^{th}$ Session of the WCRP JSC. The JSC endorsed the proposal and recommended that WGSIP take co-leadership of this activity with WGCM. If accepted, this would change the mandate of WGSIP and any decision should be reported back to the JSC through the CLIVAR SSG.

WGSIP raised concerns of coordinating any major decadal project, but T. Palmer assured the participants that this would not be a major project equivalent, at least at this stage, to the Climate-system Historical Forecast Project (CHFP – see item 4.4) experiment for example, but more an initial step to address the technical aspects of decadal prediction experiments, such as the best initialisation of an ocean model from analyses.

After some discussion, a proof of concept approach was approved by WGSIP as well as the idea to co-lead the activity and find common ground with WGCM. A number of groups mentioned that they are not in a position to run decadal prediction experiments, either for political or practical reasons. Shukla pointed out that the intention was that the IPCC groups were expected to do most of the experimentation, and WGSIP took the view that we do have something to bring to the table in terms of experience of the initial value forecasting problem.

Furthermore, climate change is inextricably part of the seasonal forecast calibration problem, and our ability to recover observed trends. Processes, for example aerosols, that are mainstream in the climate community need to be incorporated into seasonal prediction. It was also noted that the concept of testing climate change models in seasonal prediction mode has been endorsed by the climate community (see item 4.3). Given the large number of overlapping issues, it was suggested that WGSIP and WGCM should meet together, perhaps in conjunction with a mini-workshop on decadal prediction. The CLIVAR basin panels and the Global Synthesis and Observations (GSOP) Panel also need to be involved in the ocean initialisation problem.

**ACTION** WGSIP proposes to co-lead WCRP cross-cutting activity on decadal prediction activity with WGCM, bringing together their seasonal prediction and climate (IPCC) activities. Report back to JSC WGSIP decision to co-lead WCRP decadal activity (T. Stockdale to attend 11th Session of WGCM, Hamburg 3-5 Sept. 2007).

**ACTION** Discuss general ideas on decadal prediction experiment design (initially G. Boer, T. Stockdale and B. Kirtman).

**ACTION** Discuss initialisation of decadal prediction experiments with GSOP.

G. Boer presented some results from predictability studies looking at decadal multi-model potential predictability.

## 4. Review of WCRP Task Force on Seasonal Prediction (TFSP) and related activities

The TFSP, which draws on expertise from all the WCRP core projects (CLIVAR, CliC, SPARC, and GEWEX), WGNE and WGCM, together with the Sea Level Task Team, pioneered the WCRP coordinated strategy initiated in 2005. The TFSP provided a mechanism that did not previously exist to evaluate potential predictability in the cryosphere, biosphere, stratosphere and the rest of the fully coupled climate system, while also delivering focus on the value to society of seasonal prediction.

The planned two year life time of the TFSP was discussed at the 28th Session of the WCRP JSC and resulted in the recommendation that the TFSP should cease to exist after the WCRP Workshop on Seasonal Prediction held in Barcelona on 4-7 June 2007 and should continue to function through the WGSIP.

**ACTION** WGSIP to retain TFSP mission (agreed by all).

### 4.1. Overview and recommendations from the WCRP Workshop on Seasonal Prediction, Barcelona 4-7 June 2007-07-30

The TFSP in collaboration with the core programs of the WCRP (CLIVAR, CliC, SPARC and GEWEX) and the WMO World Climate Programme (WCP) organized the First WCRP Seasonal Prediction Workshop, which was hosted by the Catalonian Meteorological Service on June 4-7 2007 at the World Trade Centre in Barcelona Spain. The Workshop was also co-sponsored by the Climate Information and Prediction Services (CLIPS) Project of the WMO WCP, the European Science Foundation (ESF), the Spanish Ministry of Science, the School of Physics of the University of Barcelona, the Catalan Water Agency, US CLIVAR, the US National Science Foundation (NSF), the US National Aeronautics and Space Administration (NASA) and the US National Oceanic and Atmospheric Administration (NOAA). Attendance was very high, with around 180 people coming from over 30 countries.

The main objective of the Workshop was to enable the TFSP to make an assessment of current skill in seasonal prediction, with particular emphasis on surface temperature and precipitation. A WCRP statement of the present state-of-the-art of seasonal forecasting, together with a forward-looking road map of recommendations for the seasonal forecasting community in being produced as an outcome of the Workshop (*see Appendix C for a summary, also published in CLIVAR Exchanges Issue 43*). 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 will outline a set of specific recommendations for improving seasonal prediction skill and enhancing the use of seasonal prediction information for applications.

The Workshop focused on addressing two basic overarching questions:

1. What factors are limiting our ability to improve seasonal predictions for societal benefit?
2. What factors are limiting our ability to use seasonal predictions for societal benefit?

The Workshop was arranged into three main sessions focusing on the following objectives:

1. To assess the quality, as well as value, of forecasts, model fidelity and ocean-atmosphere interactions.
2. To address 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. This Session was co-organized by other WCRP Projects (SPARC, GEWEX and CliC).
3. To assess seasonal forecasts at a regional scale and highlight issues important for interfacing seasonal forecasts with applications including calibration, downscaling and validation, looking at whether there is an emerging consensus on approach and methodology. This session was co-organized by the CLIVAR Monsoon Panels (VAMOS, AAMP and VACS).

**ACTION** Circulate WCRP Workshop on Seasonal Prediction position paper for further review (B. Kirtman).

### 4.2. Seasonal Prediction Assessment and the IPCC Process

WGSIP discussed whether or how the seasonal prediction assessment initiated by the Workshop might become a regular process. Given the interest in some quarters for an assessment of seasonal prediction skill to be part of future IPCC assessments (see later) one possibility might be to undertake future reviews of skill within the IPCC process. However, it was generally felt that a regular evaluation of seasonal forecast skill would work better as a grass-roots approach, occurring as and when the community feels it is required. This would avoid the process becoming political and allow for more freedom to focus on scientific issues, particularly in light of changing predictability in a changing climate. Nonetheless, no decisions were taken on this topic, and the question as to how to develop the seasonal forecast assessment process remains open.

A. Henderson-Sellers, Director of WCRP, is lobbying for the seasonal prediction community to become directly involved in the IPCC process and to be actively part of any future AR5. WGSIP should instigate this push, as the TFSP is no longer active. It was agreed that WGSIP does not want to take on any direct AR5 responsibility but could play a ‘consultancy’ role, for example by providing input through a statement on the current scientific state of the art of seasonal and decadal prediction and its role in climate change assessment. While the science of the role of seasonal and decadal prediction in climate change assessment is premature, it will probably be incorporated into an AR5. If WGSIP is not willing to become involved now, it is likely to be marginalised in future discussions and there is perhaps a risk that the scientific quality of what is done may suffer.

**ACTION** WGSIP to request an active but modest role in the IPCC process, especially in terms of influencing who is selected to represent seasonal prediction in any future AR5. Communicate the fact that WGSIP is willing to make someone available for the AR5 scoping meeting if required. (B. Kirtman, T. Stockdale, and A. Henderson Sellers)

### 4.3. Studying climate model error with seasonal integrations

One idea that emerged at the WGNE Workshop on Systematic errors was the value of testing "climate change" models using seasonal integrations (see item 6 for a full report on the meeting). There are two aspects to this, the first relates to model biases. Coupled integrations that are initialised as close to reality as possible can be evaluated more exactly against observations in terms of degree of model drift and the evolution of errors, than longer runs which are supposed to represent "mean climate". Model integrations are needed only over the appropriate timescales: some biases are inherent in just the atmosphere model and develop within a few days, other biases involving coupled feedbacks develop over a period of months to a year (of course, there may be other biases related to slow processes in the climate system, which are not addressed by seasonal timescale integrations). A second aspect of testing climate models with seasonal integrations is to look at ENSO forecast skill (and perhaps seasonal forecast skill in general). The first aspect is particularly helpful in the development phase of climate models, since the modest integration time allows rapid turnaround of experiments. The second aspect may be helpful in deriving metrics to help judge the credibility of projected changes in ENSO and dynamically forced regional climate change.

Several strategies for developing these ideas further were discussed. The first is to ensure that as many "IPCC" class models as possible take part in the CHFP experimentation. This was one of the original hopes behind the development of the CHFP plans, and WGSIP encourage as many groups as feel able to participate in CHFP.

A separate strategy is the concept of developing some written guidance on simple but effective ways of initialising and testing coupled models in seasonal forecast mode. The guidance would be targeted in particular at those groups who do not have access to suitable ocean initial conditions, and might be particularly helpful during the development phase of models. There are several ways in which initial conditions can be created. E. Guilyardi proposed a 'poor man's' coupled initialisation procedure, where runs are started from an optimised state where initial conditions are generated from realistic SST and nudged wind forcing and then the model is run in coupled mode. Other community initiatives proposing such coordinated simulations will be explored and the suitability of their protocol to the objectives discussed above will be assessed.

**ACTION** WGSIP requests a one page write up to digest proposed simplified protocol for studying model error (E. Guilyardi).

### 4.4. Status of the Climate-system Historical Forecast Project (CHFP - formerly TFSP Experiment): Experiment Protocols

The TFSP Seasonal Prediction Workshop sessions organised by the different WCRP Projects gave an overview of what is currently known on the potential predictability to be gained from including more Earth system coupled processes in seasonal forecasting. An outcome of the Workshop was that, in addition to the baseline CHFP experimental design, there is plenty of scope for further sensitivity studies addressing stratospheric, land surface and cryospheric processes.

Talks presented by the SPARC community at the WCRP Seasonal Prediction Workshop did not actually demonstrate any enhanced predictive skill. More needs to be done on developing diagnostics, on testing higher vertical resolution and the sensitivity to ozone changes. Only a subset of models in the CHFP Experiment will have the possibility of having a full or reduced stratosphere and an implementation of time-varying ozone.

Issues that should be addressed to assess the role of cryospheric processes on seasonal forecasts include:

- assessing the quality of real time snow initial conditions
- the impact of Antarctic sea ice anomalies on the atmosphere
- comparing snow cover from observations and forced model estimates
- assessing the impact of specifying observed sea ice cover in coupled forecast runs by models that can or cannot have an active sea ice model
- how existing sea ice models should be initialised.

**ACTION** WGSIP to discuss role in defining protocols in coordinated experimentation in CHFP sensitivity experiments.

- SPARC: The SPARC community needs to be challenged to demonstrate predictive skill, and encouraged to look at the CHFP data (G. Boer).
- GEWEX: Support continued investigation through the GLACE project into the quality of real-time soil moisture initialisation (R. Koster).
- CliC; To consider how the role of cryospheric processes should be assessed (B. Kirtman).

### 4.5. CHFP Data Handling Strategy

*Appendix D outlines the proposed CHFP Data Handling Strategy (T. Stockdale).*

The CHFP Level 1 experiment data output will consist of 1-2 Tbytes (half due to the ocean). All the proposed CHFP experiments for one model will lead to 8 T-bytes of data. A 20 T-byte disk will be enough to comfortably handle the level 1 output from ten models.

A twin track approach is proposed to facilitate data exchange:

1. Original grid output for exchange between the main centres, in international format (GRIB, NetCDF).
2. Data on a standard 2.5 x 2.5 degree grid for exchange with the scientific user community, in a standard NetCDF format with specified metadata to allow proper handling of multi-model ensemble forecasts. Data access for the wider community will be via an openDAP interface, which will provide data from multiple models in a completely standard format, and support both downloading of user-specified sub-regions and interfaces to other software such as the Climate Explorer.

A present limitation is that a single software solution does not exist that can handle both requirements seamlessly, so that the data on the original grid will have to be served separately to the data on the standard 2.5 grid (i.e., in separate files at least). To keep data volumes manageable, it is proposed that the original grid data will generally be served only by the originator of the data. It was reported that the plot software Grads is currently being developed to handle the new ensemble dimension of NetCDF, which was welcomed.

There was discussion on the importance and value of data centre(s) willing to host data from a variety of model producers. Even though the data model is fundamentally distributed, it was recognized that one or more active data centres would be of great value, both in terms of being able to serve substantial quantities of data from producers who would rather not do so themselves, but also in providing human and software resources to help with task of properly coding the required NetCDF metadata, and checking the data and metadata. The APCC might be able to provide some resources towards acting as a data centre, and it was agreed that the details of this would be explored. Other options will be pursued as necessary.

Other considerations were that some groups (e.g. VAMOS) will require daily fields, if not output that resolves the diurnal cycle for at least some of the integrations. Accessibility to the wider user community would benefit from some limited data, such as some a limited number of surface fields at lower resolution, being provided in ascii format. If the CHFP data can be accessed via openDAP protocols by software such as the Climate Explorer, this would go a long way toward satisfying such user needs, since such software makes it easy to download limited datasets in ascii.

> **ACTION** CHFP Data Handling Strategy: recommend that software be adapted to keep up with CHFP output format. Recommend that data be served on a 2.5x2.5 grid and, if possible, to keep the option of retrieving data on the original grid. Document needs to be circulated (T. Stockdale).

> **ACTION** Need to develop a webpage to give directions to CHFP data. (T. Stockdale, A. Pirani).

## 5. On-going WGSIP activities

### 5.1. Understanding forecast skill

B. Kirtman presented some results from the NCEP Climate Forecast System (CFS) run with a random interactive ensemble approach to look at what determines the overall limit of predictability, what limits predictability (uncertainty in initial conditions: chaos within non-linear dynamics of the coupled system, or uncertainty as the system evolves: external stochastic effects) and the role of model dependence (error).

COLA has performed a set of identical twin ENSO experiments that are specifically designed to examine how uncertainty in the oceanic initial condition versus uncertainty (or internal atmospheric noise) as the forecast evolves limits ENSO predictability. These experiments were performed with the NOAA CFS model using the operational coupling strategy and the interactive ensemble coupling strategy developed at COLA. The interactive ensemble coupling strategy is used to control the amplitude the variability in the air-sea fluxes due internal atmospheric dynamics. With this new coupling strategy we can isolate the relative roles of uncertainty due to internal atmospheric dynamics versus uncertainty in the oceanic initial condition.

The results of these experiments suggest that:

- (i) The idealized limit of ENSO predictability is considerably longer than suggested by the actual forecast error.
- (ii) Excessive noise as the forecast evolves is a significant contributor to the loss of predictability at all lead times, but
- (iii) Uncertainty in the ocean initial condition leads to rapid initial error growth that exceeds the growth due to excessive noise in the evolution.

### 5.2. WGSIP standard hindcast verification system

T. Ose presented a draft proposal prepared by M. Sugi and T. Ose (*See Appendix E*), recommending that the anomaly correlation coefficient should be used to measure skill for temperature and precipitation at standard global grid points. The intention is that this should be a measure used within WGSIP to help assess forecast model performance. This is intended to complement the WMO verification standards that help national weather services and end users to choose which global forecasting centre products should be used.

The anomaly correlation coefficient has the advantage of being a good initial assessment of skill that can be easily implemented. However, it can be misleading as good correlation can result from a few good cases in a large set of forecasts so that it still needs to be used as part of a range of skill scores. Another issue is that the use of correlation maps to assess skill in precipitation can be problematic as is the direct comparison of observed and modelled precipitation, where additional dynamical verification, such as an analysis of flow regimes, would go further in assessing whether a model is correctly capturing the processes, though not the exact observed location of precipitation. Other possible skill measures are the Mean Squared Skill Score (MSSS), which is similar to ACC but includes a sensitivity to the amplitude of anomalies, root mean square (RMS) error, and Taylor diagrams.

A. Morse pointed out that the user perspective should be included by looking at abstracted user metrics such as covariance between temperature and precipitation.

> **ACTION** WGSIP should take the proposal for a standard hindcast verification system further by comparing some plots of ACC, MSSS and RMS error for temperature and precipitation at the next meeting. This should ideally be done together with plots from applications models, perhaps for a limited area (all).

### 5.3. The status of SMIP

SMIP is now closed and the data is to be considered as a sub-set for the CHFP Experiment. SMIP 2-tier can carry on and if groups want to submit more results, COLA will take the data.

> **ACTION** Need to put information on the status of SMIP onto the webpage (G. Boer, A. Pirani).

G. Boer presented some of the lessons learnt from SMIP2. Participation in this intercomparison experiment has been modest despite efforts to recruit contributors. Reasons for this could be due to a lack on interest in the experiment, perhaps due to SMIP2 being similar to SMIP1 and since no specific scientific output, such as an analysis or report, was planned. Other reasons include the lack of a clear policy on data with no dedicated data repository, the absence of funding or organisational support, and a clash with other commitments such as the IPCC AR4. Future WGSIP projects will need a clear scientific goal and data policy, funding and organisational support in terms of maintaining a website and for the organisation of workshops and meetings.

### 5.4. Activities related to applications

A. Morse underlined the importance of data availability for the casual user. Daily data would be of considerable use, or better still real time data, even if only for a limited number of grid points were made available in the public domain. If real time data cannot be released then a compromise is needed to show the range of forecast products that are available. If forecasts are provided to the user, they must also be provided with the model climate, the predictability, forecast quality, for example a reproducibility index (Stern and Miyakoda), and the skill of the ensemble compared to the model climate.

Feedback is necessary from users to the forecasting community to develop tailored products, such as histograms of break cycles and the fact that users are generally more interested in intraseasonal variability instead of monthly totals. User development and capacity building is necessary and could be directed through the CLIVAR regional panels. A link needs to be made with decision makers, with information needing to be made more publicly available.

The application of downscaling is being advanced through the ENSEMBLES project, in particular by the work of A. Cofino and J.-M. Gutierrez from the University of Cantabria, Spain. Downscaling data are available for the user community, subject to registration. What is less quantified is quality according to the ENSEMBLES user community.

> **ACTION** A white paper is in preparation this summer on seamless range and approaches for user applications (A. Morse).

## 6. Report from the Working Group on Numerical Experimentation (WGNE) Workshop on Systematic Errors in Climate and NWP Models

M. Déqué reported on the WGNE Workshop held in February 12-16 2007 in San Francisco that addressed how to identify and possibly correct biases in atmospheric and coupled ocean-atmosphere models. The scope of the meeting was to:

Item 1: examine systematic errors in operational NWP models
Item 2: investigate the potential of systematic error reduction in seasonal forecasting
Item 3: evaluate the coupled AOGCMs in the IPCC-AR4 exercise

The Tropical Pacific received particular attention in terms of the first two bullet points, though issues concerning the third bullet point dominated the meeting.

In numerical seasonal forecasting, refinements in models sometimes lead to improved climate simulation (bias decrease) and sometimes to improved scores (correlation increase). But it is much more frequent to get a double degradation than to get a double improvement.

The following approaches were presented at the meeting:

- enhancing the horizontal resolution:
- reducing SST bias via two-tier forecast
- reducing atmospheric bias by carefully evaluating initial tendency errors.

In terms of Item 1, improvements in some aspects of the bias or some scores are obtained, but degradations are observed too. With regards to Items 2 and 3, the scores are marginally affected, although the biases are significantly reduced.

-

The following recommendations were made:

- IPCC modellers should develop, or ‘tune’, the atmospheric and oceanic components separately.
- The first tests in coupled mode should use a validation technique close to what is used for seasonal forecasts with short runs. The aim is not for the best skill score, as a large ensemble would be necessary, but to evaluate skill in simulating coupled processes.

These recommendations aim to avoid the approach of taking the coupled model as the evaluation starting point and to add more stringent tests of model skill instead of focusing on model drift and the mean climate. Assessing the model in seasonal forecast mode is more objective than just timeseries analysis. See item 4.3 for WGSIP discussion on this topic.

## 7. Reports from CLIVAR regional Panels

### 7.1. Pacific Implementation Panel (PIP)

M. Balmaseda gave an overview of the main topics of focus for the PIP:

- understanding and predicting ENSO
- developing ENSO metrics of scientific and societal relevance
- ENSO sensitivity to climate change
- ENSO and Westerly Wind Burst (WWB) interactions
- understanding the South Pacific Convergence Zone (SPCZ)
- improving model biases in the Tropical Pacific
- understanding inter-basin connections on interannual to multidecadal timescales

An integral part of the panel work is to assess the observational requirements to address these issues. The panel is involved in the development of the Southwest Pacific Ocean Circulation and Climate Experiment (SPICE), whose goal is to observe, model and understand the role of the Southwest Pacific Ocean in large-scale decadal climate ENSO modulation, the Tasman Sea area and the generation of local climate signatures. More information is available at www.ird.nc/UR65/SPICE.

M. Balmaseda also presented a proposal for a Pacific Ocean modelling experiment that would span interests across the PIP, the CLIVAR Working Group on Ocean Model Development (WGOMD) as well as WGSIP. The WGOMD is developing a Coordinated Ocean Reference Experiment (CORE-II) forcing ocean/ocean-ice models with the Large and Yeager (2004) interannual forcing fields for the period 1950-2004. The PIP is interested in this experiment as it will demonstrate the ability of ocean models to represent aspects of the Pacific mean state and variability. The PIP proposal aims to complement CORE-II by running additional integrations to explore the sensitivity of the models to the forcing fields, different resolution (horizontal and vertical), bulk formulae and parameterisations. The aim of these sensitivity studies is to determine what are the dominant sources of errors that currently lead to well known deficiencies in modelling the Pacific Ocean such as the incorrect representation of upwelling off the South American coast, the westward penetration of the Cold Tongue, the east-west gradient of the Equatorial thermocline and the Equatorial heat content.

Modelling groups would be invited to participate and run the sensitivity experiments of their choice from the suite presented in this proposal. WGSIP supported the extension of the CORE-II experiments proposed here.

> **ACTION** WGSIP recommends that the Pacific Implementation Panel coordinates ocean model error analysis experiments and sensitivity of models to resolution, bulk formulae etc, to include CORE forcings and to report back to WGSIP (M. Balmaseda, representing PIP).

## 7.2. Asian-Australian Monsoon Panel (AAMP)

The AAMP has expressed interest in SMIP-2 data and some initial results were presented at the WCRP Seasonal Prediction Workshop in Barcelona. The analysis of SMIP data has been somewhat limited and should be further encouraged by WGSIP.

> **ACTION** Encourage AAMP to make use of SMIP and CHFP data, and to do additional tier-1 (uncoupled) and tier-2 (coupled) experiments in monsoon region as required.

## 7.3. Variability of the American Monsoon System (VAMOS) Panel

P. Nobre, representing the VAMOS panel, outlined the main science themes according to the modelling component of VAMOS. These are the role of the diurnal cycle, the role of the surrounding oceans, the lifecycle of the monsoon and the role of intraseasonal variability. These themes will be addressed using available seasonal prediction data and results from the CHFP Experiment.

## 7.4. Variability of the African Climate (VACS) Panel

VACS organised a training workshop on Southern and Eastern African Climate Predictability that was hosted on the 10-13 July 2006 by the Tanzanian Meteorological Agency. Scientists from each National Meteorological Service in southern and East Africa and from the operational ocean agencies in the region, were trained in the use of the IRI Climate Predictability Tool (CPT). A follow on to this event will be a workshop to further disseminate training in using the CPT so that each African sub-region has a CPT expert. The workshop participants also provided input into CLIVAR research programmes of relevance for this region, for example SAGRADEX.

W. Landman presented the SAGRADEX project, which is being proposed to study the role of the strong surface gradients in topography, rainfall, soil moisture, vegetation in southern Africa, as well as in SST and winds in the neighbouring oceans, in determining the climate, variability and predictability of this region. It is envisaged that this project will involve the analysis of currently available data as well as targeted model and field studies. The project will feed directly into seasonal to interannual prediction efforts and the associated societal impacts over southern Africa. Models currently have difficulty in adequately resolving these gradients, so hampering the understanding of regional climate variability, seasonal forecasting and climate prediction efforts and in assessing the likely impacts of climate change over the southern African region. Funding has been obtained from the UK Royal Society and the South African National Research Foundation to hold a workshop to develop a White Paper to present to funding agencies.

W. Landman also informed WGSIP that fully coupled model forecasts are being added, in collaboration with the UK Met Office, to the South African Weather Service (SAWS) multi-model system.

> **ACTION** Encourage VACS to use data from CHFP Experiment. SAWS could provide output from multi model output that is planned at SAWS to DMC in exchange for rainfall data. Propose that this becomes a VACS project that can be endorsed by WGSIP (W. Landman).

# 8. Review of national and regional activities

## 8.1. Current operations at CPTEC (P. Nobre)

CPTEC currently issues monthly seasonal climate forecasts for South America. The forecasts are based on an ensemble of 90 members of CPTEC AGCM and 10 members of CPTEC CGCM, described below. Also they use 5 members of the Eta regional model over South America.

Forecast variables are three month average precipitation and air temperature for months 2-4 of the forecast runs listed below. Atmospheric initial conditions (IC) are one day apart NCEP analyses from the previous month (e.g. December IC to predict Feb-Mar-Apr). Global Tropics SSTA are also forecast with the CPTEC coupled model up to eight months ahead of the IC. Ocean initial conditions are obtained from forced OGCM (MOM3) runs without ocean data assimilation. On extended weather forecasting, CPTEC runs its CGCM (T126L28-RAS AGCM; 1/4th degree deep tropics, 65S-65N, L20 OGCM) up to 30 days, with two members daily.

Seasonal Climate Prediction at CPTEC:

AGCM runs (T062L28):

- 15 members with KUO convection scheme and persisted SSTA
- 15 members with KUO convection scheme and predicted SSTA
- 15 members with GRELL convection scheme and persisted SSTA
- 15 members with GRELL convection scheme and predicted SSTA
- 15 members with RAS convection scheme and persisted SSTA
- 15 members with RAS convection scheme and predicted SSTA

Coupled Ocean-Atmosphere GCM (T062L28 Atmos - 1/4th degree deep tropics Ocean)

- 10 members

Regional Eta Model over South America

- 5 members, nested on CPTEC AGCM (KUO convection, persisted SSTA)

Statistical prediction models:

CCA forecast models of Tropical Atlantic SSTA and Nordeste and South Brazil seasonal rainfall.

Climate Monitoring at CPTEC:

- daily rainfall, max & min air temperature over South America
- daily OLR and SST globally
- pentad positioning of the Atlantic ITCZ
- monthly global scale atmos. and ocean fields

### 8.2. Progress with the POAMA/ACCESS SP System at BMRC (G. Wang)

POAMA is the current operational dynamical seasonal prediction system at the Bureau of Meteorology. The system was developed jointly with CSIRO and Land and Water Australia.

A new version, POAMA-1.5, has been built and is undergoing operational trials. POAMA-1.5 has several enhancements, the most significant of which is a new Atmosphere/Land Initialisation system. One of the objectives was to produce a system that when run in real-time would have the same configuration as in hind-cast mode, particularly in relation to the initialisation of the ocean, land and atmosphere. For the first time it will provide rainfall forecasts in addition to El Nino forecasts. A comprehensive set of 10-member monthly ensembles has been completed covering the period 1980-2006. Plots of hind-casts and real-time forecasts are available on the POAMA experimental web site at:

http://poama.bom.gov.au/experimental/poama15/r_gen.htm

Preliminary investigation of skill from POAMA-1.5 shows very promising results. In terms of SST the new system is significantly better than the old one in both the Pacific and Indian Oceans. Regional predictions of rainfall and temperature over Australia also show skill levels comparable to or better than statistical systems, particularly at short lead-times.

Because the system is initialised in real-time it also doubles up as an intra-seasonal prediction system. A suite of products on intra-seasonal time scales have been developed, including regional rainfall and temperature, and an MJO index.

The hindcasts from POAMA are freely available for research purposes on an open-Dap server. More information is available at http://poama.bom.gov.au.

A new version, POAMA-2, is also under development. The most significant improvement is a new ocean data assimilation system based on the ensemble Kalman filter. A 30-year ocean re-analyses is underway and some preliminary hindcasts have been performed.

### 8.3 Development of high resolution models and experiments at the MRI and the FRCGC (T.Ose)

Within MRI, the tropical cyclones simulated in an AMIP experiment with the 20km-mesh AGCM have been compared with the observed ones in the aspects of the inter-annual variability and long-term trend of occurrence of tropical cyclones.

At the FRCGC, it has been shown that the 3.5km-mesh NICAM (Nonhydrostatic ICosahedral Atmospheric Model) has the ability to make a nearly one-week forecast of tropical clouds or cyclones by comparison with infrared satellite images (Miura *et al.,* 2007, *GRL*, 34, L02804, doi:10.1029/2006GL027448). The climate sensitivity of the NICAM was also examined (Iga *et al*., 2007, *GRL*, doi:10.1029/2007GL031048, in press).

### 8.4. Recent progress in seasonal to interannual prediction at IRI (D. DeWitt)

The International Research Institute for Climate and Society (IRI) has recently been working to generate retrospective forecasts for the atmospheric general circulation models (AGCMs) used in their real-time multi-model ensemble forecasting system. Such retrospective forecasts are particularly important in the context of evaluation of the seasonal forecast skill when applied to sectoral climate risk models such as disease vectors, water resource management and agricultural applications.

They have also been conducting experiments with higher resolution versions of one of the AGCMs used in their multi-model ensemble forecast system so that they can evaluate the benefit of moving to higher resolution when weighed against the increased computing cost. Finally, work is being conducted on the methodology for combining models in the IRI multi-model ensemble and the selection of sea surface temperature (SST) scenarios for use in this.

### 8.5. Recent and planned NCEP climate modelling activities (H. -L. Pan)

NCEP's Environmental Modelling Center (EMC) and Climate Prediction Center (CPC) will produce a reanalysis of the atmosphere, the ocean, and the land for the satellite era (1979 – 2009) using a high-resolution, state of the art, data assimilation system of the atmosphere (the Gridpoint Statistical Interpolation analysis system), the ocean (the Global Ocean Data Assimilation System), and the land (the Global Land Data Assimilation System). The use of satellite radiance as well as a fully coupled model to provide first guess fields for atmosphere, ocean, and land are the major advances from previous NCEP reanalysis efforts (the NCEP/NCAR reanalysis and the DOE/NCEP reanalysis). This reanalysis effort is being developed to provide initial conditions for the next major upgrade of the NCEP Climate Forecast System by providing initial conditions for a complete reforecast to provide calibration and downscaling for monthly and seasonal predictions. Plans are underway to make the products available to the weather and climate research scientists of the world.

### 8.6. CCSM3.0 Prediction Experiments: Testing an IPCC class model for seasonal to interannual forecasting (B. Kirtman)

To assess the potential predictive skill of the CCSM3.0, a large sample of retrospective forecast experiments have been made and compared to available observations. The retrospective forecasts cover the period 1982–1998. A 12-month hindcast is initialized each 1 January and 1 July during this 17-yr period. For each initial month, an ensemble of six hindcasts is run, yielding a total of 204 retrospective forecasts to be verified. The hindcast ensembles are generated by atmospheric perturbations only and no attempt has been made to find optimal perturbations. The ocean initial state for each ensemble member is identical. We acknowledge that with this approach we may underestimate the uncertainty in any individual forecast. We emphasize that these particular hindcasts were designed as a “proof of concept” in terms of developing a national multi-model prediction system.

To form a multi-model ensemble, we use the six CCSM3.0 members initialized on 1 January for each year 1982-1998, and for CFS we use the five ensemble members initialized on 30 December – 3 January for each year 1982-1998. The correlation coefficient for each model (we have chosen not to identify which model is which) and the 11-member multi-model ensemble was shown. The systematic error for each model is calculated in the same way as is based on the limited sample from 1982-1998. In calculating the correlation coefficients we use the ensemble means. There are several points to note:

(i) The multi-model ensemble mean (black curve) has the highest correlation for most lead times;
(ii) The multi-model correlation is higher than simply averaging the correlation from the two different models;
(iii) Most notably the large drop in skill for Model A for lead times 4-6 has only a small impact on the multi-model skill;
(iv) Based on 11-member sub-sampling of the CFS data (not shown), the overall multi-model improvement is better than a same-sized ensemble from a single model (this is consistent with the results from the DEMETER project).

This suggests that the correlation coefficient for the multi-model ensemble is generally higher than either model alone, although we need to use larger ensembles, more forecast cases and ensure consist use of lead-time. Moreover, as lead-time increases the multi-model ensemble has a larger impact on the correlation. Similar results are found with the root mean square error. These results are quite encouraging in terms of developing a US national multi-model ensemble prediction system.

### 8.7. Canadian research efforts in prediction and predictability of the global atmosphere-ocean system from days to decades (G. Boer)

A new "Global Ocean-Atmosphere Prediction and Predictability (GOAPP)" Network (http://www.goapp.ca) has been initiated comprising a loose amalgamation of 16 co-investigators from 9 Universities, and from Environment Canada (EC) and the Department of Fisheries and Oceans (DFO). The broad aim of GOAPP is to improve predictions of the ocean and atmosphere on time and space scales of days to decades, and tens of km to global scales through:

- Enhanced modelling and data assimilation capabilities for the ocean and the coupled system;
- A new appreciation of the sources of, and limits to, predictability in the coupled system;
- A better understanding of the physical processes underlying seasonal and longer time-scale predictions made with coupled models;
- Decadal-scale reanalyses of the North Atlantic and North Pacific oceans and a coupled historical forecasting project;
- Analysis techniques and tools for maximizing the utility, and assessing the value, of forecasts.

The GOAPP Research Network complements an incipient governmental operational environmental prediction system involving EC, DFO and DND (Department of National Defence) consisting of three inter-related tracks: (i) an operational track with a 'fast start' achieved by importing the Mercator ocean model and data assimilation from France and coupling it with the Global Environmental Multi-scale (GEM) atmospheric model of EC; (ii) a research and
development track consisting of long-term government research and complementary academic research, including the Network mentioned above; and (iii) a products track that will identify, develop and disseminate relevant products and outputs.

The intent is that by linking of the research Network with the incipient Canadian operational initiative, and ongoing seasonal forecasting activities and climate change projections, the scientific and methodological advances produced will be quickly used to improve forecasts and generate new products.

## 9. WGSIP business and future activities

**ACTION** Membership actions to be discussed by email (T. Stockdale and B. Kirtman).

**ACTION** Consider organising the next meeting possibly in Salvador, Brazil, in early 2009 (T. Stockdale, B. Kirtman).

## Appendix A: List of Attendees

***Panel Members***
Ben Kirtman
Willem Landman
Pablo Nobre
Michel Deque
George Boer
Andy Morse
Dave DeWitt
Tim Stockdale
Chung-Kyu Park
T. Ose
Hualu Pan

***Guests and Observers***
Guomin Wang
Franco Molteni
Eric Guilyardi
Bill Merryfield
Ann Henderson Sellers
Jagadish Shukla
Tim Palmer
Jim Kinter
Ken Mooney
Magdalena Alonso Balmaseda

***ICPO***
Anna Pirani

***Panel Members not present***
Scott Power
Chris Gordon
Randy Koster

## Appendix B: Agenda of the 11th WGSIP Session, Barcelona Spain, 7-8 June 2007

**Thursday 7 June 2007 1.45 - 5.45 pm**

1. **Welcome and review of Agenda** (T. Stockdale and Ben Kirtman (co-chairs, WGSIP))

2. **Brief review of relevant CLIVAR and WCRP activities (pre-circulated)**
2.1. Report from the CLIVAR IPO (Anna Pirani)
2.2. Reports from the CLIVAR SSG (April 2006) and the WCRP JSC (March 2007)
2.3. Reports from CLIVAR regional panels: (VAMOS, AAMON, VACS, Pacific Panel, Atlantic Panel, Indian Ocean Basin Panel)
2.4. Reports from JSC/CLIVAR Working Group on Coupled Modelling (WGCM) and WGOMD

3. **New and proposed WGSIP collaborations**
3.1. Decadal prediction experiments, together with WGCM, as proposed by the JSC. (Tim Palmer to introduce)
3.2. Definition of reference ocean spin-up for use in basic seasonal forecast experiments (collaboration with Pacific Panel and WGOMD) (M. Balmaseda, Pacific Panel)
3.3. Definition of recommended protocol for testing coupled GCMs in seasonal forecast mode (collaboration with WGNE and WGCM, and using the reference spin-up defined in 3.1).
3.4. Diagnostics and data handling for collaborations 3.2 and 3.3.

4. **Ongoing WGSIP activities (check which of these are still relevant)**
4.1. WGSIP Standard hindcast and verification system (T. Ose)
4.2. Model experimentation and outputs standards project (T. Stockdale)
4.3. Expert Team for Long Range Forecast Verification (S. Power)
4.4. AA-Monsoon Collaboration (???)
4.5. VAMOS Modelling Collaboration (P. Nobre and B. Kirtman)
4.6. Pacific Panel Collaboration (S. Power)
4.7. Interactions with GEWEX (R. Koster)

**Friday 8 June 2007 9.00am - 5.00pm**

WCRP Model Errors Workshop Report (M. Déqué)

5. **Review of Seasonal Prediction Workshop and TFSP experiments**
5.1. Review of workshop and its recommendations (B. Kirtman)
5.2. Progress report from TFSP data committee (D. DeWitt, T. Stockdale, G. Boer)
5.3. Discussion and decisions on data issues
5.4. Discussion and decisions on additional experimental protocols

6. **Review of national and regional activities**
6.1. Relevant reports from regional or national CLIVAR Committees (e.g., US CLIVAR).
6.2. Update on multi-model prediction studies and operational systems (US, Europe, Asia)
6.3. Developments at national institutes and elsewhere (opportunity for all to present)

7. **Action items and organization of future activities** (T. Stockdale and B. Kirtman).
7.1. Review of action items from last meeting
7.2. Future WGSIP work related to applications (A. Morse)
7.3. Agreement on overall work plan, including new topics and continuation of existing activities
7.4. Action items for the coming year
7.5. Membership changes
7.6. Possible dates and places for next WGSIP session.

## Appendix C: Summary of the Position Paper produced by the participants of the First WCRP Seasonal Prediction Workshop

### The First WCRP Seasonal Prediction Workshop

*Prepared by the Workshop Participants*

*Coordinated by B. Kirtman and A. Pirani*

This article describes 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 applications experts. The article also summarises the WCRP Position Paper on Seasonal Prediction that is currently being prepared by the Workshop participants. The main purposes of the Workshop were to describe the current status and main limitations regarding seasonal forecast skills 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.

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 live 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 how the forecasts can be used in order to make useful improvements to forecast products.

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 the Task Force on Seasonal Prediction (TFSP) to assess 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.

As part of the seasonal prediction capability assessment, the TFSP in collaboration with the core programs of the WCRP (CLIVAR, CliC, SPARC and GEWEX) and the World Climate Programme (WCP) organized the First WCRP Seasonal Prediction Workshop, which was held June 4-7 2007 in Barcelona Spain. This article summarises the key outcomes and recommendations of this workshop. The WCRP Position Paper on Seasonal Forecasting that is being prepared by the Workshop participants, is intended to go beyond merely summarizing the workshop presentations; indeed we specifically avoid this sort of summary. The main purpose 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 our ability to use 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 brought together the diverse seasonal prediction community. This includes researchers of the physical climate system and forecast methodology, operational forecast providers and forecast applications 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 funding agencies were in attendance.

This summary of the workshop is organized as follows:

(1) We present some critical common language regarding the assessment of seasonal prediction. Developing a common language for assessing seasonal prediction is critical to successful interaction among forecast providers, forecast users and forecast researchers, and the importance of this agreed upon language cannot be over stated.

(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, and carry the full weight of the seasonal prediction community.

(3) The workshop identified a simple set of metrics for assessing seasonal prediction quality, which provide a key benchmark of evaluating future improvements. These metrics are presented here; however, the workshop participants recognized the

importance of allowing these metrics to be refined over time. In fact, the workshop report is a "living" document that will be refined and updated in the future and made available online to the entire community.

**A Common Language for Assessing Seasonal Prediction**

The need for an authoritative statement on the skill of seasonal, and other extended-range, predictions has long been recognised. Several audiences for such a statement exist, including development scientists, 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 development scientists and users, although others will gain benefit.

The term 'skill' covers a complex array of issues; just two will be covered here: *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 amongst other information, and while of fundamental concern to the user should also stimulate scientists.

**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.

(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 procedure for assessing seasonal prediction skill is critical for documenting future improvement.** 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. The Position Paper 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 the Position Paper, 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 TFSP experimental design. 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; however, international and nation coordination and commitment is seriously lacking.

(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 applications models.** 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.

(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 forecasts.

(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.

(10) **The research community has not adequately quantified what is and what is not predictable.** Indeed, the Position Paper, as it evolves over time should include, where possible, unambiguous statements regarding what is predictable and what is not predictable.

## Assessing Seasonal Prediction Quality

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 (i.e., 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 source 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 two metrics for an overarching assessment the quality of seasonal prediction. Here we summarize results from one of these metrics. It was clearly acknowledged that these metrics were not sufficient for use in applications or to improve the quality of the forecasts. These are the multi-model Brier Skill Score, described in more detail below, and the quality of the predicted SST in the Eastern Pacific (i.e. the Nino3.4 region). Much more detailed information is required, but well beyond the scope of the Position Paper. 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.

*Multi-model Bier Skill Score (BSS)*
This 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). This particular forecast quality metric is discussed in detail in a companion paper in this volume (see Palmer et al, 2007) and in the peer review literature (see Palmer et al. 2008). 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 BSS based on DEMETER data (Palmer *et al., 2004*) is summarized in Table 1. 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 reliable forecasts and underlined values indicate greater than 90% confidence in the reliability. Negative underlined values indicate that the multi-model ensemble reliably 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 (i.e., Central America, Amazon Basin, Western Africa), although there are sub-tropical regions of considerable forecast quality (i.e., Tibet). While some regions can be reliably predicted in both JJA and DJF, there is significant seasonality in 2m temperature forecast quality.

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 reliability. The fact that forecast quality can be improved using these techniques indicates that models and predictions can and should be improved.

**Conclusion**

The workshop, the follow-on WCRP Position Paper on Seasonal Prediction, and the TFSP prediction experiments (http://www.clivar.org/organization/wgsip/tfsp.php) represent the necessary steps in a comprehensive (quality and value) seasonal prediction assessment. The consensus statements that are part of the WCRP seasonal prediction Position Paper carry the full weight of the entire seasonal prediction community. The consensus statements were discussed and vetted in great detail at the workshop and have been made available for comment from scientists and researchers who were unable to attend the workshop. Indeed, the feedback from both the workshop participants and those unable to attend has lead to significant refinements.

As noted above, the workshop participants felt very strongly that the WCRP seasonal prediction Position Paper be a living document, not only to embellish upon the assessment metrics, but also to clearly document future seasonal prediction improvements. In particular, the TFSP experiments require a much more “comprehensive” view of seasonal prediction. By comprehensive, we mean that the TFSP is hypothesizing that there will be substantive improvements in seasonal prediction if and only if we include all the interactions among the components of the climate system. This WCRP seasonal prediction paper necessarily needs to document the results of the TFSP experiments that should become available in late 2009.

**References**

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

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

Palmer, T. N., F. J. Doblas-Reyes, A. Weisheimer, and M. J. Rodwell, 2007: Seasonal Forecast Datasets – A resource for Calibrating Regional Climate Change Projections? *CLIVAR Exchanges*, **43**, 6-7.

Palmer, T. N., F., J. Doblas-Reyes, A. Weisheimer, and M. J. Rodwell, 2008: Towards Seamless Prediction: Calibration of Climate Change Projections Using Seasonal Forecasts. *Bull. Amer. Meteor. Soc., accepted.*

| 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).

## Appendix D: Proposed CHFP Data Handling Strategy (T. Stockdale)

C.1. Analysis of general seasonal prediction data needs: Data formats and grids for international exchange

There are two separate sets of requirements for supplying multi-model seasonal forecast data.

1. Exchange of data between forecast producers and data centres.

- intelligible data formats (GRIB or netCDF)
- minimal loss of data (original grid where possible)
- data to be readily processible (rectangular lat/long or Gaussian grids only)
- automatic real-time data transfer possible (e.g. remote FTP)

2. Supply of data to "user" community

- supply of multi-model data on a common grid
- supply of multi-model data in single file
- suggested standard grid of 2.5 by 2.5
- possibility to request data on different lat/long grids (e.g. higher resolution)
- possibility to request regional data (subsetting)
- openDAP interface
- netCDF format

At the moment, and for the immediate future, there is no single software solution that meets both of these sets of needs. In particular, software which can serve data on any requested grid, including the original, through an aggregated openDAP server interface to allow multiple models in a single file, and which also allows remote batch mode access, does not exist.

Thus in the short term, we recommend a twin track approach.

1. For data producers, data should be made available via any commonly used and batch-accessible method (e.g. FTP). Data should be on the original grid (if lat-long or gaussian), or else converted to an appropriate regular grid if the original grid is too irregular (e.g. for ocean models). Data format should either be GRIB or netCDF. We should work towards standardized metadata, filename conventions etc; this is not critical if only one or two organizations exchange data, but will become more important if the number of data producers grow.

2. For users, data should be available via openDAP on a standard 2.5 deg grid, for manual download, and using an aggregation server. A multi-model forecast can be downloaded as a single file. In the future, and/or as software allows, users should be able to access data on the original grid or any other grid of their choice.

C.2. Proposed technical details for serving CHFP data

1. Data should be made available in netCDF
2. CF compliance should be observed to the extent possible
3. Atmosphere data should be available on a 2.5 deg grid (plus, optionally, on the original model grid)

4. Ocean data should be available on a 1 deg Levitus grid.

5. If possible, data should be served using an openDAP aggregation server such as THREDDS, following the example of the ENSEMBLES data served at ECMWF.

http://www.ecmwf.int/research/EU_projects/ENSEMBLES/data/data_dissemination.html

Here the data from different models is aggregated into a single multi-model ensemble data structure. The individual non-aggregated netCDF files can also be accessed.

(At the time of writing the aggregated data includes 9-member ensembles from four different models, giving 36 forecast members in total. The data arrays "realization", "experiment_id", "source" and "institution" are vectors which return information on each forecast in the ensemble, saying where it came from etc. There was still a small bug affecting some of the contents of this metadata, but it should all be fixed by mid-April).

6. Feedback on the ECMWF data server is welcome. In particular we know that we need to improve the user documentation to make this "user friendly" - the effort so far has been on the overall design and technical implementation. [Improvement has since been made]

7. For reference, approximate data volumes are estimated as follows:

- 1 atmos field = 144*71 pts * 4 bytes = 41 kbytes
- 1 ocean field = 360*180 pts * 4 bytes = 259 kbytes

- Pressure level: 5*(3-6)* 215 days = 3225-6450 fields / integration
- Surface: 40 fields * 215 days = 8600 /integration (approx)
- Ocean monthly: 90 * 7m = 630/integration
- Ocean daily: 7 * 215 = 1505/integration (daily averaged fluxes) (5375 with 6h fluxes)

- Level 1 : 1000 integrations -> 1 - 2 TBytes per model
- Level 2a: 2000 integrations -> 2 - 4 Tbytes per model (on top of level 1)
- Level 2b: 800 integration -> 0.8 - 1.6 Tbytes per model (on top of level 1)

The biggest factor influencing data volumes is the ocean data, particularly the surface fluxes. If these are provided as daily averages (instead of 6h), it cuts the total data requirement in half. At ECMWF, because our coupling is made with 24h averaged fluxes, we only archive the 24h averaged fluxes in any case.

If atmosphere data is provided on the original grid as well as the 2.5 deg grid, the data volumes will increase, particularly so for higher resolution models.

**APPENDIX E**

## DRAFT PROPOSAL

### WGSIP Standard Hindcast and Verification System (for Modellers)

M. Sugi and T. Ose

Meteorological Research Institute, Tsukuba, Japan

The most remarkable progress in SI forecast research during the last decade may be the development of multi-model ensemble (MME) forecast systems. The MME forecast system experiment projects such as SMIP, SMIP2/HFP, PROVOST, DEMETER, APCN have been conducted successfully with many fruitful results. These projects have demonstrated significant skill improvement by the MME method, the importance of systematic hindcasts, the need for demonstration of the value of SI forecast applications, the need for objective evaluation of the individual model's skill and for skill improvement, and so on.

In order to make further progress in SI forecast research, we should consider the following three steps of the SI forecast and its application:

1) To develop good models for global dynamical forecasts,
2) To develop optimum post processing, including calibration, MME and downscaling,
3) To develop application technology to meet the end user's needs.

A key to all of these three steps may be the establishment of a Standard Hindcast and Verification System. To improve a model, modellers need to know the skill of their model and the advantages and disadvantages of their model compared with other models. For model skill inter-comparison, a standard hindcast and verification system is indispensable. To develop an optimum post processing, of course, a hindcast and verification system is also needed. By employing a standard system, many different models can be incorporated to develop optimum post processing. To develop application technology, we first need to demonstrate the value of the SI forecast information for users based on a standard hindcast and verification system. Users need to know which models and which MME system is most useful for their own application.

The standard hindcast and verification system should play a major role in developing common understanding of the skill of SI forecasts among modellers and users. It should be a communication tool between them. For this purpose, the standard hindcast and verification system has to be as simple as possible. As a measure of the skill, we propose to use the anomaly correlation coefficient (ACC) for surface temperature and precipitation at standard global grid points. There are a number of reasons to use ACC as measure of skill in the system:

1) Temporal correlation is the most fundamental skill measure for SI forecasts. The square of correlation is the ratio of forecast signal variance to the total variance. It shows how much of the total variance is explained by the forecast.
2) Correlation is invariant in linear transform. A basic post processing is a linear calibration (bias correction and amplitude correction, or linear regression) and correlation is invariant in the linear calibration. It represents an intrinsic skill of the

model. It should be noted that a linear calibration is equivalent to minimizing the RMSE and it also can have an effect of variance inflation for probability forecast.

3) Correlation is widely used and most users are familiar with it and it is easy to understand. The users can easily compare the model's skill with conventional statistical forecast skill at the user's point, and evaluate how useful the model forecast is.
4) In order to evaluate the value of probability forecast for a particular user's application, a more sophisticated skill measure may be needed. However it should be noted that the SI forecast skill basically originates from the skill of the forecast of the predictable signal in the SI variation, and therefore, generally the higher the correlation skill, the higher the probability forecast skill.

**WGSIP Standard Hindcast and Verification System**

Variable: Tsfc(T2m) +SST, Precipitation

Verification Data: JRA-25, ERA-40, NCEP/DOE R2;
HadiSST, Raynolds SST, COBE SST
GPCP, CMAP

Verification Period: 1981-2005 (25 years) - This should be extended to 1981-2010(30 years) in the future.

Verification Score: temporal anomaly correlation (ACC) on a 2.5° x 2.5° global lat-lon Grid

Standard Hindcast Output: monthly mean ( +seasonal mean) of forecast variables of each member

National Oceanography Centre, Southampton
University of Southampton Waterfront Campus
European Way, Southampton SO14 3ZH
United Kingdom
Tel: +44 (0) 23 8059 6777
Fax: +44 (0) 23 8059 6204
Email: icpo@noc.soton.ac.uk