WORLD METEOROLOGICAL ORGANIZATION

COMMISSION FOR BASIC SYSTEMS

OPAG ON
DATA PROCESSING AND FORECASTING SYSTEMS

MEETING OF THE

JOINT EXPERT TEAMS ON LONG-RANGE FORECASTING

(INFRASTRUCTURE AND VERIFICATION)

ECMWF, 3-7 APRIL 2006

FINAL REPORT

EXECUTIVE SUMMARY

The meeting of the Joint Expert Teams on Long-Range Forecasting (Infrastructure and Verification) of the CBS OPAG on DPFS was held at the kind invitation of ECMWF in Shinfield Park, Reading from 3 to 7 April 2006. Fifteen experts representing eleven operational Long-range Forecast producers and one regional centre participated in the meeting. Dr Dominique Marbouty, Director of ECMWF opened the meeting and welcomed the participants. The Director stressed that ECMWF was fully operational in Long Range Forecasting, and was pleased to host this meeting, where experts from other producers of long-range forecasts will be able to share ideas and contribute to the building of international infrastructure necessary for LRF. ECMWF was pleased to make available to WMO Member states its LRF products, which can be accessed via Internet.

The Meeting reviewed the scientific benefits of the development of multi-model ensembles (MMEs) for long-range forecasting and considered the status of MME centres. The Team was given presentations of the EURO-SIP, ENSEMBLES and APEC Climate Centre projects. Considering the results of the experiments with multi-model ensembles, the Team agreed that:

- MMEs provide the opportunity for improved reliability over that available from single model ensembles alone.
- MMEs provide the opportunity to estimate uncertainties in LRF, and to particularly identify limitations of LRF.
- MMEs provide a means to a "confidence builder" in the area of LRF.

The Team agreed that there were needs for Lead Centre(s) for collection of globally available LRF to build MMEs. The Team stressed that Lead Centre(s) would remain at the same level of status as all the other GPCs. The Team appreciated the offer by KMA (Korea Meteorological Administration), possibly joint with NCEP (USA), to become Lead Centre for LRF MMEs. The Team agreed to a set of functions for a Lead Centre for LRF MMEs

Larger improvements in skill can be achieved from the use of multi-model ensembles, rather than increasing the ensemble size for a single model. The Team agreed that the required ensemble size depends on the size of the signal to be detected, for example, in the mid-latitudes, one needs bigger ensembles for detecting small shifts in probability. Skill should really be the key point in deciding the ensemble size. In order to estimate uncertainty, a large ensemble size is desirable. The definition of optimum ensemble size should be decided in the context of the way of ensemble data usage (it is expected to be different for assessment of means, probabilities and extremes).

The Team reviewed the data needs for global LRF, in particular for specific land surface data, like soil moisture and also for ocean sub-surface data. The Team reviewed the statement of guidance, which contained observational data requirements for long-range forecasts and proposed a new text. The team recommend that this list of Observation Data Needs for Producing Global Long – Range Forecasts be considered by the CBS/OPAG on IOS, by CCl and by GCOS.

At the request of some members, the application of WMO Resolution 40 related to LRF products and data was requiring clarification, especially for GPCs' products needed by RCCs and NMCs. In the spirit of Resolution 40, the Team proposed that GPCs' products defined in the Appendix II.6 of the Manual on GDPFS shall be considered as essential and given free without condition to NMCs and RCCs. Other data or products should also be given by GPCs at request of RCCs, NMCs or Lead Centre(s), especially for the purpose of enabling them to perform their tasks, provided the RCCs, NMCs and Lead Centre(s) adhere to the conditions, if any, attached by the GPCs to the data or products.

The Team proposed to define in a simple manner the role of GPCs and RCCs in the Manual on GDPFS. It also reviewed the minimum list of products to be made available by GPCs. The Team kept the substance of the list, but added some necessary clarifications. The Secretariat Representative recalled the requirements to seek recognition as a Global Producing Centre (GPC), as defined by CBS XIII. He informed the Team that a letter recalling the procedures for designation of GPC had been addressed to the Permanent Representatives and at the same time the communication of the Centres' verification data to the Lead Centre for Standard Verification System for Long-Range Forecast (SVSLRF) located in Melbourne was requested.

Dr Andre Kamga from ACMAD gave a presentation on the needs of climate prediction products for the West African region. Products should be "downscaled" to the concerned region. The impact on agriculture and other parties concerned should be taken into account for delivering the right information, by adapting the products to the needs. The forecasts must be available at a certain time that depends on the region, for example March for rainy season in West Africa. Indication on onset, cessation and length of rainy season are needed, but this could be more in the domain of extended range (weeks to month forecasts). Dr Madhavan Nair RAJEEVAN followed by presenting the overall needs of RCCs and NMCs. A general remark on behalf of the users of the products was that GPCs should strive to improve LRF skill, focusing on the

main forecast parameters required by the RCCs. The following experimental products are desired by RCCs and NMCs.:

- o Averages, accumulations or frequencies over 1-month period to 3-month period.
- o Probabilities of exceeding some threshold values (e.g., seasonal rainfall totals above a range of thresholds)
- o Risk of extreme climate anomalies that may help in warning of e.g. occurrence of heat and cold waves over a particular region.
- o Predicted generalized indices of drought, monsoon etc.
- o Dry and wet spells: frequency and duration (with one month lead time)
- o Probable date of onset of main rainy seasons (over a region, like South Asia, East Asia, southern Africa, GHA etc).
- o The need to have first month (0-lead) averages was expressed.

The products are preferred in GRIB 2 format rather than NetCDF, especially for downscaling. The requirements are as follows:

- o Forecast data for downscaling algorithms; this is likely to require more than monthly mean data, e.g.:
    - ▪ Statistics on daily variability
    - ▪ Anomalies for some or all ensemble members
    - ▪ Hindcast data
- o Data for RCM boundary and initial conditions (including SST data).
- o Data for calculating regional specialized indices (drought).
- o Analyzed fields of surface and upper air parameters for use in empirical models as predictors.
- o Observed and predicted global weekly values of SST.
- o Daily satellite precipitation analysis for use in monitoring through the season.

Regional climate centres/ NMCs may not have expertise in all aspects of Long range forecasts. They will need assistance in training from GPCs in the following main areas:

- o Interpretation and use of GPC LRF products
- o Downscaling techniques (both statistical and dynamical)
- o Verification techniques (for local verification of RCC generated products and application outputs)
- o Development of local user applications from RCC downscaled products
- o Use and implementation of regional climate models.

Dr Simon Mason from IRI presented the Climate Predictability Tool (CPT). The CPT is a software package developed by the IRI designed for making downscaled seasonal climate forecasts by RCCs and NMCs.
Dr David Jones on behalf of Dr Andrew Watkins gave an update on the Lead Centre activities. The Team assigned one additional responsibility to the Lead Centre concerning the development of software for graphical display of confidence level information.

The Team discussed matters related to the verification. At the moment the Team prefers to exclude verification of forecasts that were post-processed using for example Model Output Statistics or Perfect Prog approaches but GPCs are encouraged to display verification on their post-processed forecasts on their own web sites. Significance levels can be derived from either standard significance statistical test or bootstrapping techniques. The ET agreed on the general principle that if standard significance tests for a given score are available and valid, given the assumptions about the data, it will be preferable to use them. The ET agreed that the cross-validation should be mandatory for both calibrated and re-calibrated forecasts. There is a need to define a rigorous cross-validation procedure that can be used as part of the standard SVS guidelines. The ET recognized that identifying whether there is a correlation between the accuracy of a forecast and the ensemble spread is not an optimal way of identifying whether there is any information in the ensemble distribution. A more successful approach would involve comparing the quality of the forecasts given the observed ensemble spread / distribution with the quality achieved by keeping the ensemble spread / distribution constant. The ET recognized the inherent difficulty of verifying forecasts of extreme events because of the small sample sizes involved. The only option is to perform verification and to indicate the uncertainty in the calculation of these scores.
It was decided that the recommended period of hindcast should be 1981-2002 for submission to the Lead Centre web site. The ET recommends the GPCP data as the official data set for precipitation verification.

TABLE OF CONTENTS

PAGE

1. OPENING OF THE MEETING

1.1 The meeting of the Joint Expert Teams on Long-Range Forecasting (Infrastructure and Verification) of the CBS OPAG on DPFS was held at the kind invitation of ECMWF in Shinfield Park, Reading from 3 to 7 April 2006.

1.2 Dr Dominique Marbouty, Director of ECMWF opened the meeting and welcomed the participants (see list in Annex to this paragraph). The Director stressed that ECMWF was fully operational in Long Range Forecasting, and was pleased to host this meeting, where experts from other producers of long-range forecasts will be able to share ideas and contribute to the building of international infrastructure necessary for LRF. ECMWF was pleased to make available to WMO Member states its LRF products, which can be accessed via Internet. Within the items to be discussed by this meeting, the Multi Model Ensemble was a very important one and ECMWF will invest a lot in this activity, which was still at an experimental stage. The Verification required also a lot of work, necessary to make a proper use of LRF. ECMWF was also expanding its activities for the monthly forecast. It was also working at a unified model system for all the different ranges. ECMWF would indeed seek to be recognized as an official Global Producing Centre (GPC) for Long-range Forecast.

1.3 The representative of the Secretariat expressed the gratitude and appreciation of WMO to ECMWF for the kind offer to host this joint meeting of the two CBS Expert Teams on the Infrastructure for Long-Range Forecasting and on Standard Verification System for Long Range Forecasts. He also thanked all the staff of ECMWF who have helped with the planning and arrangement for the excellent facilities for this event, especially Dr Laura Ferranti. He welcomed all the experts and the scientists of ECMWF attending this meeting. He stressed that it was most appropriate, in a way, to hold this meeting at ECMWF, since nowadays, the techniques for long-range forecasting have their main roots in the medium-range forecasting, for which ECMWF is a centre of excellence.

During its last session (February/March 2005), the WMO Commission for Basic Systems stressed the need to move forward in the generation, exchange and verification of long-range forecast products, especially among Global Producing Centres of long-range forecasts (GPCs). A workshop of Global Producing Centres of long-range forecasts was held in Jeju Island, the Republic of Korea, last October. The workshop produced a certain number of recommendations, but also generated questions and new requests. The main subjects of this meeting will be: the efficient building and good use of ensemble of ensembles with multi-models, the observations needs for global LRF, the procedures for efficient dissemination and exchange of LRF products, including the official recognition of GPCs and finally the system of verification techniques. An important reason for the existence of the World Weather Watch Global Data Processing and Forecasting System is to ensure that GPCs' products are fully used to provide predictions to WMO Member Countries (that is their National Meteorological Services) to contribute to disaster prevention and mitigation (like severe climatic conditions), and to contribute to better social-economic planning that accounts for variable climatic conditions. That is why the needs of Regional Climate Centres and National Meteorological Centres who are the main users of LRF products, must also be well taken into account. The WMO Representative wished the meeting would be useful for the World Weather Watch Programme and the World Climate Programme and would produce good, thorough and realistic recommendations to be used as "input" for examination by the Implementation Coordination Team on DPFS in June in Geneva, and finally by the next Commission for Basic Systems in November 2006 (planned to be hosted also by the Republic of Korea in Seoul). He thanked all the experts for their contributions, and especially the two chairmen Willem Landman from Pretoria for the Expert Team on Infrastructure for Long Range Forecasting and Normand Gagnon from Montreal for the Expert Team on Standardized Verification System for Long-Range Forecasts, who have a leading role.

2. ORGANIZATION OF THE MEETING

2.1 Adoption of the agenda

2.1.1 The Meeting adopted the agenda given in Annex to this paragraph.

2.2 Other organizational questions

2.2.1 The Meeting agreed on its working hours, mechanism and work schedule. Dr Willem Landman agreed to chair the meeting discussing items 1 to 6, and Mr Normand Gagnon to chair the meeting for discussion on items 7 and 8.

3. PROGRESS ON EXCHANGE OF ENSEMBLES PRODUCTS AND DEVELOPMENT OF MULTI-MODEL ENSEMBLES (MME)

3.1 Status of MME projects:

The Meeting reviewed the scientific benefits of the development of multi-model ensembles (MMEs) for long-range forecasting and considered the status of MME centres.

3.1.1 Towards a multi-model real-time ensemble system for seasonal prediction: the EURO-SIP consortium (ECMWF, Met Office and Météo-France)

Building on the success of DEMETER, a European consortium has been developing and operating a real-time multi-model ensemble forecasting system, as described in Annex to this paragraph by Dr Laura Ferranti and Dr Mike Davey. Forecasts are produced each month. Most of the computation (running the model ensembles, processing and archiving output) is carried out at ECMWF, making use of a common framework.

To meet differing requirements, the consortium members run various post-processing packages. The basic multi-model products use a simple average of the 3 probability distributions associated with each individual ensemble system. For some experimental products hind-cast data are used to adjust the probabilities: currently the Met Office applies discriminant analysis for this purpose. At a later stage Bayesian techniques, such as those developed at ECMWF, will be extended to the multi-model.

3.1.2 The European ENSEMBLES project

The EURO-SIP members are also participating in a European Union Framework 6 project called ENSEMBLES. Dr Francisco Doblas-Reyes presented the project to the meeting. This project contains a substantial effort on seasonal to decadal forecasting that is relevant to multi-model development. As was the case in DEMETER, this project also contains strong links to applications in e.g. the agriculture and health sectors. Downscaling tools are developed for that purpose. The definition of extreme events will depend on the regions, and thresholds identifying events will have to be defined. The experiments also showed the importance of taking into account the greenhouse gases for skilful simulation of the last 40 years.

3.1.3 Status of APEC Climate Center (APCC)

Dr Won-Tae Yun from the Republic of Korea presented the status, goals, functions, structure and future plan of the APEC Climate Center (APCC) (see Annex to this paragraph). The APEC Climate Center is a server, as a hub of regional climate network for realization of concerted and systematic effort to produce a skillful climate prediction and information service. There is open access to APEC Member economies to enhance the capacity building of Member economies in climate prediction and its applications for disaster prevention and sustainable socio-economic growth. The vocation of the APCC, in contrast with GPCs, interfacing with RCCs and NMCs, is the direct interface with various Members Institutions.

3.1.4 Statements on Multi-Model Ensemble (MME) Prediction Systems

Considering the results of the experiments with multi-model ensembles, the Team agreed that:

- MMEs provide the opportunity for improved reliability over that available from single model ensembles alone.
- MMEs provide the opportunity to estimate uncertainties in LRF, and to particularly identify limitations of LRF.
- MMEs provide a means to a “confidence builder” in the area of LRF.

3.2. Lead Center(s) for LRF Multi Model Ensemble Predictions

The results of the sub-groups discussions held during the Workshop of Global Producers of Long Range Forecasts in Jeju Island from 10 to 14 October 2005, related to the need for multi-model ensemble Lead Centre(s), were revisited by the Team.

3.2.1 The Team agreed that there were needs for Lead Centre(s) for collection of globally available LRF to build MMEs. The Team recognized the valuable contributions currently being made by APCC, IRI, ECMWF, and other organizations as prototype Centres for LRF MMEs. The Team agreed that, under the auspices of WMO, procedures and standards for Lead Center(s) for LRF MMEs be established. Data Volume of LRF products is an issue with very large amounts involved (100 terabytes) and there was a need for (a) lead centre(s) for data collection. There was also a need to have the GPC forecasts used correctly by RCCs and RCOFs. The availability of appropriate maps was essential. For this purpose, time resolution of the data sets could be monthly averages and probabilities as defined in the GPCs list of functions. The Team stressed that Lead Centre(s) would remain at the same level of status as all the other GPCs.

The Team appreciated the offer by KMA (Korea Meteorological Administration), possibly joint with NCEP (USA), to become Lead Centre for LRF MMEs.

3.2.2 Definition of functions of Lead Centre for LRF MMEs

The Team agreed to a set of functions for a Lead Centre for LRF MMEs:

- Be a global collector of hind-casts and forecasts from GPCs of LRF and LRF MME, and making them available to GPCs, RCCs, and NMHSs, as registered users (with password protected access);
- Promote the exchange of research and experience on MME, and provide documentation on MME;
- Work, under the auspices of WMO, at the establishment of standards for MME products (e.g. NetCDF, GRIB 2);
- Provide a repository of different MME techniques for the generation of MME in support of GPCs and RCCs;
- Provide display of GPCs forecasts in a common format based on agreed standards, including WMO standards, to RCCs, NMCs and GPCs, with password protected access.

The Team recommended that the Commission for Basic Systems propose the recognition of Lead Centre(s) for LRF MMEs, if performing the required functions.

3.2.3 The Team agreed that the data requirements for the Lead Centre(s) and Multi-model ensemble activities should be discussed further.

3.3 Definition of minimal or optimal ensemble size for practical considerations (for hindcast runs vs. real-time runs, and for nesting limited area models)

The larger, the ensemble size, the better the skill, especially for a single model. Given that, is the notion of optimal ensemble size, the correct one?. An alternative may be to ask the question about the ensemble size with respect to “return of investment,” i.e., beyond a point, skill of a seasonal ensemble prediction system asymptotes, and a larger ensemble results in a smaller gain in skill. Larger improvements in skill can be achieved from the use of multi-model ensembles, rather than increasing the ensemble size for a single model.

The Team agreed that the required ensemble size depends on the size of the signal to be detected, for example, in the mid-latitudes, one needs bigger ensembles for detecting small shifts in probability. Skill should really be the key point in deciding the ensemble size. In order to estimate uncertainty, a large ensemble size is desirable. The definition of optimum ensemble size should be decided in the context of the way of ensemble data usage (it is expected to be different for assessment of means, probabilities and extremes).

The same question is valid for the length of hindcasts. Depending on resources, the Team agreed it is preferable to increase the length of the hindcast rather than the number of ensemble members. The hindcast period should extend back at least to 1981 (see paragraph 8.3.5). For the purpose of statistical downscaling it is recommended that rigorous cross-validation procedures be applied.

4. DATA NEEDS FOR PRODUCING GLOBAL LRF

The Team reviewed the data needs for global LRF, in particular for specific land surface data, like soil moisture and also for ocean sub-surface data. The Team reviewed the statement of guidance, which contained observational data requirements for long-range forecasts and proposed a new text as listed in Annex to this paragraph. The team recommend that this list of Observation Data Needs for Producing Global Long – Range Forecasts be considered by the CBS/OPAG on IOS, by CCl and by GCOS.

5. OPERATIONAL PROCEDURES FOR DISSEMINATION AND EXCHANGE OF PRODUCTS, TERMS AND CONDITIONS

The results of the sub-groups discussions held during the Workshop of Global Producers of Long Range Forecasts in Jeju Island from 10 to 14 October 2005, relating to the operational procedures for dissemination and exchange of products, and their terms and conditions, were reviewed by the Team.

5.1 Resolution 40 clarification

At the request of some members, the application of WMO Resolution 40 related to LRF products and data was requiring clarification, especially for GPCs’ products needed by RCCs and NMCs. In the spirit of Resolution 40, the Team proposed that GPCs’ products defined in the Appendix II.6 of the Manual on GDPFS shall be considered as essential and given free without condition to NMCs and RCCs. Other data or products should also be given by GPCs at request of RCCs, NMCs or Lead Centre(s), especially for the purpose of enabling them to perform their tasks, provided the RCCs, NMCs and Lead Centre(s) adhere to the conditions, if any, attached by the GPCs to the data or products.

5.2 Clarify role of GPCs and RCCs

Obviously, GPCs will continue to conduct further research into improving forecast skill and developing new forecast products (as specified in chapter 6). New products should be developed in close collaboration with scientists from RCCs and NMCs. The Team considered the functions of the RCCs having in mind the interest of the GPCs and of the NMCs, in the view of the forecasters rather than the public or other users external to the GDPFS system. The team was informed that the last CCl recalled the possibility that a Regional Association could choose to establish a Regional Climate Centre under the terms of Volume II of the Manual on the Global Data Processing and Forecasting System (GDPFS) (WMO-No. 485). However, the CCl recognized that some Regions will want to formally designate RCCs through Volume I of the GDPFS and requested that CCl and CBS begin work as soon as possible to undertake the necessary revisions to the Manual on the GDPFS to incorporate new text related to the roles and functions of RCCs and climate prediction. In that spirit, the Team suggested that:

- RCC nominations be accelerated so that GPCs can know to whom products should be provided – RCCs are potentially major users of GPC products.
- Staff of designated RCCs using GPC products should have some minimum prerequisite background knowledge (training, experience, etc.).
- RCCs should be encouraged to express their needs in relation to GPC products.
- GPCs need feedback from users (RCCs), especially on which fields they require in addition to the minimum set for applications modelling (e.g. crop models require daily data, and only certain fields not archived as part of minimum set, for example radiation data). RCCs should provide GPCs with feedback from RCC users.
- Description of attributes and shortcomings of GPC products should be provided (including skill) to RCCs and NMCs.
- RCCs require GPCs to make their products available in digital format.

For the purpose of clarifying the relative roles of GPCs and RCCs, the Team proposed some additions to the Manual on GDPFS to clarify the role of the different types of centre:

Add two notes to 1.4.1.2: (b):

Notes: 1) Centres producing global long-range forecasts, and recognized as such by CBS, are called Global Producing Centres for Long-range forecasts (GPCs). They are not necessarily among the WMC, RSMC or NMC GDPFS centres. The functions of GPCs and the list of official recognized GPCs can found in Appendix II-NN.

2) Regional Climate Centres can be recognized by CBS at request of Regional Associations, if they perform the functions as defined in Appendix II-MM. They are not necessarily among the WMC, RSMC or NMC GDPFS centres.

5.3 Review of Appendix II.6 of the Manual on GDPFS describing the minimum list of LRF products to be made available by GPCs.

The Team reviewed the minimum list of products to be made available by GPCs. It found that requirements were unclear in some places and needed some sharpening up. The Team kept the substance of the list, but added some necessary clarifications. The suggestions are annotated on the Annex to this paragraph in tracked changes.

5.4 Status of GPCs and their nomination

The Secretariat Representative recalled the requirements to seek recognition as a Global Producing Centre (GPC), as defined by CBS XIII. He informed the Team that a letter recalling the procedures for designation of GPC had been addressed to the Permanent Representatives of

Australia, Brazil, Canada, China, France, Japan, Republic of Korea, Russian Federation, South Africa, United Kingdom, USA and the Directors of ECMWF and IRI. At the same time the communication of the Centres' verification data to the Lead Centre for Standard Verification System for Long-Range Forecast (SVSLRF) located in Melbourne was requested. The procedures are recalled below.

The Commission for Basic Systems (CBS), during its Thirteenth Session in St. Petersburg, 23 February to 3 March 2005, noted that significant progress had been made over the last few years in long-range forecasting. The Commission recommended that Global Producing Centres (GPCs) for Long-Range Forecasts (LRF) should be officially designated. That would allow institutions outside the WWW system with demonstrated capabilities in LRF production and services to be officially recognized as such and to make their products available. The Commission agreed that the procedure (defined in Appendix I-2 of the WMO publication no 485, *Manual on the GDPFS*) for broadening the functions of existing RSMCs and for designating new RSMCs, should be applied to the designation of GPCs for LRF. In order to be officially recognized as a GPC, the candidate centre should, as a minimum, adhere to the following criteria:

a) Fixed production cycles and time of issuance;

b) Provide a limited set of products as determined by the revised Appendix II-6 of the *Manual on the GDPFS*; (reproduced in Annex II)

c) Provide verifications as per the WMO Standard Verification System for Long-Range Forecast (SVSLRF);

d) Provide up-to-date information on methodology used by the GPC;

e) Make products accessible through the GPC Web-site and/or disseminated through the GTS and/or Internet.

The Team noted that the passing of verification data to the Lead Centre for SVSLRF, was strongly recommended, as it is included as a statement containing a "should" in the Appendix II.6, but that it was not a "shall". The Team thus found that some centres, having difficulties to pass at this stage their verification data, but planning to do so in the future, should not be hampered by this temporary problem and should seek already recognition as GPCs.

The Team recommended that a document be passed to the CBS for assisting a probable Sub-Committee that will be assessing the achievements of the candidate GPCs. The document should clarify the requirements, in particular for the passing of verification data to the Lead Centre.

5.5 Progress Reports

The Team was pleased to receive presentations of status and progress reports by some Centres, who are all looking in the near future for recognition as GPC. Dr David Jones (Australia) presented the activities of Melbourne, Dr Arun Kumar (USA) presented the activities of NCEP (Washington), Dr Richard Graham (UK) presented the activities at Exeter, Dr Peiqun Zhang (China) presented the activities of Beijing, Dr Laura Ferranti presented the activities of ECMWF and Dr Dimitry Kiktev presented the activities of Moscow. The corresponding submitted documents can be found in Annex to this paragraph.

6. NEEDS OF RCCs and NMCs

To introduce the subject Dr Andre Kamga from ACMAD gave a presentation on the needs of climate prediction products for the West African region. Products should be “downscaled” to the concerned region. The impact on agriculture and other parties concerned should be taken into account for delivering the right information, by adapting the products to the needs. The forecasts must be available at a certain time that depends on the region, for example March for rainy season in West Africa. Indication on onset, cessation and length of rainy season are needed, but this could be more in the domain of extended range (weeks to month forecasts). There is a need to have verification maps per region. There is a need to have specific indices like the MJO signal. Diagrams showing simply forecasts versus climatology probability distribution function (pdf) are better for understanding.

Dr Madhavan Nair RAJEEVAN followed by presenting the overall needs of RCCs and NMCs. A general remark on behalf of the users of the products was that GPCs should strive to improve LRF skill, focusing on the main forecast parameters required by the RCCs.

6.1 Needs for Processed Products

Most of the GPC products are in the form of forecast maps. These maps, as currently provided, need few changes. There is higher requirement for data products (GRIB-2 format), so that NMCs can further do downscaling to meet their requirements. However, data products should be restricted to well established and tested output. Thus, website products may be expanded to include new experimental products. In addition to the minimum list of LRF products as stated in Appendix II-6 (Manual on GDPFS), the following experimental products are desired by RCCs and NMCs.:

- Averages, accumulations or frequencies over 1-month period to 3-month period.
- Probabilities of exceeding some threshold values (e.g., seasonal rainfall totals above a range of thresholds)
- Risk of extreme climate anomalies that may help in warning of e.g. occurrence of heat and cold waves over a particular region.
- Predicted generalized indices of drought, monsoon etc.
- Dry and wet spells: frequency and duration (with one month lead time)
- Probable date of onset of main rainy seasons (over a region, like South Asia, East Asia, southern Africa, GHA etc).
- The need to have first month (0-lead) averages was expressed.

It is however, recognized that development of some of these products will require further research by GPCs. In particular for the last two products, which are more connected to extended-range forecast between 1 to 4 weeks. It is wished that products from GPCs be more fitted to limited geographical area.

The Team recognized that downscaling products are needed for users. RCCs need to define the exact products they need to derive forecasts of the onset of monsoon. RCCs should do some studies and define the set of products needed to make the onset forecasts. There is some evidence that monthly forecasts have good signal of the onset in some regions.

The Team recommended that RCCs work and collaborate with GPCs through RCOFs or other bilateral or multilateral means to explore the predictability and define the products to be made available.

6.2 Needs for Data Products

The products are preferred in GRIB 2 format rather than NetCDF, especially for downscaling. The obvious requirement of data is for forecast products. However, there may be some observed products that GPCs are in a good position to provide. The requirements are as follows:

- Forecast data for downscaling algorithms; this is likely to require more than monthly mean data, e.g.:
  - Statistics on daily variability
  - Anomalies for some or all ensemble members
  - Hindcast data
- Data for RCM boundary and initial conditions (including SST data).
- Data for calculating regional specialized indices (drought).
- Analyzed fields of surface and upper air parameters for use in empirical models as predictors.
- Observed and predicted global weekly values of SST.
- Daily satellite precipitation analysis for use in monitoring through the season.

6.3 Requested Scores

- Scores for minimum list products (as in Appendix II-6 of Manual on GDPFS) should be readily available.
- Means of assessing skill for the new products may need special consideration by the expert teams for CBS and CCl on verification
- Scores should be user friendly, which can be understood by forecasters in NMCs.

6.4 Training

Regional climate centres/ NMCs may not have expertise in all aspects of Long range forecasts. They will need assistance in training from GPCs in the following main areas:

- Interpretation and use of GPC LRF products
- Downscaling techniques (both statistical and dynamical)
- Verification techniques (for local verification of RCC generated products and application outputs)
- Development of local user applications from RCC downscaled products
- Use and implementation of regional climate models.

Training might take place in the form of attachment of RCC staff to GPCs (for 2-3 months), exchanges of visiting scientists (for 2-3 months) or capacity building workshops. The Team suggested that WMO undertake to organize training seminar(s) (minimum 6 days) with order of programme similar to those for the Medium-range products training seminars. The Team also suggested that WMO may support participants for attending appropriate courses.

6.5 Coordination among GPCs.

It would benefit the RCCs if GPCs could converge in forecast formats, issuance times etc. Also establishment of a kind of clearing house (a possible lead centre) for collection of all available GPC products would help efficient transfer of forecast data to RCCs.

6.6 Climate Predictability Tool (CPT)

Dr Simon Mason from IRI presented the Climate Predictability Tool (CPT). The CPT is a software package developed by the IRI designed for making downscaled seasonal climate forecasts by RCCs and NMCs. The software is an easy-to-use downscaling tool that runs on Windows 95+, and is specifically designed to produce statistical forecasts of seasonal climate using either the output from a GCM or from fields of sea-surface temperatures. It thus acts both as a statistical forecasting package as well as a tool for conducting model output statistics (MOS) corrections to downscaled GCM predictions. A source code version of the software is available for use on other computer systems, but this version does not have the graphical user interface or graphics output facilities of the Windows version.

The software was developed primarily to address some of the problems that were being experienced in producing consensus forecasts at the various climate outlook forums (COFs). Specifically it was developed to address the following issues:

a. Slow production time: the time taken to construct the regression-based statistical forecasts at the COFs was requiring long and expensive pre-forum workshops, and prohibited capacity building for further development of forecasting and verification capabilities;

b. Artificial skill: some of the statistical models developed were subject to artificially inflated skill estimates, and the limited time available for performing rigorous validation prohibited the detection of such problems;

c. Dependence on one model: the effort invested in constructing the statistical model tended to encourage an over-confidence in the prediction from this single model, and inputs from other sources (most notably the dynamical predictions from GPCs) were largely down-played because of the lack of a sense of ownership of these additional products;

d. Unreliable probabilities: the statistical methods used to obtain forecast probabilities are known to be unreliable given the small sample sizes typical of most seasonal climate forecasting systems.

The software has been introduced to most of the COFs and is now used fairly extensively, especially in the Southern African COF (SARCOF), where it has been successfully in promoting the consideration given to GPC products. It is being used increasingly in other areas, including South America and South-East Asia where the source code version is also being used. Some training workshops in the use of the software have already been held, including under the auspices of the CLIPS activities.

The software requires a hindcast set as well as the current forecast, and the hindcast set is used to downscale the GCM predictions typically to station data provided by the NMS. A forecast is constructed using either canonical correlation analysis or principal component regression. Extensive diagnostic statistics are provided, including most of the scores and procedures recommended by the SVSLRF (including the calculation of significance levels and error bars), and the scores are calculated in cross-validation mode, as well as there being an option to calculate retroactive performance measures.

There are options that provide considerable flexibility to tailor forecasts for specific user requirements, including options to define the above- and below-normal categories at percentiles other than the terciles, or to define them using absolute values or as anomalies. Forecasts can be produced in a variety of formats, and detailed information is provided so that the forecast can be communicated to the end users in easy-to-understand terms.

The Team congratulated Dr Mason for this very useful and easy to implement tool, which is freely downloadable in the Windows version and at request in Linux version (http://iri.columbia.edu/outreach/software/).

7. ONGOING COORDINATION AND SUPPORT OF LEAD CENTRE ROLE

Dr David Jones on behalf of Dr Andrew Watkins gave an update on the Lead Centre activities. WMC Melbourne/Australian Bureau of Meteorology and RSMC Montreal/Meteorological Service of Canada co-hosts of the Lead Center for LRF Verification achieved a successful launch of the Lead Center website in early 2006. The WMO has invited Global Producing Centres for LRFs to submit their verification results for inclusion on this website in a letter sent to PR on February 6. A detailed description of the status of the Lead Center is to be found at ET CBS-OPAG/DPFS/ET/LRF/Doc. 7(1), with the website available at http://www.bom.gov.au/wmo/lrfvs/.

The Lead Centre has so far processed some verification information from 4 long range forecast models, originating from 3 separate institutions (JMA (2), MSC and UKMO). These results are currently displayed on the website. They have also demonstrated to the Lead Centre the practicalities (or otherwise) of some of the procedures in processing and displaying the verification information. Such interactions have resulted in a streamlining of procedures at the Lead Centre. Many centres have expressed their intention to submit verification to the SVSLRF web site in the coming months.

The Team assigned one additional responsibility to the Lead Centre concerning the development of software for graphical display of confidence level information. Once this development work is completed the guidelines included in the appendix II.8 will be updated.

8. VERIFICATION TECHNIQUES

8.1 This chapter deals with all the issues brought to our attention by the previous meeting of the ICT. The discussion between the ET members on these issues are summarised below.

8.1.1 "Clarification on whether or not verification should be carried out on post-processed output"

In the appendix II-8 it is not stated explicitly. Most centres are currently posting forecasts derived from a simple calibration and so, for the sake of comparison on the Lead Centre web site, it was decided that scores for forecasts that were calibrated are to be submitted (see updated Attachment II-8 in Annex to paragraph 8.1.2). At the moment the Team prefers to exclude verification of forecasts that were post-processed using for example Model Output Statistics or Perfect Prog approaches but GPCs are encouraged to display verification on their post-processed forecasts on their own web sites.

8.1.2 "Development of more information on error bars and significance levels to be made available in the documentation, and consideration of the best means of displaying such information"

Significance levels can be derived from either standard significance statistical test or bootstrapping techniques. The ET agreed on the general principle that if standard significance tests for a given score are available and valid, given the assumptions about the data, it will be preferable to use them. The ET needed further discussion before recommending specific methods to assess significance levels. Therefore the requirement for the submission of significance levels to the Lead Centre web site should be postponed. The GPCs are still encouraged to perform significance level testing and display their results. The ET has updated the appendix II.8 to reflect this decision (See Annex to this paragraph). The Lead Centre will carry development on this subject with help from ET members and report at the next meeting.

8.1.3 "Calculation of the area under ROC curves (use of fitted curves or not)"

In the appendix II-8, it is stated explicitly at the section 3.3.3 that: “For the core SVSLRF the area under the ROC curve should be calculated using the Trapezium rule”. There was no ambiguity there.

8.1.4 “Responsibility for display of real time monitoring information”

In the appendix II-8 it is stated that it is the responsibility of each GPC to display real-time performance monitoring. The Team was of the opinion that the SVSLRF is really for hindcast verification. The Team felt that there is need to improve the provision of near real time global observed data set for verification purpose.

8.1.5 “Need for more guidance on the prescription of the cross-validation procedure and its appropriateness for individual dynamical models”

The ET agreed that the cross-validation should be mandatory for both calibrated and re-calibrated forecasts. It is clearly unavoidable for training of empirical models and statistical post-processing as well as multi-model combination schemes if the data set used is not large enough to be divided in 2 parts (training and then validation). There is a need to define a rigorous cross-validation procedure that can be used as part of the standard SVS guidelines.

8.1.6 “Specification of ENSO years”

In 2005 an expert team was formed by the Commission for climatology (CCl) to prepare a catalog of El Nino and La Nina indices and definitions used around the world. This team was led by Fiona Horsfall. The team have submitted a report at the Beijing meeting in November 2005. This report essentially recommends that another expert team should be formed to do more work on this. This recommendation was accepted and the new team is named “expert team on El Nino and La Nina” (OPAG 3). The team is led by Luc Maitrepierre and is composed of 7 international scientists. Our hope is that they will provide us with a list of El Nino and La Nina years to allow for a stratification of the data according to these criteria. The ET on SVSLRF should follow closely the work of this new ET. The ET recommends that the SVS verification need not be stratified according the ENSO years until we have a clear official definition of the phenomenum.

8.2 According to the ICT the future work of the team should involving the topics cited in the excerpts below:

“...6.25 The Meeting recommended that the expert team on SVSLRF continue its work for the next period. Areas that may need future consideration to augment the SVSLRF are:

- Development of scores to measure skill in the ensemble spread
- Assessment of multi-model ensembles
- Standardising methods for defining terciles, etc.
- Verification of extremes (such as the outlying quintiles)
- Standardising of hindcast period
- Standardising verification data sets
- Ongoing coordination and support of Lead Centre role
- Clarification of issues arising from the broader use of SVSLRF...”

8.2.1 The Team needs to review options for new scores to be introduced to the SVSLRF that can be applied to probabilistic forecasts when forecasts are expressed on a continuous scale. These procedures would be relevant to forecasts that are expressed either as a parameterized distribution or by fitting a kernel (*a smoothing function that acts as a distribution*) to an ensemble. There are three properties of probabilistic skill scores for such forecasts that should be considered:

1. Propriety: the score should not encourage the forecaster to hedge (i.e. adjust the forecast to improve the score), but rather to issue forecasts consistent with his/her belief;
2. Effectiveness: the expected score must be a strictly decreasing function of the quality of the forecast;
3. Locality: the score should depend only on the probability assigned to the verification.

The desirability of the third property requires further consideration by the team since it has implications for recommending the exclusion of scores such as the ranked probability score (RPS) that do not have this property. The team should consider the suitability of scores such as the log probability score (including the ignorance score) as candidates for inclusion in the SVSLRF.

8.3 Here is a summary of the discussion that the ET members had during the meeting on these specific issues listed at the beginning of 8.2:

8.3.1 "Development of scores to measure skill in the ensemble spread"

The ET recognized that identifying whether there is a correlation between the accuracy of a forecast and the ensemble spread is not an optimal way of identifying whether there is any information in the ensemble distribution. A more successful approach would involve comparing the quality of the forecasts given the observed ensemble spread / distribution with the quality achieved by keeping the ensemble spread / distribution constant. The ET needs to provide detailed guidelines for conducting such tests.

8.3.2 "Assessment of multi-model ensembles"

No new scores are required specifically for assessing the quality of multi-model ensemble forecasts (except for the need for probabilistic scores on continuous scales), but the ET needs to consider making recommendations for minimizing problems associated with the dangers of over-estimating forecast performance given a large number of models ("multiplicity"). Specifically the ET should establish some guidelines for conducting rigorous out-of-sample validation.

8.3.3 "Standardising methods for defining terciles, etc. "

Two approaches for defining quantiles are in common usage: parametric methods based on assumption of a distribution (eg. tercile boundaries can be estimated at +/- 0.43s.d for data with a Gaussian distribution), and ranking or counting methods (eg. the lower tercile separates the data ranked in the lowest third of the sample). Parametric methods require choice of the most appropriate distribution and parameter estimation procedures, and incorrect choices can lead to pathological results (e.g. lower quantiles for precipitation may have negative values). The ET therefore recommended that counting methods should be used in preference to parametric methods.

There are various ways of applying the counting method, the differences lying in the details of interpolation from the two data points surrounding the quantile (the simplest method being an unweighted average of the two surrounding values). The ET recommends that the relative benefits of the different interpolation methods be explored before defining a recommended method for the SVS.

8.3.4 "Verification of extremes (such as the outlying quintiles)"

The ET recognized the inherent difficulty of verifying forecasts of extreme events because of the small sample sizes involved. The only option is to perform verification and to indicate the uncertainty in the calculation of these scores. The uncertainty in these scores will be unavoidably large. While the existing SVSLRF contains adequate procedures for verification of probabilistic forecasts of extreme events, the ET recognizes the importance of estimating confidence limits for these verification scores. Appropriate procedures for calculating these confidence limits need to be added to the SVSLRF manual.

The ET needs to identify appropriate procedures for verification of deterministic forecasts expressed as estimates of the frequency of extreme weather events during the season. The team should consider the following options:

a. Data transformation: can the counts be transformed to have normal distribution, and if so would the current scores for deterministic forecasts in the SVSLRF be appropriate?

b. Categorization: should the counts could be categorized, and the current scores for categorical forecasts in the SVSLRF be used? If so the ET needs to consider guidelines for the categorization.

c. New scores: would a new set of scores be more appropriate, such as percentage error instead of mean bias, and non-parametric measures of association instead of Pearson's correlation?

8.3.5 "Standardising of hindcast period"

After discussion, it was decided that the recommended period of hindcast should be 1981-2002 for submission to the Lead Centre web site. The beginning of the period was chosen to be 1981 because it is the first year where good ocean observation data were made available. The end of the period was chosen because it is the ending year of the ERA-40 data set. The specification of this period will be reconsidered by the Expert Team as available observations data sets evolved. However, in-homogeneity and incompleteness of reference observation data impose constrain on increasing the hindcast periods. In future, changes in the details of SVS recommendations, such as hindcast period, will be communicated through the Lead Centre web site to avoid frequent updating of Appendix II-8.

8.3.6 "Standardising verification data sets"

The precipitation field is the only parameter for which more than one data set are currently listed in the appendix II-8. Effectively in the document and on the web site both the GPCP and the Xie-Arkin (CMAP) data sets were recommended. A paper by Yin et al. (2004) suggests that the GPCP has fewer flaws than the Xie-Arkin one. In addition the ET have been informed that the Xie-Arkin may be discontinued in the near future. The ET recommends the GPCP data as the official data set for precipitation verification.

9. CLOSURE OF THE MEETING

After having lead the joint Teams with diplomacy and efficiency the two chairmen, Dr Willem Landman and Mr Normand Gagnon closed the Meeting at 12.45 on 7 April 2006.

ANNEX TO PARAGRAPH 1.2

LIST OF PARTICIPANTS

CHAIRMEN

Mr Normand **GAGNON**
Chairman of WMO/CBS Expert Team on Standardized Verification System for Long-Range Forecasts
Canadian Meteorological Centre
2121 Transcanada Highway
H9P 1J3
Dorval Québac
Canada

Tel: +1514 421 47 12
Fax: +1514 421 46 57

Email: normand.gagnon@ec.gc.ca

Dr Willem A. **LANDMAN**
Chairman of WMO/CBS Expert Team on Infrastructure for Long Range Forecasting
South African Weather Service
Private Bag X097
0001
Pretoria
South Africa

Tel: +2712 367 6003
Fax: +2712 367 6189

Email: willem@weathersa.co.za

PARTICIPANTS

Dr David A. **JONES**
Australian Bureau of Meteorology
GPO Box 1289X
3001 Melbourne-Victoria
Australia

Tel: +613 9669 40 85
Fax: +613 9669 46 78

Email: d.jones@bom.gov.au

Dr Peiqun **ZHANG**
China Meteorological Administration
46 Zhongguancun Nandajie
Haidian District
BEIJING 100081
China

Tel: +(86 10) 6840 7175
Fax: +(86 10) 6840 7175

Email: zhangpq@cma.gov.cn

Dr Madhavan Nair **RAJEEVAN**
National Climate Center
India Meteorological Department
Shivaji Nagar
411 005 Pune
India

Tel: +9120 2553 58 77
Fax: +9120 2553 54 35

Email: rajeevan61@yahoo.co.in

Dr Tomaki **OSE**
Climate Prediction Division
Japan Meteorological Agency
1-3-4, Otemachi
Chiyoda-ku
TOKYO 100-8122
Japan

Tel: +(81 3) 3211 8341 ext. 4224
Fax: +(81 3) 3211 8406

Email: tomoaose@mri-jma.go.jp

Dr Won-Tae **YUN**
Climate Prediction Division
Koea Meteorological Administration
460-18 Sindaebang-dong Dongjak-gu
Seoul 156-726
Republic of Korea

Tel: +(82 2) 2181-0842
Fax: +(82 2) 832-6018

Email: wtyun@kma.go.kr

Dr Dmitry **KIKTEV**
Hydrometcentre of Russia
11-13 Bolshov Predtechensky per.
123242 MOSCOW
Russia

Tel: +7495 255 13 54
Fax: +7495 255 15 82

Email: kiktev@mecom.ru

Mr Richard **GRAHAM**
Met Office
B2 W015 FitzRoy Road
EX1 3PB Exeter
U.K.

Tel: +44 1392 886 361
Fax: +44 1392 885 681

Email: richard.graham@metoffice.gov.uk

Dr Arun **KUMAR**
Climate Prediction Center (W/NP51)
National Centers for Environmental Prediciton
World Weather Building, Room 800
5200 Auth Road
MD20746-4304 CAMP SPRINGS
USA

Tel: (+1 301) 763-8000 x7579
Fax: (+1 301) 763-8125

E-mail: arun.kumar@noaa.gov

Dr Simon **MASON**
IRI (representing CCl)
61 Route 9W
Palisades
New York 10964-8000
USA

Tel: +1 845 680 45 14
Fax: +1 845 680 48 65

Email: simon@iri.columbia.edu

Dr Laura **FERRANTI**
ECMWF
Shinfield Park
RG2 9AX Reading
U.K.

Tel: +44 118 949 96 01
Fax: +44 118 986 94 50

Email: laura.ferranti@ecmwf.int

Dr Anna **GHELLI**
ECMWF
Shinfield Park
Reading, Berkshire
RG2 9AX
U K

Tel: ++44 (0)118 949 9425
Fax: ++44 (0)118 986 9450
Email: a.ghelli@ecmwf.int

Dr Francisco **DOBLAS-REYES**
ECMWF
Shinfield Park
Reading, Berkshire
RG2 9AX
U K

Tel: ++44 (0)118 949 9655
Fax: ++44 (0)118 986 9450
Email: f.doblas-reyes@ecmwf.int

Dr Tim **STOCKDALE**
ECMWF
Reading, Berkshire
RG2 9AX
U K

Tel: ++44 (0)118 949 9117
Fax: ++44 (0)118 986 9450
Email: T.stockdale@ecmwf.int

WMO Secretariat
7 bis avenue de la Paix
Case postale 2300
1211 GENEVE 2
Switzerland

WWW website

www.wmo.int/web/www/www.html

Mr Joël **MARTELLET**

Tel: +(41 22) 730 8313
Fax: +(41 22) 730 8021

Email: jmartellet@wmo.int

ANNEX TO PARAGRAPH 2.1.1

AGENDA

1. OPENING OF THE MEETING

2. ORGANIZATION OF THE MEETING
   - 2.1 Adoption of the agenda
   - 2.2 Other organizational questions

3. PROGRESS ON EXCHANGE OF ENSEMBLES PRODUCTS AND DEVELOPMENT OF MULTI-MODEL ENSEMBLES
   - Exchange of ensemble products
   - Status of MME centres
   - Definition of minimal ensemble size for practical considerations (for hindcast runs vs. real-time runs, and for nesting limited area models)

4. DATA NEEDS FOR PRODUCING GLOBAL LRF
   - Land surface data: e.g. soil moisture
   - Ocean sub-surface data

5. OPERATIONAL PROCEDURES FOR DISSEMINATION AND EXCHANGE OF PRODUCTS, TERMS AND CONDITIONS
   - Resolution 40 clarification
   - Publication of scores
   - Central Facilitation Centre
   - Clarify role of GPCs, RCCs, NMCs

6. NEEDS OF RCCs-NMCs
   - Relative to products (e.g. access to boundary conditions for downscaling)
   - Relative to data
   - Relative to training

7. STATUS OF LEAD CENTRES FOR VERIFICATION OF LRF

8. VERIFICATION TECHNIQUES
   - New scores for ensemble
   - New scores for multi-model ensemble
   - Standard way to define terciles
   - Verification of extremes
   - Skill for ensemble spread
   - Standardised data sets (to be put on Web site)
   - Standardised hindcast period

9. CLOSURE OF THE MEETING

ANNEX TO PARAGRAPH 3.1.1

Towards a multi-model real-time ensemble system for seasonal prediction: the EURO-SIP consortium (ECMWF, Met Office and Météo-France)
(Submitted by Mike Davey and Laura Ferranti)

1. Background

The ocean, land and atmosphere models used in seasonal forecast systems contain biases and errors that contribute substantially to forecast uncertainty. With a multi-model ensemble it is possible to represent some of these uncertainties and hence improve the quality of forecasts. A multi-model ensemble-based system for seasonal-to-interannual prediction was developed in the European DEMETER (Development of a European Multi-model Ensemble system for seasonal to inTERannual prediction) project, supported by the European Union Fifth Framework Environment Programme. Results from DEMETER show enhanced reliability and skill for the multi-model ensemble over a more conventional single-model ensemble approach. (See Palmer et al. 2004, http://www.ecmwf.int/research/demeter , and the Tellus 2005 special issue.)

Building on the success of DEMETER, a European consortium has been developing and operating a real-time multi-model ensemble forecasting system, as described below. Forecasts are produced each month. Most of the computation (running the model ensembles, processing and archiving output) is carried out at ECMWF, making use of a common framework.

2. The real-time EURO-SIP system and its implementation

The EURO-SIP (European Seasonal to Interannual Prediction) multi-model system currently has three participants (ECMWF, Met Office and Météo-France), who have each developed an independent coupled general circulation model.

The ECMWF model is system2 (soon to be upgraded to system3). The atmospheric component has a spatial resolution of ~200km and 40 levels in the vertical. Ocean initial conditions are taken from the ECMWF ocean analysis system. The forecast ensemble size is 40, and hindcasts integrations cover the period 1987-2001. The hind-cast ensemble size is 5 (soon to be increased to 11). Documentation of the ECMWF operational system is available at
http://www.ecmwf.int/products/forecasts/seasonal/documentation .

The Met Office model is GloSea. The atmospheric component has a horizontal spatial resolution of 2.5x3.75 degrees and 19 levels, while the ocean has 40 levels. Ocean initial conditions are taken from the Met Office ocean analysis system. The forecast ensemble size is 41, and 15-member hindcasts for the period 1987-2001 are used for calibration. Documentation of the Met Office GloSea system is available at http://www.meto.gov.uk/research/seasonal/glosea.html .

The Météo-France model is Arpege/ORCA. Arpège4 (the atmospheric component) has 31 vertical levels and a spatial resolution of about 300Km. ORCA (the oceanic component, OPA8.2) is an ocean model developed at LOCEAN in Paris. The ocean initial conditions are prepared by MERCATOR in Toulouse. Forecast ensemble size is 41, and hind-casts cover the period 1993-2004 with a 5 member ensemble.

All the ocean components have enhanced near-equatorial resolution.

The ensemble sizes are constrained by computational resources. It is difficult to indicate an optimal hind-cast ensemble size that is the best compromise between computer demand and adequate sampling of forecast probability distribution and skill. (Indeed, the optimal size could well be model-dependent.) Sub-sampling of hind-cast years in the DEMETER dataset (1959-2001) reveals a lack of stability in skill scores. This indicates that the uncertainties associated with the length of the hind-cast period are significant in forecast skill estimates.

Further European centres may join the consortium in the future.

2.1 Operational implementation and archive

The EURO-SIP system is fully operational. Data and products from the multi-model system are produced by 12Z on the 15th of each month.

Forecast and hind-cast data from the three coupled systems, are archived in the ECMWF Meteorological Archival and Retrieval System (MARS), and thus are available by this route to European member states and other centres with access to MARS. At some stage probabilities might also be archived.
The EURO-SIP data are archived in MARS in a similar fashion to that used for the operational ECMWF seasonal forecast data: for details see:
http://www.ecmwf.int/products/forecasts/seasonal/documentation/ch3_2.html. Data from the Met Office and Météo-France systems can be retrieved by using an additional descriptor "origin". Met Office data has origin=egrr, and Meteo-France data has origin=lfpw.

As each individual EURO-SIP component system is upgraded in the future, new sets of hind-casts will be produced. Such model changes will not be synchronised. A descriptor “system” is used to distinguish the model upgrades of each EURO-SIP system. (Currently the Met Office uses system=3 and ECMWF and Météo-France use system=2.)

Note that the data are not generally accessible to all NMHSs by this route. The topic of wider distribution of digital data is under review by the consortium.

2.2 Multi-model products distribution and verification

To meet differing requirements, the consortium members run various post-processing packages. The basic multi-model products use a simple average of the 3 probability distributions associated with each individual ensemble system. In some cases hind-cast data are used to adjust the probabilities: currently the Met Office applies discriminant analysis for this purpose. At a later stage Bayesian techniques, such as those developed at ECMWF, will be extended to the multi-model.

Multi-model products are now available on the Met Office web site (see http://www.metoffice.gov.uk/ , currently using Met Office and ECMWF data), and will also be available on the ECMWF web site (http://www.ecmwf.int). The sets of products are similar to those produced for the individual systems – see the web sites for details. For example, temperature and precipitation probabilities for various sets of 3-month-average categories out to six months ahead are provided. (The Météo-France website is under development: products are presently made available to NMHSs via an ftp site: ftp://ftp.meteo.fr .)

Estimates of the multi-model skill levels and of the skill of each individual component are made using the methods agreed by the LRFSVS ET. Documentation of the multi-model products and of each individual component will be made available.

2.3 Exchange of ensemble products

As described above, the provision of multi-model products via websites is at an advanced stage of development. The consortium will aim to meet the recommendations for global producers, the details of which are on the agenda for this meeting of the ET LRF.

The provision of multi-model products by other routes, e.g. the ECMWF catalogue, is under discussion, recognizing that there is commercial interest in such data.

3. The European ENSEMBLES project

The EURO-SIP members are also participating in a European Union Framework 6 project called ENSEMBLES (http://www.ensembles.org). This project contains a substantial effort on seasonal to decadal forecasting that is relevant to multi-model development. As was the case in DEMETER, this project also contains strong links to applications in e.g. the agriculture and health sectors.

The first stage has recently been completed, with testing of new systems and methods used to create ensembles. Several sets of hindcasts, to 6-month, 12-month and decadal range, have been produced by the leading European climate modelling centres.

Following the successful developments during the DEMETER project, the multi-model ENSEMBLES hindcasts are archived in MARS using a common format, which will allow comprehensive verification of the forecast quality at the different time scales. The results will be displayed on the ECMWF external website.

ANNEX TO PARAGRAPH 3.1.3

STATUS OF APEC CLIMATE CENTER (APCC)

1 Current status of APCC

APCC, "APEC Climate Center," is a regional climate program aimed at realizing the APEC vision of regional prosperity through the reduction of economic losses due to abnormal climate. Asia-Pacific Economic Cooperation (APEC) is the premier forum for facilitating economic growth, cooperation, trade and investment in the Asia-Pacific region. APEC was established in 1989 by the endorsements of the summits from 12 Member Economies including the United States and Korea. Currently, APEC has 21 members, which account for more than 2.5 billion people, a combined Gross Domestic Product of 19 trillion US dollars, and 47 percent of world trade.
APCC produces real-time operational climate prediction information based on a well-validated multi-model ensemble system. APCC contributes to enhancing capacity-building in the monitoring and prediction of unusual weather and climate in the Asia-Pacific region by sharing high-cost climate data and information. Those countries without the capacity to produce climate predictions are able to access optimized, high-cost global climate predictions produced by APCC.

2 Goals

The APCC is designed to set up an institutionalized communication channel for more effective exchanges of regional climate information among APEC member economies. Therefore, the APCC aims at realizing the APEC vision of regional prosperity through the enhancement of economic opportunities, the reduction of economic loss and the protection of life and property through:

- Production of skillful real-time climate predictions;
- Facilitating the share of high-cost climate data and information;
- Enhancing the capacity building in prediction and sustainable social and economic applications of climate information;
- Minimizing climate related damages;
- Capitalizing on non-preventable damages; and
- Accelerating and extending socio-economic innovation.

3 Functions

The mission of APCC is to enhance the socio-economic well-being of member economies by utilizing up to date scientific knowledge and applying innovative climate prediction techniques through:

- Developing a value-added reliable real-time climate prediction system, through a state-of-the-art multi-model climate prediction system utilizing model predictions from member economies;
- Acting as a center for climate data and related information with open access to member economies;
- Helping build the capacity of member economies in producing and using reliable climate predictions;
- Developing improved methods of utilizing socio-economic innovation to mitigate and adapt to climate fluctuations and change and guide member economies towards optimum utilization of APCC climate prediction information; and
- Coordinating research toward the development of an APEC integrated climate-environment-socio-economic system model (ultimate and longer-term scope).

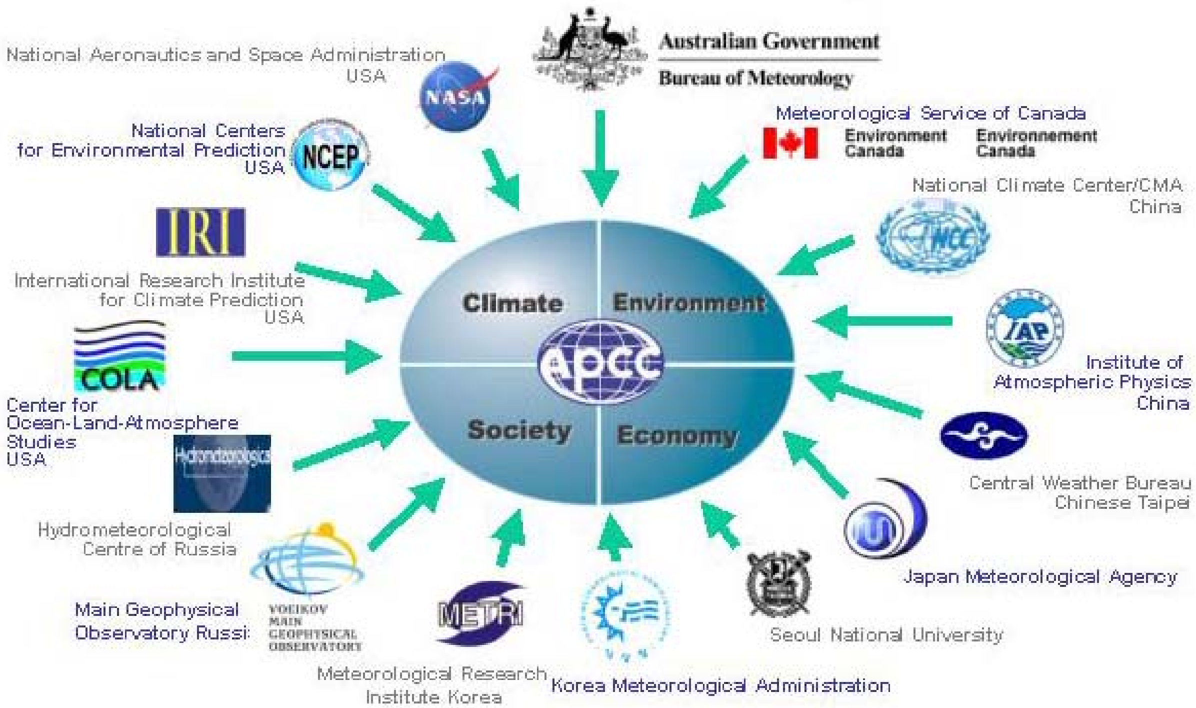
4 History

At the fourth APEC Ministers Conference in March 2004, the Ministers recognized the proposed initiative of the APEC Climate Center (APCC), and then the member economies supported the establishment of APCC at the twenty-seventh ISWTG meeting held in Singapore in September 2004. APCC was identified as the Modeling and Data Processing Center of Global Earth Observation System of Systems (GEOSS), representing the Asia-Pacific region, and recorded in the GEOSS 10-Year Implementation Plan Reference

Document adopted at the third Earth Observation Summit held in Brussels in February 2005. In recognition of the decision made by the ISTWG, the establishment of APCC was endorsed at the first APEC Senior Officials' Meeting (SOM I) held in Seoul, Korea in March 2005. The APCC functions and operations were endorsed at the twenty-eighth APEC ISTWG Meeting held in Gwangju, Korea in March 2005. Hosted by the Deputy Prime Minister and the Minister of Science and Technology of the Republic of Korea, the APCC Opening Ceremony was held in Busan to commemorate its start on the occasion of the 13th APEC Economic Leaders' Meeting in Busan, Korea on 18-19 November 2005. The APEC Climate Center is located in Busan, Korea and started its reinforced work after the Opening Ceremony.

5 Participating organizations

The participating organizations and institutes involved in the APCC-MMES are as follows. Australian Government-Bureau of Meteorology, Meteorological Service of Canada, China Meteorological Administration, Institute of Atmospheric Physics of China, Central Weather Bureau of Chinese Taipei, Japan Meteorological Agency, Korea Meteorological Administration, Meteorological Research Institute of Korea, Russian Federal Service for Hydrometeorology and Environmental Monitoring, Main Geophysical Observatory of Russia, Center for Ocean-Land-Atmosphere Studies of the USA, national Aeronautics and Space Administration, National Center for Environmental Prediction of the USA, and the International Research Institute for Climate Prediction of the USA.



The participating organizations and institutes involved in the APCC-MMES are as follows.

Country	Model	Organization	Resolution
Australia	POAMA	Bureau of Meteorology Research Centre	T47L17
Canada	MSC	Meteorological Service of Canada	1.875° ×1.875° L50
China	NCC	National Climate Center/CMA	T63L16
	IAP	Institute of Atmospheric Physics China	4° ×5° L2
Chinese Taipei	CWB	Central Weather Bureau	T42L18
Japan	JMA	Japan Meteorological Agency	T63L40
Korea	GDAPS/KMA	Korea Meteorological Administration	T106L21
	GCPS/SNU	Seoul National University	T63L21
	METRI/KMA	Meteorological Research Institute	4° ×5° L17
Russia	MGO	Main Geophysical Observatory	T42L14
	HMC	Hydrometeorological Centre of Russia	1.125° ×° 1.40625L28
USA	COLA	Center for Ocean-Land-Atmosphere	T63L18
	NCEP	NCEP Coupled Forecast System	T62L64
	NSIPP/NASA	National Aeronautics and Space	2° ×2.5° L34
	-	International Research Institute	-

6 Organizational structures

The APEC Climate Center is an server as a hub of regional climate network for realization of concerted and systematic effort to produce a skillful climate prediction and information with open access to member economies, and to enhance the capacity building of member economies in climate prediction and its applications for disaster prevention and sustainable socio-economic growth. Full-time staffs are composed of the Executive Director (1) and the staffs of three divisions: Administration (7), Science (8), and System (5).

7 Future plan

Objectives set in the APCC Science Plan include development of (i) a well-validated multi-model ensemble climate prediction system (ii) an integrated climate-atmospheric environmental monitoring and prediction system (iii) user application methods, particularly decision making models using climate prediction (iv) seamless climate prediction system and application system for socio-economic benefit in a long-term plan.

:

ANNEX TO PARAGRAPH 4

OBSERVATION DATA NEEDS FOR PRODUCING GLOBAL LONG –RANGE FORECASTS
(updated April 2006)

This Statement of Guidance (SOG) was developed through a process of consultation to document the observational data requirements to support seasonal-to-interannual (SIA) climate prediction. This version was prepared originally by the ET-ODRRGOS with experts from the NWP community, and was subsequently updated in consultation with a number of experts from the climate community through AOPC and by the CBS ET on Infrastructure for Long-Range Forecasting. It is expected that the statement will be reviewed at appropriate intervals by the OPAG on Data Processing Forecasting Systems to ensure that it remains consistent with the current state of the relevant science and technology

1. Introduction

Coupled atmosphere-ocean models are used to produce seasonal-to-inter-annual forecasts of climate. While empirical and statistical methods are also used to predict climate conditions a season ahead, the present assessment of how well observational requirements are met relates only to the coupled model inputs. It is noted that historical data sets also play an important role in SIA prediction by supporting calibration and verification activities.

Whilst such forecasting is still subject to much research and development, many seasonal forecast products are now widely available. The complexity of the component models ranges from simple models to full general-circulation-model representations of both the ocean and atmosphere. There is also large variation in the approach to the assimilation of initial data, with some of the simpler models assimilating only wind information while the more complex models usually assimilate subsurface temperature information and satellite surface topography and temperature data. Indeed, major challenges remain in the development of assimilation techniques that optimise the use of observations in initialising models.

The time and space scales associated with seasonal-to-interannual variability (large scale, low frequency) suggest the key information for forecasts will derive mostly from the slow parts of the climate system, in particular the ocean, but also the land surface. When considering impacts such as rainfall deficiencies or increased temperatures over land, however, there are very good reasons for considering variables associated with the land surface conditions. In particular, land surface moisture and vegetation should be specified and predicted. The models should also include up-to-date radiative forcing (e.g. greenhouse forcing), which are important for maximising skill in forecasts of land surface air temperature anomalies relative to recent historical reference-normal periods.

In this list of observation needs, the requirements for SIA forecasts are based on a consensus of the coupled atmosphere-ocean modelling community. It builds on the requirements for Global NWP and represents in addition those variables that are known to be important for initialising models or for testing and validating models. For the most part, aspects that remain purely experimental (i.e. unproven) are not included. There is some attempt to capture the impacts aspects; that is, those variables that are needed for downscaling and/or regional interpretation.

2. Data Requirements

The following terminology has been adhered to as much as possible: marginal (minimum user requirements are being met), acceptable (greater than minimum but less than optimum requirements are being met), and good (near optimum requirements are being met).

2.1 Sea surface temperature

Accurate SST determinations, especially in the tropics, are important for SIA forecast models. Ships and moored and drifting buoys provide observations of good temporal frequency and acceptable accuracy, but coverage is marginal or worse over large areas of the Earth. Instruments on polar satellites provide information with global coverage in principle, good horizontal and temporal resolution and acceptable accuracies (once they are bias-corrected using *in situ* data), except in areas that are persistently cloud-covered (which includes significant areas of the tropics). Geostationary imagers with split window measurements are helping to expand the temporal coverage by making measurements hourly and thus creating more opportunities for finding cloud-free areas and characterising any diurnal variations (known to be up to 4 degrees C in cloud free regions with relatively calm seas). Microwave measurements provide acceptable resolution and accuracy and have the added value of being able to 'see through' clouds. Blended products from the different satellites and *in-situ* data can be expected to be good for SIA forecasts.

There is a requirement for high quality, fast delivery SST (ideally with accuracy < 0.1 deg C on 100 km spatial scale and < 0.25 deg C on 10 km spatial scale, available within 24h (by SST we mean eg bulk temperature at 2m depth).

2.2 Ocean wind stress

Ocean wind stress is a key variable for driving ocean models. It is important to recognise the complementarity between surface wind and surface topography measurements. Current models use winds derived from Numerical Weather Prediction (NWP), from specialist wind analyses or, in some cases, winds inferred from atmospheric models constrained by current SST fields. The tropical moored buoy network has been a key contributor for surface winds over the last decade, particularly for monitoring and verification, providing both good coverage and accuracy in the equatorial Pacific. Fixed and drifting buoys and ships outside the tropical Pacific provide observations of marginal coverage and frequency; accuracy is acceptable.

Satellite surface wind speed and direction measurements are now the dominant source of this information. Currently their data reach SIA models mostly through the assimilated surface wind products of NWP, where their positive impact is acknowledged. Overall, a two-satellite scatterometer system, or its equivalent, would provide good coverage and acceptable frequency, and it would complement the ocean-based systems. At this time, continuity and long-term commitment are a concern. Improved integration of the data streams and operational wind stress products from NWP and other sources will be needed to achieve acceptable or better coverage, frequency and accuracy.

High quality scatterometer winds are the best products available at the moment and need to be maintained operationally. Additional data would always be useful. For example data to allow better estimates of heat-fluxes and P-E (precipitation minus evaporation) could help give a better definition of the mixed layer structure.

2.3 Subsurface temperature

Many, but not all, SIA forecast models assimilate subsurface temperature and salinity data, at least in the upper ocean (down to ~500 m depth). The Tropical Atmosphere Ocean (TAO) / TRITON moored buoy network provides data of good frequency and accuracy, and acceptable spatial resolution, of subsurface temperature for the tropical Pacific, at least for the current modeling capability. The tropical moored network in the Atlantic (PIRATA) is better than marginal but does not yet have the long-term resource commitments and stability to be classified as acceptable. There is no array in the Indian Ocean. The Ships-of-Opportunity Programme (SOOP) provides data of acceptable spatial resolution over some regions of the globe but the temporal resolution is marginal. It is noted that SOOP is evolving to provide enhanced temporal resolution along some specific lines. The *Argo* Project is providing global coverage of temperature and salinity profiles to ~2000 m, mostly with acceptable-to-good spatial resolution, but only marginal temporal resolution in the tropics. In all cases the accuracy is acceptable for SIA purposes.

Ocean observation system over Equatorial Atlantic is deficient in moorings. Moorings at and near the equator are important. Equatorial moorings in the Indian Ocean are also useful.

2.4 Salinity

Salinity is becoming an important parameter. Some model are starting to make use of such data in the ocean data assimilation. ARGO is a major source of salinity observations. It provides global coverage of temperature and salinity profiles to ~2000 m, mostly with acceptable-to-good spatial resolution, but only marginal temporal resolution in the tropics. Valuable data also comes from the tropical moorings although data coverage is too limited. Surface salinity will be measured by satellite in the forthcoming research mission. There will be a need for continuity of those measurements.

2.5 Ocean topography

Ocean altimetry provides a measure of the sea surface topography relative to some (largely unknown) geoid (or mean sea surface position) that in turn is a reflection of thermodynamic changes over the full-depth ocean column. In principle, the combination of altimetry, tropical mooring and *Argo* will provide a useful system for initialising the thermodynamic state of SIA models. Long term commitments for satellite altimetry are required. Research satellites are providing a mix of data with acceptable accuracy and resolution and data with good spatial resolution (along the satellite tracks) but marginal accuracy and frequency. The "synoptic" global coverage, particularly beyond the tropical Pacific, is an important requisite. Ocean altimetry data can currently only be used to look at variability in the sea state. There are plans to make use of geodetic data to obtain information about the geoid and the mean state of the oceans. It is expected that geodetic data will become available from satellites; GRACE and CHAMP are flying missions; GOCE will be an important addition.

2.6 Surface heat and freshwater fluxes

There are a few sites in the tropical ocean where the data on surface heat flux are of value for validation and are required at a number of sites in the tropical oceans. NWP products (derived from analysis from short range forecast), in principle, have good resolution but the accuracy is at best marginal. Satellite data provide prospects for several of the components of heat flux, particularly shortwave radiation, but at present none is used on a routine basis for SIA assimilation. Precipitation estimates are important for validation because of the fundamental role of the hydrological cycle in SIA impacts. They also have importance in initialisation because of the links to salinity. However, there remain significant uncertainties in estimates of rainfall over the oceans. In addition the fresh water run off information from rivers (large estuaries) will become important in coastal areas and regional parts of the oceans, eg the Gulf of Bengal.

2.7 Ocean current data

Models generally do not currently assimilate ocean current data, perhaps in part because data is limited. However, because of the central importance of dynamics and advection, current data are important for testing and validation. For example, experimental fields of surface current for the tropical Pacific and Atlantic are now being produced routinely by blending geostrophic estimates from altimetry with Ekman estimates from remotely-sensed wind observations. Inferred surface currents from drifting buoys are acceptable in terms of accuracy and temporal resolution but marginal in spatial coverage. Satellite altimetry is also being used to infer the distribution of ocean currents. Moored buoys are good in temporal coverage and accuracy, but marginal otherwise.

2.8 In-situ sea level

In-situ sea level measurements provide an additional time-series approach (good temporal resolution and accuracy; marginal spatial coverage), particularly for testing models and validating altimetry.

2.9 Atmospheric data

Since several SIA systems are driven by winds and, in several cases, surface heat flux products from operational analyses, the global (atmospheric) observing system is fundamental for SIA forecasts and their verification.

2.10 Land surface

- Snow cover. Snow cover and depth are important, particularly at short lead times (intraseasonal-to-seasonal). Snow depth observations are marginal.
- Soil moisture and terrestrial properties:
  - Soil moisture are still very marginal although soil moisture initial conditions are a crucial element in the forecast performance in mid-latitudes spring/summer (Beljaars, 1996) and might extend predictability over land in the monthly to seasonal range (Koster et al., 2004a, b). Soil moisture drifts are ubiquitous in NWP models, due to deficiencies in land surface models and/or the forcing precipitation and radiative fluxes (Viterbo, 1996).
  - Due to its extended memory, the relevant quantity to initialise is the soil water in the root layer. There are no existent or planned direct observations of such quantity with global or even regional coverage. Soil moisture analysis relies on proxy data. Such data covers 3 main groups:
    - Observations related to the surface-atmosphere feedback, or the partitioning of available energy at the surface into sensible and latent heat fluxes (e.g. Screen-level temperature and humidity and early morning evolution of IR radiances in the window channels in geostationary platforms)
    - Observations related to the soil hydrology, such as microwave remote sensing; radiances are sensitive to water in the first top few cm of the soil.
    - Remote sensing observations related to plant phenology, such as leaf area index (LAI), fraction of available photosynthetically active radiation (fAPAR), broadly based in the contrast in reflectances between the visible and NIR. In as much as the phenological evolution of plants depends on available water, there is a soil water related signal in the LAI and/or fAPAR; conversely, assimilation of such quantities will constrain the model evaporation, impacting on the background soil moisture.
  - Without careful constraints, the use of one of the 3 classes of observations presented above will alias information into the analysed soil moisture. A strong synergy is expected from combining observations from each of the 3 classes above, because they sample "complementary directions" in the physical space.

2.11 Sea Ice cover and thickness

Sea ice cover is important for high latitudes. It is implicitly included in the leading SST products. Sea ice thickness is important for fluxes and would be useful for initialisation. Too few ice thickness measurements are presently available.

2.12 Other data

There are many other data sets that may play a role in future-generation SIA forecast models. Because these roles are largely unknown, it is premature to discuss the adequacy of observing systems to meet these needs; generally speaking, they are not expected to rank near the above data in terms of priority. These data sets include:

- Ocean colour. Ocean transparency is already included in several ocean models and is thought to be a factor in SIA models (helping to determine where radiation is absorbed). Ocean colour measurements provide a means to estimate transparency.
- Clouds. Poor representation of clouds remains a key weakness of most SIA models. Better data are needed to improve parameterisations but these needs are adequately specified under NWP and elsewhere.
- Aerosols data such as volcanic ash is also required. Continuity of satellite observations of volcanic aerosols is important.
- Stratospheric ozone concentration data might be of interest in the future for seasonal forecasting.

ANNEX TO PARAGRAPH 5.3

Proposed Amendments to *Manual on the GDPFS* - Appendix II-6

Minimum list of LRF products to be made available by global scale producing centres

1. Forecast Products

Note: it is recognized that some centres may provide more information than the list including for example daily data or hindcast data.

Basic properties

Temporal resolution.
~~**Monthly averages/accumulations for a season**~~**.**

Averages, accumulations or frequencies over 1-month or longer periods (seasons)

Spatial resolution.
2.5° x 2.5° (note: selected to match resolution of current verification data)

Spatial coverage. **Global**
(separate areas of interest to users, down to sub-regions of a continent or ocean basin, may be provided on special request from Members)

Lead time. ~~**0-4 months for seasonal forecasts.**~~ **Any leadtimes between 0, and 4 months (**definition of lead time: for example, a three-monthly forecast issued on 31 December has a lead time of 0 months for a January-to-March forecast, and a lead time of 1 month for February-to-April forecast, etc.)

Issue frequency. **Monthly or at least quarterly**

Output types. Either rendered images (eg forecast maps and diagrams) or digital data.~~Gridded numerical values, area-averaged values and indices, and/or images~~. GRIB-2 format should be used for products posted on FTP-sites or disseminated through the GTS.

Indications of skill including hindcast **should be provided**, in accordance with recommendations from CBS on the Standardised Verification System (Attachments II-8). The minimum required is level 1 and level 2 verification. The verification of Nino3.4 index will only apply to those centres producing such indices. However GPCs are encouraged to provide level 3 verification. Verification results over the hindcast period are mandatory.

Content of basic forecast output: (some products are intended as directly meeting NMS requirements with regard to information needed for end-user applications [direct or further processed]; others are to assist the contributing global centres in product comparison and in the development of multimodel ensembles. These products are regarded as feasible from current systems).

A. **Calibrated outputs from ensemble prediction system showing the mean and spread of the distribution for:**
   - **2 metre temperature over land**
   - **sea surface temperature**
   - **precipitation**
   - **Z500, MSLP, T850**

Notes:
1. These fields are to be expressed as departures from normal model climate.
~~**2. SST used as boundary conditions for (two-tiered) AGCM predictions should be made available.**~~

B. **Calibrated probability information for forecast categories.**

- **2 metre temperature over land**
- **SST (Atmospheric coupled models only)**
- **Precipitation**

Notes:

- B is the minimum requirement. A should be provided, at least, by request.
- **Tercile categories should be provided**, consistent with present capabilities. Information for larger numbers of categories (e.g. deciles) is foreseen, however, as capabilities increase and to match better the anticipated end-user requirements. These targets are implied also for forecasts from statistical/empirical models.
- **Information on how category boundaries are defined should be ~~included.~~made available.**
- "Calibrated" implies correction based on systematic errors in model climatology, using at least 15 years of retrospective forecasts.

ANNEX TO PARAGRAPH 5.5

ACTIVITIES AS A GLOBAL PRODUCING CENTRE FOR LONG-RANGE FORECASTS

Long-Range Forecast and Related Products, Capacity Building and Training
WMC Melbourne/Australian Bureau of Meteorology

Submitted by
David Jones (d.jones@bom.gov.au) *& Andrew Watkins* (a.watkins@bom.gov.au)

Operational Long-Range forecast material

- Sea-surface temperature forecasts from the Predictive Ocean Atmosphere Model for Australia (POAMA) coupled atmosphere-ocean general circulation model (version 1).
  Daily updated plumes for NINO 1, 2, 3, 3.4 and 4.
  Daily updated spaghetti diagrams for Sea Surface Temperature anomalies.
  Daily updated histograms of likely anomalies, providing probabilities of warm/neutral/cool conditions in the Pacific.
- 3-month seasonal climate outlook products (Australia only).
  Rainfall, maximum and minimum temperatures, analogue years. Distributed through the media/website and printed booklet.
  Extensive hindcast verification and forecast validation of all seasonal outlook products.
- South Pacific Seasonal Outlook Reference Material.
  A monthly report emailed to Pacific Island NMHSs which represents a collation of Bureau and other seasonal forecast information and background observational material.
- Participation in the Island Climate Update process with NIWA (NZ).
- A survey of 12 international ENSO forecast models.
  Prepared operationally. Summary used in media releases, and made available through web.
- Outgoing Long Wave Radiation and intraseasonal predictions based on the MJO.
  A substantial project is underway to unify statically based intraseasonal and seasonal climate forecasting methodologies.
- The ENSO Wrap-up
  An operationally produced overview of the status of El Niño/La Niña, based on latest forecasts and observational material
  Produced fortnightly, and made available through the web.
- Contributions to the WMO EL Niño alerts

All of this information can be accessed at the following two websites:
http://www.bom.gov.au/climate
http://www.bom.gov.au/bmrc

Experimental Long-Range forecast material available

- Land surface temperature and rainfall fields from POAMA coupled atmosphere-ocean general circulation model (version 1) – POAMA1.
  9-month forecasts updated daily.
  Other fields such as heights, pressures, oceanic heat content etc.
  Above products in a standard format – netCDF. Can be convert to GRIB etc as needed.
- Extensive hindcast verification results are available and a range of technical and scientific reports have been written. However verification is not to the SVS standard.

- Operationally maintained – 24/7 – but fields not deemed sufficiently accurate for general distribution.

All products are calibrated relative to the model's climate calculated as a function of lead-time and start month, using hind-casts. The current research website is http://www.bom.gov.au/bmrc/ocean/JAFOOS/POAMA.

Shorter Term Goals: 1-3 years

Launch of Predictive Ocean Atmosphere Model for Australia coupled atmosphere-ocean general circulation model (version 2) – POAMA2 (see Annex). This offers the possibility of *accurate* direct model predictions of rainfall/temperature etc. It is expected that this system will be verified according to the SVSLRF recommendations.

- Skill assessment of short-term products as required by WMO and users (eg. ROC scores, % consistent scores etc)
- Forecasts of global atmospheric variables: 2m temperature, precipitation, Z500, MSLP, T850, SOI + NAO indexes, etc
- Down-scaling/statistical correction of POAMA2 output to provide location specific forecasts.

Longer Term Goal: 3-5 years.

Development has commenced on the new Australian Community Climate Earth System Simulator (ACCESS). This system is being developed as a collaborative project between the Australian Bureau of Meteorology, CSIRO, The Australian Greenhouse Office, and national universities as an integrated modelling system for nowcasting through to climate change time predictions. A multidisciplinary team of scientists is being assembled for this project, while model components are being ported from the U.K. Meteorological Office (Atmospheric Model and 4DVAR Assimilation) and Australian Bureau of Meteorology and CSIRO (AusCom Ocean Model, Sea Ice Model and Land Surface/Carbon Cycle Model). It is planned that this modelling system will move into real-time operational service in approximately 2-3 years.

Capacity Building and Training

Pacific Island Climate Prediction Project (PICCP)

An Australian Bureau of Meteorology – Australian AID (AUSAID) project delivered through partnership with Meteorological Services in Fiji, Tuvalu, Tonga, Samoa, Cook Island, Nuie, Kiribati, Solomon Islands, and Vanuatu. Seen the development of a PC based "*down-scaling*" tool for delivering tailored Seasonal Climate Outlooks for Pacific Island Countries called SCOPIC (Seasonal Climate Outlooks for Pacific Island Countries). SCOPIC includes verification, historical data analysis, and an automated text generation tool for producing public seasonal outlook statements.

Involved very extensive in-country training, stakeholder workshops, and now stakeholder projects with the Fiji Sugar Industry, Solomon Islands Water Authority, and Media (Vanuatu, Tonga, Cook Islands).

Australia is exploring, in collaboration with other Pacific Basin countries, the various options for implementing a distributed system for providing climate information, including Long-Range forecast material, throughout the Pacific. These matters are being taken up within the contexts of various bi-laterals and the WMO Regional Climate Centres framework.

Other relevant activities include a Pacific Data Rescue Project, and developing plans to become a southwest Pacific GCOS lead-centre. Negotiations continue with various partner and funding agencies which have been very positive. It is expected that the successful PICCP activities will be rolled into a more permanent capacity building commitment for the South Pacific.

ANNEX

POAMA Development plans - POAMA-2

***Purpose/objectives*:**

- Improve forecasting of El Niño
- Provide ensemble set of hind-casts
- Preliminary trial rainfall products
- Preliminary trial intra-seasonal products

Timescale

- Operational mid 2006, large hind-cast set 2006/7

Ocean model

- ACOM2 (some retuning to reduce noise in the East Pacific)

Ocean data assimilation

- Reanalyse 1979-present, fix bugs in forcing and use new ENACT forcing (cures some SST biases)
- Use new ocean data assimilation scheme based on ensemble multi-variate optimum interpolation
- Assimilate salinity data
- Make salinity increments based on temperature data and dynamical adjustment
- Introduce the direct assimilation of in-situ SST data (optional)

Atmosphere model

- Use new version BAM4
- Increase resolution T63/95L34/60? (increase boundary layer res. and lift top of atmosphere into stratosphere). Final vertical and horizontal resolution being finalised
- Use new physics, SES, EC land surface and boundary layer, new gravity wave etc
- Tunning/testing presently being done

Atmospheric initialisation

- Use GASP for real-time
- Use ERA-40 for hind-casts (note: POAMA-1 used AMIP)

Coupler

- OASIS 3.4 (2.4 in POAMA-1)

Land surface

- Use new ECMWF land surface scheme with variable surface parameters (POAMA-1 used bucket model)
- Basic land surface initialisation procedure will be developed Options: (a) AMIP style (b) ERA-40 for hind-casts and GASP in real-time (c) used land-surface model in standalone with observed precipitation and surface forcing. Aiming for option (c) but (a), (b) are fallback positions if resources not found.

Hind-casts

- 5-10 member ensemble per month over the period 1980-2005, ensembles to be generated by varying the start date of the initial conditions by a day

Real-time forecast system

- As POAMA-1, option to increase ensemble size to around 50.

Products/output

- Provide range of products for downscaling and user applications, contribution to international projects COPES and to the APCN. Need to ensure output is consistent with international standards - e.g. COORDS compliant, and provide enough data to force regional models. Work on downscaling/bridging as part of SEACI.

The Australian Bureau of Meteorology

www.bom.gov.au

Long-range Forecasting (LRF) Progress Report

January 2004 – December 2005

1. A brief summary of research and development connected with applications and main operational changes in LRF related issues

i. A dynamic coupled seasonal forecast system – Predictive Ocean Atmosphere Model for Australia (POAMA) is now operational and is the basis for seasonal forecasts for Sea Surface Temperatures (SSTs) in the NINO-3, 3.4 and 4 areas on a daily basis: http://www.bom.gov.au/climate/coupled_model/poama.shtml
ii. A new version of this forecast system (POAMA-2) is currently being developed with the aim of experimental implementation in mid-2006. This will include, besides improved forecasting of El Niño, a trial seasonal rainfall prediction system.
iii. The development and implementation of an experimental intra-seasonal prediction system.
iv. The provision of seasonal outlook verification tools on the internet for hindcasts period and for current forecasts.

2. Research and development in LRF specialised data processing

As outlined in the previous report (WMO/TD-No. 1279), and briefly summarised here, the current POAMA system uses near real-time ocean analyses for input. Each day the ocean state is integrated forward one day using the ocean model. The ocean model is forced with six-hourly fields from the Bureau of Meteorology GASP Numerical Weather Prediction (NWP) system. Every three days observations, including subsurface observations, are assimilated into the ocean model. Surface temperature observations are not assimilated. Instead the ocean model surface temperature is relaxed to the SST analysis field used in the GASP system with an e-folding time scale of 3 days. Every day a 9-month coupled model forecast is produced in real-time using the very latest ocean state and the latest atmospheric state from the GASP NWP analysis. The daily forecasts are combined to form a 30-member monthly ensemble and an ensemble of the last 30 forecasts (updated daily on the POAMA web site).

One feature of the POAMA system is the ability of the atmosphere model to represent the MJO. This, together with the use of real-time ocean and atmospheric data, means that POAMA can also produce forecasts of intra-seasonal variability out to a few weeks lead-time. Experimental intra-seasonal forecasts are now available on the POAMA web site and are updated daily (http://www.bom.gov.au/bmrc/ocean/JAFOOS/POAMA).

The second generation of POAMA (POAMA-2) is due for experimental release in mid 2006. It is likely that the POAMA-2 system will increase the number of forecasts made each month and a 25-year hindcast period will be constructed. Other changes will include the introduction of a stand-alone land surface initialisation scheme and modification to the ocean assimilation scheme, which will be further detailed in Section 4.

3. Outstanding research and development activities related to the LRF specialised analysis forecast system in operational use in the previous year

i. As outlined in the previous report (WMO/TD-No. 1279), one of the current seasonal forecast systems in use for seasonal forecasting in Australia is the Predictive Ocean Atmosphere Model for Australia (POAMA). This first version (POAMA-1) is a seasonal to interannual forecasting system, which was jointly developed by the Australian Bureau of Meteorology Research Centre (BMRC) and CSIRO Marine Research (CMR). This system is now operational and is updated daily (http://www.bom.gov.au/climate/coupled_model/poama.shtml). The current focus of the POAMA-1 system is the prediction of tropical SST anomalies and these are presented in the form of a forecast plumes line graph (e.g. Figure 1), distribution plots (e.g. Figure 2) and horizontal spatial plots. All forecast anomalies are calculated relative to the model hind-cast climatology over the period 1987-2001, using all the hind-casts starting at the same time of the year.

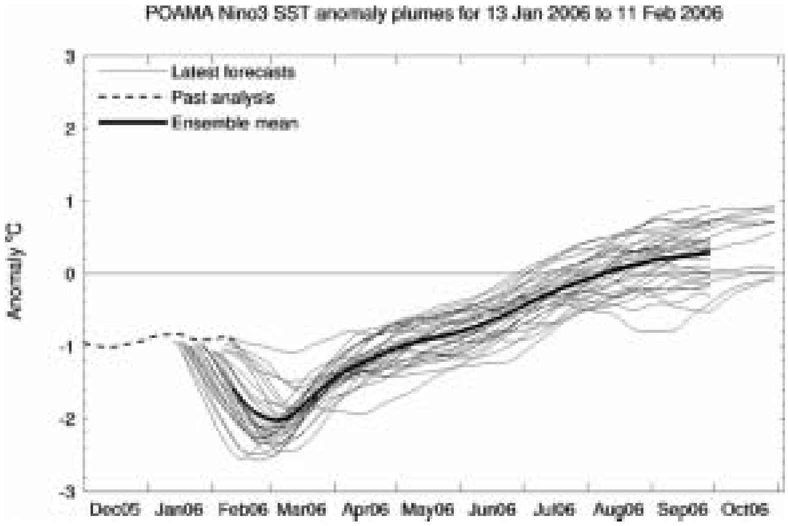


Figure 1. Forecast **SST anomalies in the NINO3 region. The latest 30 daily forecasts are shown (as available on day of writing, the start dates of the forecasts are shown in the plot heading). The dashed line is an estimate of the observed SST anomaly from the BMRC ocean analysis system. The dark line is the ensemble mean. All anomalies are relative to the 1987-2001 period.**

Components of the POAMA-1 system are:

a Atmosphere model: The POAMA-1 system uses the latest version of the Bureau of Meteorology unified atmospheric model (BAM version 3.0d). It uses a modified convection closure that allows the model to have a good representation of the MJO. It has a horizontal spectral resolution of T47 and has 17 vertical levels. The performance of this model forced with observed SST is described in Colman et al., (2005).

b. Ocean Model: The ocean model component is ACOM2. It was developed by CMR, and was based on the Geophysical Fluid Dynamics Laboratory Modular Ocean Model (MOM version 2). The grid spacing is 2 degrees in the zonal direction. The meridional spacing is 0.5° within 8° of the equator, increasing gradually to 1.5° near the poles. There are 25 levels in the vertical, with 12 in the top 185 metres. Technical details of ACOM2 are given in Schiller et al., 1997 and Schiller et al., 2002.

c. Coupler: The ocean and atmosphere models were coupled using the Ocean-Atmosphere-Sea Ice-Soil (OASIS) coupling software (developed by CERFACS, France; Valcke et al., 2000).

d. Ocean data assimilation: The ocean data assimilation scheme is based on the optimum interpolation (OI) technique described by Smith et al., (1991). Only temperature observations are assimilated and only measurements in the top 500m are used. There are several improvements over the scheme described by Smith et al., (1991). The OI scheme is used to correct the model background field every 3 days using a 3 day observation window, one and a half days either side of the assimilation time. Ocean current increments are calculated by applying the geostrophic relationship to the temperature corrections, similar to the method described by Burgers et al., (2002).

e. Atmospheric initial conditions: For the real time forecasts the atmospheric component is initialised with weather analyses from the Bureau of Meteorology's operational NWP system (GASP). This means that the seasonal forecast model knows about the latest intra-seasonal variability in the tropical atmosphere.

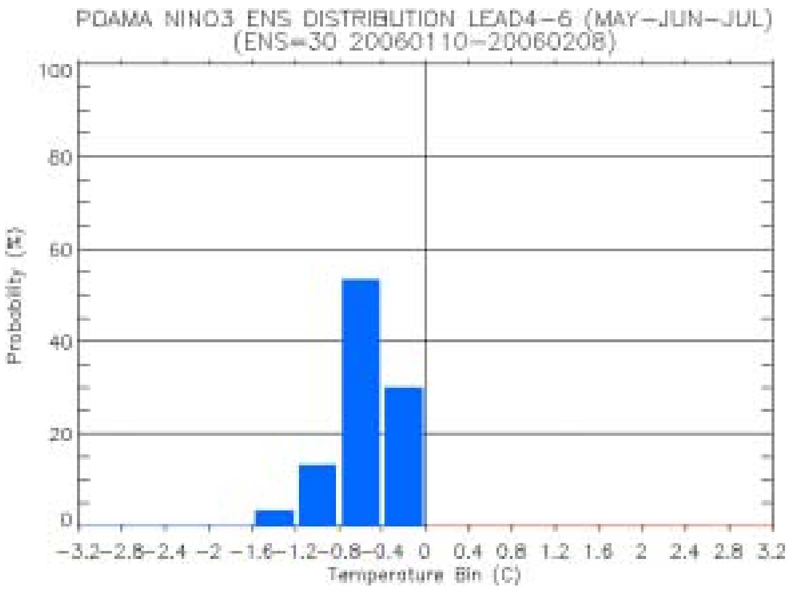


Figure 2. NINO3 SST anomaly distribution for lead times 4-6 months. These show the percentage of ensemble members lying within the specified 0.4°C temperature anomaly bins.

ii. One of the main areas of new research within the POAMA-1 system during the 2004-2005 period has been in the area of intraseasonal prediction. The tropical MJO has been investigated for its use as a predictor at this sub-LRF range (i.e. less than 30 day lead). Aside from forecasts made with the coupled model, a seasonally independent index for monitoring the MJO has been developed based on a pair of empirical orthogonal functions (EOFs) of the combined fields of near-equatorially averaged 850 hPa zonal wind, 200 hPa zonal wind and satellite-observed outgoing longwave radiation (OLR) data (Wheeler and Hendon 2004). The pair of principal component (PC) time series that form the index vary mostly on the intraseasonal time scale of the MJO only and are called the Real-time Multivariate MJO series 1 (RMM1) and 2 (RMM2). These indices are able to closely monitor the current state and position of the MJO (see Figure 3). Another application of the RMM series is through prediction of onset dates of the Australian and Indian monsoons, as onset, whilst occurring at any time during the convectively enhanced phase of the MJO cycle, rarely occurs during the suppressed phase. Another application is through the prediction of extreme weekly rainfall: in an examined area around Darwin in northern Australia the probability of receiving an upper-quintile weekly rainfall event increased threefold from the dry to the wet MJO phase (Wheeler and Hendon 2004).

Experimental forecasts of climate and weather variations related to the MJO using the RMM indices are currently available: http://www.bom.gov.au/bmrc/clfor/cfstaff/matw/maproom/RMM/index.htm

In collaboration with researchers from Queensland Department of Primary Industries and Fisheries, RMM indices are used to provide forecast guidance to Northern Australian farmers. Guidance based on the state of the MJO and links to the MJO forecast is also included in the Australian Bureau of Meteorology's "Weekly Tropical Climate Note": http://www.bom.gov.au/climate/tropnote/tropnote.shtml

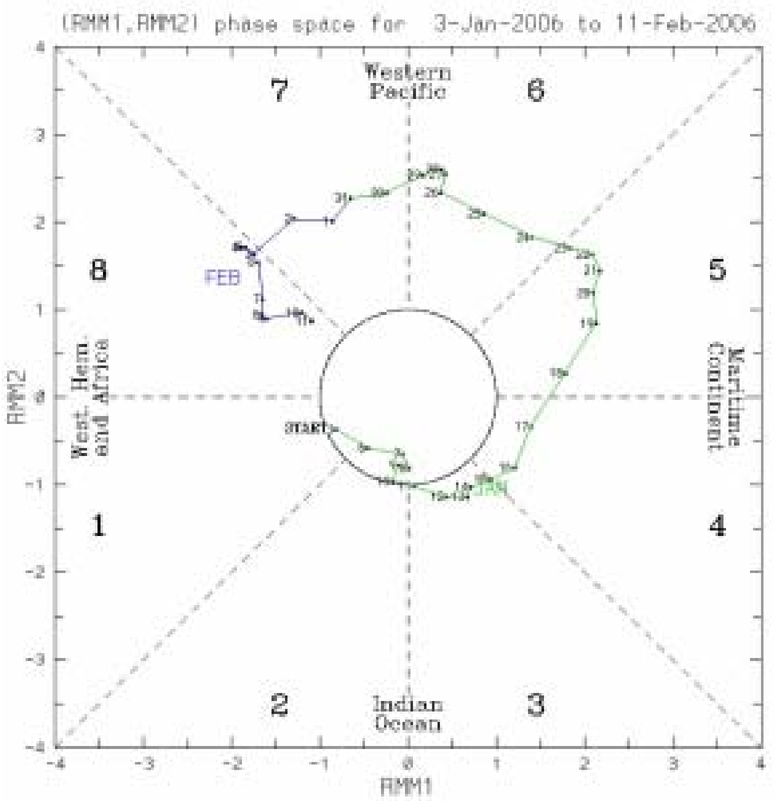


Figure 3. Phase diagram of RMM1 and RMM2 from observations for the latest 40 days (at time of writing). The circle represents one standard deviation, with any projection outside the circle indicative of possible MJO activity in that sector. The distance from the circle indicates the strength of the event, with points further away indicating stronger events.

4. Plans for future research and development activities related to the improvement of the LRF oriented operational system

i. The second generation of POAMA (POAMA-2) is due for experimental release mid-2006. A significant improvement for POAMA-2 is the atmospheric model. There are improvements in both horizontal and vertical resolutions and the introduction of several new physical parameterization schemes.

A new ocean data assimilation scheme has been developed. The multivariate ensemble OI scheme of Oke et al., (2005) is being used as the basic analysis technique. Model error covariances are taken from an ensemble of short forecasts performed each analysis time (Alves and Robert, 2005). The ensembles are generated by perturbing the surface forcing fields using random estimates of the forcing errors. These will provide time and flow dependent multivariate error statistics. These statistics will be used to generate salinity and current corrections based on the temperature observations. In addition, salinity data will also be assimilated into the ocean model.

A large set of hind-casts will be produced with POAMA-2. Present plans are for at least a 10-member ensemble each month over the past 25 years. The hind-casts will be initialised with true atmospheric initial conditions from the ERA40 re-analysis. This means that the intra-seasonal forecast skill can also be assessed. The POAMA model is able to simulate variability characteristic of the MJO and the POAMA-1 real-time forecasts showed some skill in forecasting intra-seasonal variability. Intra-seasonal forecast products will be generated from POAMA-2 and their skill assessed.

ii. Longer range plans for climate forecasting in Australia include the development of the Australian Community Climate Earth-System Simulator (ACCESS), which is a coupled climate and earth system simulator to be developed as a joint initiative of the Australian Bureau of Meteorology and CSIRO in cooperation with the university community in Australia.

5. Development in verification procedures including performance statistics

i. During 2004-2005, statistical based seasonal forecasts were issued by the Australian Bureau of Meteorology for rainfall and maximum and minimum temperatures in the Australian region. Forecasts indicate the probabilities of receiving wetter/warmer than average rainfall/temperatures over a three month period. This forecast is based on the statistical relationship of Australian climate variables with sea surface temperatures in the tropical Pacific and Indian Oceans. These forecasts are available via the internet at: http://www.bom.gov.au/climate/ahead

Results of cross validation of a hindcast period are currently provided for the seasonal forecasts of rainfall and temperature. Results are displayed as a map representing percent consistent values for each forecast period. Tables of individual long range forecasts are then available for any location in Australia: http://www.bom.gov.au/silo/products/verif/
In the first half of 2006, this site will be updated to show the results of operational forecasts. Other verification methods include LEPS skill scores (see example in Figure 4), Brier Scores and Reliability diagrams. The forecast scheme has also been extensively documented and verified in Fawcett et al. (2005).

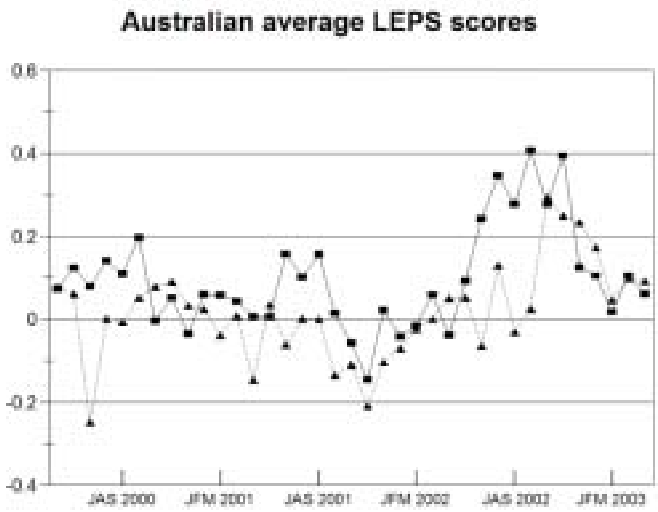


Figure 4. Australian average LEPS skill score (MAM 2000 to MAM 2003) for seasonal maximum temperature (solid line with squares) and seasonal minimum temperature (dashed line with triangles). Positive averages are seen throughout most of the 2002/03 El Niño. Periods of low skill generally correspond to periods of low forecast signal.

ii. Hindcasts from POAMA-1 have been used to measure the forecast skill via the anomaly correlation skill for NINO 3 SST anomalies (see Figure 3 in the previous report: WMO/TD-No. 1279). The forecast skill for a dipole of heat content in the Indian Ocean, an important indicator of the state of the Indian Ocean, is shown in Figure 5 with the anomaly correlation of the forecast beating that of persistence at lead times of approximately three to nine months.

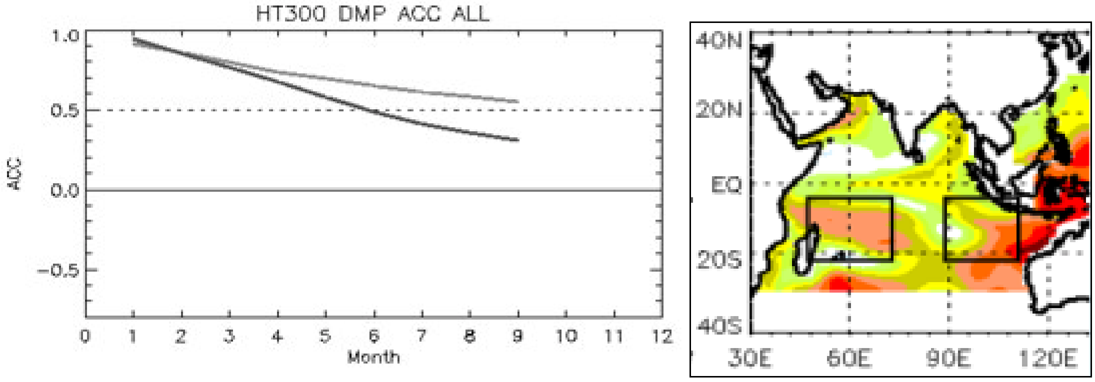


Figure 5. The heat content anomaly correlation, for a box in the west Indian Ocean minus a box in the east, as a function of lead time for 60 forecasts, one per season, starting during the period 1987-2001. Light grey – POAMA coupled model, Dark grey – persistence. The area of the two boxes are given in the figure on the right.

iii. The MJO RMM forecast system discussed in Section 3 and forecasts from the POAMA model have been included in 'The MJO Experimental Prediction Project', run by the Climate Diagnostics Centre with the aim: "*To provide a method to access and compare MJO forecasts, and to analyze the effects of MJO events on tropical and mid-latitude weather forecasts*" (http://www.cdc.noaa.gov/MJO/). For 11-day forecasts of the tropical velocity potential anomaly, the simple MJO statistical-based RMM forecasts have been shown to be as skillful as an ensemble of forecasts from a modern NWP model (NCEP's ensemble GCM).

6. Publications

Alves, O., Balmaseda, M.A., Anderson, D. and Stockdale, T. 2004. Sensitivity of dynamical seasonal forecasts to ocean initial conditions. Q. J. Roy. Met. Soc., 130, 647-67.

Colman, R. 2004. Coupled modelling in BMRC: present status and future directions. BMRC Research Report No. 98, Bur. Met. Australia.

Fawcett, R.J.B and D.A. Jones and G.S. Beard, 2005: A verification of publicly issued seasonal forecasts issued by the Australian Bureau of Meteorology: 1998-2003. *Aust. Met. Mag*., **54**, 1-13.

Wang, G., Alves, O and Smith N. 2005. BAM3.0 tropical surface flux simulation and its impact on SST drift in a coupled model. BMRC Research Report No. 107, Bur. Met. Australia.

Wheeler, M.C., and H.H. Hendon, 2004: An all-season real-time multivariate MJO Index: Development of an index for monitoring and prediction. *Mon. Wea. Rev.*, 132, 1917-1932.

7. References

Alves, O. and C. Robert, 2005. Tropical Pacific ocean model error covariances. Submitted to Q. J. Roy. Met. Soc.

Burgers, G., M.A. Balmaseda, F.C. Vossepoel, G. van Oldenborgh, P. van Leeuwen, 2002: Balanced Ocean-Data Assimilation near the Equator. Journal of Physical Oceanography: Vol. 32, No. 9, pp. 2509-2519.

Colman, R., L. Deschamps, M. Naughton, L. Rikus, A. Sulaiman, K. Puri, G. Roff, Z. Sun and G. Embery, 2005. BMRC Atmospheric Model (BAM) version 3.0: comparison with mean climatology'. BMRC Research Report No. 108, Bur. Met. Australia.

Fawcett, R.J.B and D.A. Jones and G.S. Beard, 2005: A verification of publicly issued seasonal forecasts issued by the Australian Bureau of Meteorology: 1998-2003. *Aust. Met. Mag*., **54**, 1-13.

Oke, P. R., A. Schiller, D. A. Griffin, G. B. Brassington, 2005. Ensemble data assimilation for an eddy-resolving ocean model of the Australian region. Submitted to Quart. J. Roy. Met. Soc.

Schiller, A., J. S. Godfrey, P. C. McIntosh, G. Meyers, N. R. Smith, O.Alves, G. Wang and R. Fiedler, 2002: A New Version of the Australian Community Ocean Model for Seasonal Climate Prediction. CSIRO Marine Research Report No. 240.

Schiller, A., J. S. Godfrey, P. McIntosh and G. Meyers, 1997: A global ocean general circulation model climate variability studies. CSIRO Marine Research Report No. 227. 60 pp.

Smith, N. R., J. E. Blomley and G. Meyers, 1991. A univariate statistical interpolation scheme for subsurface thermal analyses in the tropical oceans. *Prog. Oceanog*., **28**, 219-256.

Valcke, S., L. Terray and A. Piacentini, 2000. OASIS 2.4 Ocean Atmospheric Sea Ice Soil users guide, Version 2.4. CERFACS Tech. Rep, CERFACS TR/CMGC/00-10,85pp.

Wheeler, M.C., and H.H. Hendon, 2004: An all-season real-time multivariate MJO Index: Development of an index for monitoring and prediction. *Mon. Wea. Rev.*, 132, 1917-1932.

Activities as a Global Producing Centre for Long-Range Forecasts

National Centers for Environmental Prediction (NCEP) response to WMO requirements to be recognized as a Global Producing Center (GPC) for long-range forecasts:

Background: Dynamical seasonal forecast at NCEP are produced based on the Climate Forecast System (CFS) that includes comprehensive ocean-atmosphere-land models and a data assimilation system for initialization. Seasonal forecasts from the CFS will be the basis for NCEP to seek WMO's designation as one of the Global Producing Center for long-range forecasts. NCEP's point-by-point responses to the criterion that need to be met, and be designated as a GPC, are included below.

(a) Fixed production cycles and time of issuance: The Climate Forecast System at NCEP is an operational system and follows a fixed production cycle. In its present configuration, the CFS is run twice daily generating a forecast for a 10 month lead time. A comprehensive set of hindcasts from 1981-2005 is also available to assess the performance of the CFS, and if required, for calibration of real-time forecasts. Daily data for selected fields, and their monthly means, are archived and are made available to the users.

(b) Provide a limited set of products as determined by the revised Appendix II-6 of the Manual on the GDPFS: List of variables in Appendix II-6 of the Manual on the GDPFS includes monthly means of: 2-meter temperature over land; sea surface temperature; precipitation; 500-mb heights, mean sea level pressure; and 850-mb temperatures. Some other requirements for the variables included are: spatial resolution is $2.5^o x 2.5^o$; spatial coverage is global; lead time of forecasts is 0-4 months. The current output from the CFS already meets all the requirements listed in Appendix II-6 of the Manual on the GDPFS.

(c) Provide verifications as per WMO Standard Verification System for Long-Range Forecasts (SVSLRF): NCEP is currently working on complying with the requirements for the WMO Standard Verification System for Long-Range Forecasts. Analysis is expected to be completed within six months (September 30, 2006) and results will be submitted to the lead center for SVSLRF.

(d) Provide up-to-date information on methodology used by the GPC: Comprehensive up-to-date information for the CFS methodology and forecast configuration is provided at:

http://cfs.ncep.noaa.gov/
http://www.cpc.ncep.noaa.gov/products/people/wwang/cfs_fcst4CPC/

Any further information can be provided, and included, at relevant websites.

(e) Make products accessible through the GPC web-site and/or dissimilated through the GTS and/or Internet: Graphical products (including some verification statistics) from the CFS are currently available at:

http://www.cpc.ncep.noaa.gov/products/people/wwang/cfs_fcst4CPC/

Forecast data can also be downloaded from the NWS ftp serves, and relevant information is already available from:

http://cfs.ncep.noaa.gov/

To summarize, forecasts from the NCEP's Climate Forecast System are currently very close to be in compliance with the WMO requirements for designation as a GPC for the long-range forecast. The only missing factor is the submission of verifications per WMO standards to lead center for the verification, and this step will be completed within the next six months.

Met Office, UK
Annual Research Progress Report on Long-Range Forecasting
January 2004 - December 2005

R.J. Graham, B.D. Becker, A. Brookshaw, A.W. Colman, M.K. Davey, C.K. Folland, M. Gordon, M. Huddleston, S. Ineson, B. Ingleby, M. MacVean, P.J. McLean and A.A. Scaife

Met Office, FitzRoy Road, Exeter, UK

0. Current status with regard to WMO criteria for designation as a Global Producing Centre (GPC) for Long-range Forecasting

The current status of current long-range forecasting activities at the Met Office is detailed in the following sections 1 and 2. In this section a summary of progress is provided with regard to the GPC designation criteria (a) to (e).

a) Fixed production cycles and time of issuance.
A fixed monthly production cycle is operated. The precise time of issuance is under review. Currently forecasts are issued in the last week of each calendar month.

b) Provide a limited set of products as determined by the revised Appendix II-6 of the Manual on the GDPFS.
The majority of products listed in the manual are provided (Type B products (probabilities for tercile categories) have been provided since July 2004).
http://www.metoffice.gov.uk/research/seasonal/monthly_forecasts/single_terce2.html

c) Provide verifications as per the WMO SVS for LRF.
SVS verification diagnostics are provided on the Met Office website and provision of these to the Lead Centre for SVSLRF has now started.
http://www.metoffice.gov.uk/research/seasonal/monthly_forecasts/skill.html

e) Provide up-to-date information on methodology used by the GPC
This information is provided on the Met Office website and has been submitted to the Lead Centre of SVSLRF.
http://www.metoffice.gov.uk/research/seasonal/user_guide.html

f) Make products accessible through the GPC website and/or disseminated through the GTS and/or internet.
Products are made available (since January 1998) mainly via the Met Office website.
http://www.metoffice.gov.uk/research/seasonal/index.html

1. Dynamical prediction systems and products

1.1 Dynamical monthly prediction
Monthly forecast services to a range of users have continued as in previous years. Forecasts are generated using operational output from the ECMWF coupled ocean-atmosphere 51-member monthly-range ensemble system (Vitart, 2003). The model is run weekly from initial conditions at 00GMT Thursday. A hindcast dataset, with the same start time and valid period as the forecast, is available ahead of each forecast using a 5-member ensemble. For forecast calibration, the Met Office's post-processing uses a rolling 12-year hindcast period, ending with the year prior to the forecast year.

Met Office post-processing is performed for mean, maximum and minimum temperature, precipitation and sunshine amount averaged/accumulated over three forecast periods; days 5-11 ahead, days 12-18 ahead and days 19-32 ahead (for the UK region, forecasts for the 5-11 day period are generated using the ECMWF 10-day EPS). Forecast products include probability forecasts for various regions of the globe with additional focus on the 10 UK climate districts. Global probability products are provided in the form of 1) probability maps for tercile and outer-quintile categories of temperature and precipitation, and 2) for specific regions, probability histograms for quintile categories (well-below, below, near, above, and well-above the climate normal for the region and time of year). For the 10 UK climate districts temperature and rainfall forecasts are generated in terms of quintile categories. Tercile categories are used for sunshine. The UK forecasts are expressed both in terms of the probability of each category, and a deterministic forecast based on either the ensemble mean or the most probable quantile.

1.2 Dynamical seasonal prediction

As in previous years, seasonal forecasts to 6-months ahead have been generated each month using the Met Office's 41-ensemble coupled ocean-atmosphere global seasonal prediction system (known as GloSea). GloSea is based on the HadCM3 climate model. A performance assessment of the GloSea system is provided by Graham et al., 2005. Operational forecasts are initialised with ocean and atmosphere conditions valid for the first day of the current month. Perturbations to the initial conditions are applied to the ocean component only and are based on 5 parallel ocean assimilations, generated through application of perturbed windstress. Additional instantaneous SST perturbations are applied at initial time to generate the 41 starting states required for the ensemble. The forecasts run on the ECMWF computing facility in parallel configuration with the ECMWF system2 seasonal prediction model as part of a developing European multi-model system (the European Seasonal to Interannual Prediction Project – Euro-SIP).

GloSea forecasts are expressed relative to a model climatology defined for each month of the year from a set of 15-member ensemble integrations initialised at the beginning of each month over the 15-year period 1987-2001. A range of forecast products are made available to NMSs, Regional Climate Outlook Fora, UK government agencies, the public and commercial companies. In 2005, a major upgrade of the Met Office seasonal forecasting web pages was released. Products now available include the following. Forecasts for anomalies in 3-month-average 2-metre temperature and precipitation, at one-, two- and three-month leads - corresponding to months 2-4, 3-5 and 4-6 of the integration. A probabilistic format is used giving probabilities for equi-probable tercile categories and also for two outer-quintile categories (20^{th} and 80^{th} percentiles). In addition to these probability products, maps indicating the most probable tercile category are also provided. Forecast products for monthly-mean Sea Surface Temperature anomalies in the tropical Pacific are also made available. Products may be viewed at www.metoffice.gov.uk/research/seasonal. Verification information indicating forecast performance has been generated, using WMO guidelines, and is available on the website. Verification diagnostics used include ROC curves, ROC score maps and reliability diagrams. On the website, forecasts from the GloSea system may be compared with corresponding forecasts generated using output from the Euro-SIP multi-model forecast database, which currently includes forecast ensembles from the Met Office, ECMWF and Météo-France seasonal systems. Currently Met Office products derived from Euro-SIP comprise an unweighted combination of the Met Office GloSea forecast ensemble and the ECMWF system2 seasonal ensemble.

1.3 Empirical and hybrid empirical/dynamical real-time seasonal prediction

As in previous years, hybrid statistical and dynamical prediction schemes were used to make seasonal forecasts for selected regions of special interest including the East Africa Short rains season (October-December), the west

African Monsoon season (July-September) and the NE Brazil wet season (March-May) . The statistical schemes use historical relationships between key sea surface temperature patterns and surface meteorological conditions. For these regions, information from both statistical models and the GloSea coupled model ensemble was combined to obtain best-estimate seasonal rainfall forecasts, which were contributed to Regional Climate Outlook Fora and distributed to National Meteorological Services and drought monitoring agencies in the target regions. Forecasts were also contributed to Regional Climate Outlook Fora and published in the Experimental Long-Lead Forecast Bulletin. Statistical forecasts of July-August temperature tercile probabilities for western Europe were issued on the Met Office website in March and updated in July.

Warm summers were correctly predicted for NW Europe in 2004 and 2005 and a drier than average short rains season was correctly predicted for East Africa in 2005. Our forecasts for the Sahel have indicated above average rainfall. Whilst observed rainfall has been generally higher here than in the drought years of the 1980s and early 1990s, it has not been as wet as predicted. An explanation for this could be that the relationship between inter-hemispheric contrast in SST and rainfall, which is a major contributor to our statistical forecasts, has changed. Possible reasons for such a change include the impact of climate change and altered ecology of the Sahel region related to trends in land use. Improvements to the forecast system to take account of such changes are being investigated. Our GLOSEA model correctly predicted a slightly drier than average season in NE Brazil in 2005, but unpredicted mid season changes in South Atlantic SST anomalies were the likely cause of a poor forecast for 2004.

Forecasts have also been issued, on a monthly basis, for rainfall in the Volta river catchment in West Africa and for inflow into Lake Volta, Ghana. This forecast application was developed with the Volta River Authority Ghana, and forecasts are used to assist management of hydro-electric power generation. A particular challenge in this region is that rainfall anomalies of opposing signs are often observed between the north and south of the catchment, and this was a particularly common feature in 2004 and 2005. The total rainfall, and consequent lake inflow, is thus dependent on the positioning and relative strengths of each pole of the dipole. Slightly below average rainfall and inflow for the 2005 peak season (July-October) was successfully predicted but, except for one longer-lead forecast, predictions for 2004 were too dry.

The Met Office has issued, in June each year, a long range forecast of the state of the winter North Atlantic Oscillation (NAO) averaged over the December to February winter period. In June 2004 the forecast was for a weakly positive NAO of +0.5 s.d. and this was realised in the 2004/5 winter when the NAO index was +0.1 s.d.

This year the forecast was also used to infer a north European winter temperature anomaly. This is possible because of the strong influence of the NAO on winter temperature over the European region (Scaife et al., 2005). The method used is based on the previous May’s Atlantic sea surface temperature (Rodwell and Folland, 2002, 2003), giving a forecast with a lead time of about six months. By projecting the monthly mean SST fields for May onto a predefined North Atlantic tripole pattern, the method correctly predicts the sign of the following winter NAO in two out of three cases. We have also now established that the method predicts the correct sign of winter central England temperature anomalies and Northern European winter temperature anomalies in 2 out of 3 years. Furthermore, new ocean heat content analyses show that some of the years in which hindcasts by this method have failed can be explained by the occurrence of ENSO events which produce additional European signals, confounding the forecast. The statistics of the method are therefore improved in non-ENSO years. Finally, the method has also been shown to be more skillful than dynamical model predictions for the European region in winter. For winter 2005/6 the method indicated colder than average conditions over Northern Europe (10W-50E, 50-70N) and a strongly negative NAO of -1.1 s.d. Details are available at the following web site:

http://www.metoffice.com/research/seasonal/regional/nao/index.html.

Output from the method was combined with dynamical model forecasts and observations of the evolving North Atlantic surface and sub-surface temperature anomalies to generate Met Office statements on the prospects for winter 2005/6, first issued in August 2005 and updated monthly. The forecast stated a 2 in 3 chance of a colder-than-average winter for much of Europe and that, if this were to hold true, parts of the UK - especially southern regions – would have temperatures below normal. The forecast and its likely impacts were communicated to users in the Utilities, Finance and Insurance, Defence, Aviation and Transport sectors as well as to local authorities and regional resilience planners and charities, and gained widespread media attention.

1.4 Forecast of annual mean global surface temperature in 2004 and 2005

Each December the Met Office issues a forecast of the global mean surface temperature anomaly (i.e. a combination of global land surface air temperature and global sea surface temperature) for the year ahead. The forecast uses a statistical method that includes a variety of natural and anthropogenic forcing factors, the state of the Atlantic Multidecadal Oscillation (Knight et al, 2005) and a coupled model forecast of the state of El Nino in the first part of the year ahead. In December 2003 we issued a forecast for a global surface temperature anomaly for 2004 of 0.50+-0.12^{o}C (Folland and Colman, 2003) where the uncertainty encompasses the 95% confidence range. The observed global temperature anomaly was 0.44+-0.06 ^{o}C (95% confidence range estimate) calculated using an optimum average (Folland et al, 2001) of the HadCRUT2v data set (Jones and Moberg, 2003). So the outcome of this forecast was well within its uncertainty range. In December 2004 the forecast for 2005 was for 0.51+- 0.12 ^{o}C (Folland and Colman, 2004). The observed optimally averaged global temperature anomaly for 2005 is still being assessed, but is clearly warmer than 2004.

2. Dynamical prediction studies, model calibration and validation

Validation studies have found that, in common with other models, GloSea forecast probabilities for outer-tercile, outer-quintile and outer-decile events exhibit a bias relative to the observed frequency of the events. In general the bias indicates ‘over confidence’, such that when the chance of observing the event is relatively high, forecast probabilities are too large, and when the chance of observing the event is relatively low forecast probabilities are too small. Such biases may be corrected by calibration techniques that use the statistics of past performance. The merits of a number of such techniques have been contrasted, and the best overall method of those tested found to be discriminant analysis. Discriminant analysis may also be used to combine, in an optimal way, ensemble output from the different CGCM components of the Euro-SIP multi-model system. Investigations of (discriminant) calibrated multi-model products show the main potential for unbiased probabilistic prediction of outer-quintile and outer-decile events is in tropical regions. It is planned to release calibrated probability forecast products on the seasonal forecast website in 2006.

Further improvements were made to the statistical downscaling methods developed to improve long-range forecast skill for UK climate districts. It has been shown that the strategy of using dynamical model forecast data from non-local grid points produces significant gains in skill for monthly-average temperature forecasts at ranges of a few months. Work to perform additional assessments for categorical probability forecasts is in progress.

Investigations into dynamical seasonal predictability, with an extension to inter-annual and decadal timescales are continuing as part of the ENSEMBLES project. ENSEMBLES is an Integrated Project under the 6th Framework Programme (FP6) of the European Union (EU), coordinated by the Met Office. The central project theme is the development of multi-model ensemble-based probabilistic prediction of climate and its impacts from seasonal to decadal and longer (century) timescales. The project started on 1st September 2004 and will run for 5 years. The seasonal to decadal component of ENSEMBLES builds on the FP5 project DEMETER, which demonstrated the superior skill available at seasonal timescales from representation of uncertainties in model formulation through use of a multi-model ensemble comprising 7 European CGCMs. DEMETER also pioneered the integration and testing of user application/impact models with ensemble CGCM output, specifically for crop yield and health applications.

In addition to other improvements, ENSEMBLES models will include, for the first time, realistic concentrations of green house gases, solar forcing and (at initialisation time) volcanic dust. The Met Office DePreSys system, designed specifically for decadal prediction, will also be included in the ENSEMBLES multi-model. In addition to the multi-model approach, alternative (or complimentary) techniques for representing model uncertainties will be investigated. In this respect the Met Office is investigating the benefits of a perturbed parameter technique in which an ensemble is generated by using perturbed versions of the CGCM physics to generate each member. The perturbed model versions are constructed by using different settings (from within a plausible range) for a number of tuneable physics parameters. The method has been previously developed and used to generate ensemble-based probabilistic predictions of climate change (Murphy et al., 2004). Initial conditions for the ENSEMBLES multi-model will be generated using improved ocean analysis techniques and observation datasets formulated as part of the FP5 ENACT project and further developed in ENSEMBLES. Techniques for representing initial condition uncertainty will also be compared. As part of the development and assessment phase of the ENSEMBLES system the GloSea model has been updated to allow realistic green house gas concentrations and volcanic dust (at initialization time), and hindcasts have been run for the 11-year period 1991-2001. Experiments using both the operational method of ensemble initialization (see Section 1.2) and a lagged start method have been conducted. The integrations are made in 9-member ensembles from May and November start dates out to at least 12 months ahead. Runs from May 1965 and 1994 have been integrated to 10years ahead. An extended1960-2001 hindcast set will be employed in final hindcast production starting in 2007.

References

Folland, C.K. and A. Colman, 2003: Empirical prediction of the global temperature anomaly for 2004. *COLA, Experimental Long Lead Forecast Bulletin, **12**, (on line). Dec 2003*

Folland, C.K., Rayner, N.A., Brown, S.J. Smith, T.M. Shen, S.S. Parker, D.E., Macadam, I., Jones, P.D., Jones, R.N., Nicholls, N. and Sexton, D.M.H., 2001: Global temperature change and its uncertainties since 1861. *Geophysical. Research Letters*, **106**, 2621-2624.

Jones, P.D. and A. Moberg, 2003: Hemispheric and large-scale surface air temperature variations: an extensive revision and an update to 2001. *J. Climate,* **16**, pp.206-223.

Murphy, J.M., Sexton, D.M.H., Barnett, D.N., Jones, G.S., Webb, M.J., Collins, M., Stainforth, D.A. 2004: Quantification of modelling uncertainties in a large ensemble of climate change simulations. Nature 430, 768-772.

Rodwell, M.R. and Folland, C.K., 2002: *Quarterly Journal of the Royal Meteorological Society,* **128**, 1413-1443.

Rodwell, M.R. and Folland, C.K., 2003: *Annals of Geophysics,* **46**, 47-56.

Vitart, F., 2003: Monthly forecasting system. ECMWF Technical memorandum No. 424. 68pp.

Internal and External Publications in 2004/05 (non-Met Office authors indicated '*'):

Colman, A.W., 2004: Combined Statistical/Dynamical Forecast of 2004 Season Rainfall in the Sahel and Other Regions of Tropical North Africa using Observed Ocean and Atmosphere Information from up to June 2004. Experimental Long-Lead Forecast Bulletin 13 No. 2 (2004). http://grads.iges.org/ellfb/Jun04/colman/colman.htm

Colman, A.W., 2004: Forecast of East African Rainfall for October-December 2004 using Dynamical and Statistical Methods. Experimental Long-Lead Forecast Bulletin 13 No. 3 (2004). http://grads.iges.org/ellfb/Sep04/colman/colman.htm

Colman, A.W. and Graham R.J., 2004: Forecast of North East Brazil Seasonal Rainfall for February to May 2005 using Empirical and Dynamical Methods and Atmosphere and Ocean Data up to November 2004. Experimental Long-Lead Forecast Bulletin 13 No. 4 (2004). http://grads.iges.org/ellfb/Dec04/colman/colman.htm

Davey, M., Huddleston, M., Ingleby, N., Ineson, S. and *project partners, 2005: ENACT project: final report to the European Commission

Folland, C.K. and A. Colman, 2004: Empirical prediction of the global temperature anomaly for 2005. *COLA, Experimental Long Lead Forecast Bulletin, **13**, (on line). Dec 2004*

Graham, R.J., Gordon, M., McLean, P.J., Ineson, S., Huddleston, M.R., Davey, M.K., Brookshaw, A. and Barnes, R.T.H. 2005: A performance comparison of coupled and uncoupled versions of the Met Office seasonal prediction General Circulation Model. Tellus, 57A, 320-339.

Huddleston, M.R., Bell, M.J., Martin, M.J. and Nichols, N.K., 2004: Assessment of wind stress errors using bias corrected ocean data assimiliation. Q.J.R. Meteorol. Soc., 130, 853-871.

Huddleston, M. and Ingleby, B., 2005: ENACT: a 43-year ocean analysis set. Met Office internal (GMR) report 2005.

Ineson, S. and Dong, B., 2005: Report on HadGEM's suitability for long-range forecasting. Met Office internal (GMR) report 2005, V.C.1(04/05).

Ingleby, B. and Huddleston, M.: Quality control of ocean temperature and salinity profiles - historical and real-time data; paper submitted to Journal of Marine Systems

Knight, J., Allan R.J., Folland, C.K., Vellinga, M. and M.E. Mann*, 2005: Natural Variations in the thermohaline circulation and future surface temperature. *Geophysical. Research Letters,* **32**, L20708, doi: 1029/2005GL024233.

*Palmer, T. N., *Alessandri, A., *Andersen, U., *Cantelaube, P., Davey, M., *Delecluse, P., *Deque, M., *Diez, E., *Doblas-Reyes, F., *Feddersen, H., Graham, R., *Gualdi, S., *Gueremy, J.-F., *Hagedorn, R., *Hoshen, M., *Keenlyside, N., *Latif, M., *Lazar, A., *Maisonnave, E., *Marletto, V., *Morse, A., *Orfila, B., *Rohel, P., *Terres, J.-M., and *Thomson, M., 2004: **Development of a European multi-model ensemble system for seasonal to interannual prediction (DEMETER). Bull. Amer. Met. Soc., 85, 853-872.**

Scaife, A.A., Knight, J.R., *Vallis G.K. and Folland C.K., 2005: *Geophysical Research Letters*, **32**, L18715.

Smith D.M., A.W. Colman, S. Cusack, C.K. Folland, S. Ineson and J.M. Murphy: Predicting surface temperature for the coming decade using a global climate model. Submitted.

National Climate Center of CMA (NCC/CMA)
Annual Progress Report on Long-Range Forecasting (LRF) in 2004-2005
Activities response to WMO requirements to be recognized as a Global Producing Center (GPC) for long-range forecasts:

Dynamical climate prediction is produced at NCC/CMA based on a dynamical climate model prediction system (DCMPS) which is developed during 1996-2003 and launched into experimental operation run in 2004, and began to routinely run for monthly and seasonal forecasts from Jan. 2005 on the new platform of High Performance Computer. This system mainly consists of the Global Atmosphere-Ocean Coupled Model for Seasonal Prediction and its oceanic data assimilation system. The current status of long range forecasting activities at NCC/CMA is summarized below with regard to WMO criteria for designation as a Producing Center (GPC):

a) Fixed production cycles and time of issuance

The DCMPS at NCC/CMA is operational running for monthly and seasonal prediction as follow:

·Monthly (30-day) prediction:

Based on its monthly dynamical extended range forecast model (DERF), NCC runs its DERF model once per day with 8 member (4 for LAF and 4 for SVD) and conducts ensemble forecast every 5days (with 40 members at most), and issues the monthly (30-day) global prediction in the first day of every pentad, i.e., 1st, 6th, 11th, 16th, 21st and 26th, the prediction period includes 1-10day, 11-20 day, 21-30 day and 31-40 day, 1-30 day and 11-40 day for monthly prediction. The variables include the precipitation, 2-meter temperature, geopotential height (200hPa, 500hPa and 700hPa), sea level pressure, zonal and meridional wind (200hPa and 700hPa). For temperature and precipitation, we issue both the determined prediction and the probabilistic prediction by 3 tercile.

·Seasonal (90-day) prediction:

Based on its coupling general circulation model (CGCM), NCC real time forecast is released around 25th each month for following 0-6month (2 seasons future), which are produced from initial conditions near the end of the previous month. Total number of forecast ensemble members is 48, using 8 atmospheric initial conditions and 6 oceanic initial conditions. The 8 atmospheric initial conditions are taken from each of the last 8 days of the end of previous month. The 6 oceanic initial conditions are from a single initial state with different perturbations of ocean data assimilation system. The variables of products include the precipitation, 2-meter temperature, 500hPa geopotential height, 850hPa temperature, sea surface temperature. For temperature and precipitation, we issue both the determined prediction and the probabilistic prediction. (1) Forecast length is 7 months.

b) Provide a limited set of products as determined by the revised Appendix II-6 of the Manual on the GDPFS;

There are three temporal resolutions of NCC/CMA prediction products: 10-day mean, monthly averages and accumulations for a season. And the spatial resolution of the global prediction products is 2.5˚×2.5˚ to match the verification data.

NCC/CMA's prediction products cover the whole global and the leading time of

the seasonal forecasts varies from 0 to 6 months (in near future expend to 0-11month).
The products could be provided in the netCDF format, which can be posted on FTP-sites or disseminated through the GTS or the Internet.
The retrospective forecasts (or hindcasts) starting Jan 1983 are used for calibration and skill evaluation both for monthly and seasonal forecasts. The hindcasts are configured exactly the same way as the real-time forecasts.

c) Provide verifications as per the WMO standard verification system for long-range forecast (SVSLRF);
From 2005, NCC started to verify its CGCM forecasts in terms of the standardized verification system for long-range forecasts (SVSLRF) of WMO as specified in the new attachment II-8 to the Manual on the Global Data Processing and Forecasting System (WMO-No. 485), Volume I. The verification datasets used are the Xie-Arkin CMAP precipitation and the ERA40 re-analysis of surface air temperature anomaly at screen level (T2m). In the ROC score for the three equi-probable forecast categories (above normal, near normal and below normal) for the probabilistic summer precipitation anomaly forecasts show that NCC's seasonal forecast has a higher skill score in the tropical regions than in the extra-tropics in both the above normal and the below normal categories, especially in the East Asian. And the score in the abnormal categories are greater than that of the normal counterpart. The completed verification results are expected to be provided within 6 months and will be submitted to the lead center for SVSLRF.

d) Provide up-to-date information on methodology used by GPC;
Up-to-date information for the dynamical climate prediction methodology and forecast configuration is provided and will be enriched at:
http://bcc.cma.gov.cn/Website/index.php?ChannelID=29
http://bcc.cma.gov.cn/CSMD/Website/index.php?WCHID=3

e) Make products accessible through the GPC Web-site and/or disseminated through the GTS and/or Internet
Now, the comprehensive global prediction figures can be easily accessed at http://bcc.cma.gov.cn/products/en_md.php?WCHID=29&ChannelID=63 for monthly forecast and
http://bcc.cma.gov.cn/products/en_cs.php?WCHID=29&ChannelID=64 for seasonal forecast, and forecast data can be downloaded at NCC's website also.

Beside, a combined dynamical and statistical prediction system has been established in NCC. The empirical/statistical models of climate prediction have been developed on the basis of physical factors, such as SST, land surface condition on the Tibetan Plateau, monsoon, blocking high and subtropical high. The methods of interpreting dynamic product to produce global LRF products also have been developed at NCC to improve the predictability of global LRF.

This document reports on the Long range Forecasting activity at ECMWF

1. Introduction

Since 1997, ECMWF issues global seasonal predictions routinely every month. In 2000 the seasonal forecasts became part of the operational products and by mid-2000 the seasonal forecast products became available to the all the WMO members. The ECMWF seasonal forecast is a dynamical system consisting of a coupled atmosphere-ocean model and an ocean analysis. A brief description of the operational system is given in section 2. In sections 3 and 4 products and verification are discussed. Section 5 describes envisaged future implementations. Relevant points referring to the requirements specified in the CBS-XIII document for being recognized as a Global Producing Centre for Long range forecast are highlighted in Italic.

2. Long-Range Forecasting system

ECMWF Seasonal Forecast system is fully operational. ***The operational release date for the forecast is set at 12UTC on the 15th of each month.***

The current seasonal forecast system was introduced into operational use at the beginning of 2002. The atmospheric component is CY23R4 of the IFS with a horizontal resolution of TL95 and 40 levels in the vertical. This is the same cycle of the IFS as was used in the ERA40 re-analysis. The ocean model resolution is 0.3 degrees meridionally near the equator and to 1 degree x 1 degree at higher latitudes with 29 vertical levels. The ocean initial conditions are provided not from a single ocean analysis but from a 5-member ensemble of ocean analyses. The analyses differ in that a measure of uncertainty in the surface winds used to force the ocean is taken into account.

The ensemble ocean analysis is part of the new method of ensemble generation in the operational seasonal forecast system. Each ensemble forecast consists of 40 members all with initial conditions on the 1st of the month. The ensemble forecast's design aims to represent the most important uncertainties in the initial conditions. Uncertainties in SST values are represented by 40 different SST perturbations added to the 5 ocean analyses in order to create a 40-member set of ocean initial conditions from which the forecasts are launched. In addition, stochastic physics is used to perturb the coupled integrations throughout the forecast period. This gives a significant de-correlation of the atmospheric flow in the tropics in the first few days of the forecast, compensating for the fact that perturbations to the atmospheric initial conditions are not included. The 40-member ensemble can be run once the ocean analyses are available, generally on the 11th of each month. Because a large amount of computation is involved, and to ensure reliable delivery, the operational release date for the forecast is set at the 15th of the month. This is still a big improvement in timeliness over the original system.

As with all models, the seasonal forecast system is not perfect. One symptom of this is climate drift: the model climatology does not match the observed climatology. To account for this, the forecasts need to be referenced to the model climatology.
The estimate of the model climatology is based on an ensemble of 5 integrations spanning the years 1987- 2001. This 15 year climate gives a more stable basis for computing anomalies than the 6 year climate available in the original system. ***A comprehensive and up-dated description of the operational system is publicly available on the web at : http://www.ecmwf.int/products/forecasts/seasonal/documentation/index.html***.

3. Products

Data dissemination to Member States and commercial customers through Internet and dissemination of predicted SST monthly data to GTS are part of the operational schedule. Since mid-2000 a large part of graphic products displayed on the ECMWF web-pages are available to the WMO members. Those products include: a weekly monitoring of the oceanic conditions, the Nino plumes, spatial charts, and forecast of tropical storms. All plots can be downloaded as postscript or pdf files, as well as being viewed on screen. *Global spatial maps with horizontal resolution (2.5X2.5) of: 2 metre temperature,sea surface temperature, precipitation, 500 hPa height, temperature at 850hPa and mean sea level pressure are shown, in the form of probabilities for tercile and 15%ile categories as well as the ensemble mean anomaly and the probability of exceeding the climate median. These fields are expressed as departures from normal model climate.* The Nino SST indices include the Nino 3.4 and Nino 4 regions as well as Nino 3, and the ocean analysis plots include several meridional sections, as well as zonal and horizontal maps. The forecast lead time indicated on the web at:
http://www.ecmwf.int/products/forecasts/d/charts/seasonal/forecast/charts/ is different from the one suggested by the CBS document.

As part of the operational products ECMWF issues also predictions in the extended range using the Monthly forecast system (for more information on the monthly forecast please see http://www.ecmwf.int/products/forecast).

A large number of different model fields from seasonal forecast (both forecast and hind-cast) is archived although only a small subset of these are presently listed in the 'ECMWF catalogue' for commercial use. A full list of the output fields can be found in section 3 of the online Seasonal Forecast User Guide, at http:/www.ecmwf.int/products/forecasts/seasonal/documentation. The comprehensive data archive allows the development of a full range of sophisticated products, and in particular the synoptic variability of each ensemble member is well resolved. The upper air and surface fields should be sufficient for statistical downscaling techniques, including those that require the synoptic evolution of the system. The archive does not include the full model level data that would be required to drive regional dynamical models, since to store the full global fields for all ensemble members would be excessive. Ocean analysis data are also archived. For further details see:
http://www.ecmwf.int/products/forecasts/seasonal/documentation/ch3_2.html

4. Verification

For a correct interpretation of seasonal predictions the user needs to complement the forecast products with knowledge of the forecast skill. For this purpose ***the site at:***
http://www.ecmwf.int/products/forecasts/d/charts/seasonal/verification provides a comprehensive documentation of skill levels, using methods that have been agreed at the international (WMO) level for the evaluation of long-range forecast systems. A suite of verification scores for deterministic (e.g. spatial anomaly correlation and Mean Square Skill Score (MSSS)) and probabilistic forecasts Relative Operating Characteristics (ROC curves and areas) can be viewed. *2m temperature, precipitation, and Sea Surface temperature anomalies, requested by the SVS, are among the parameters that are verified. The SVS level 1 consists of a large scale aggregated overall measures of forecast performance. These are bulk numbers calculated by aggregating verification at grid points over three regions: Tropics, Northern extra-tropics and Southern extra-tropics. On the web these aggregated values are either displayed in the graphic form or appeared in the title of the plot. ROC curve, ROC areas, reliability and frequency histograms are part of SVS level 1. ROC curve and ROC area are provided while reliability and frequency histograms are not available at the moment. The SVS level 2 verification, consisting of graphic representation of grid point values of: MSSS with its 3-term decomposition and ROC area, are provided.*

Although we can take advantage of the experience in the medium range forecast verification, evaluating seasonal forecast skill involves dealing with a generally small signal to noise ratio and limited sample of cases. Significance testing methods are therefore particularly relevant and this is something we hope will be increasingly reflected in the verification statistics provided to the users. A constant effort is devoted to enhance/improve the verification information. By next year the full set of scores from level 1 and 2 will be sent to the lead Verification Centre in Melbourne.

5. Future implementations

In the second half of 2006 the seasonal forecasting system will be upgraded. The new system will use a more recent version of the atmospheric component with increased horizontal and vertical resolutions. A more advance version of the ocean assimilation system will be introduced: salinity and scatterometer data will be assimilated and a bias correction will be applied. The ensemble generation will include atmospheric perturbations based on the singular vector computation. With the new system the hind-cast period will be extended to 1981-2005 and the hind-cast ensemble size will be 11 members. This 25 year climate will provide a more stable basis for computing anomalies than the 15 year climate available in the present system.

As for the previous upgrade, the new operational system will run in parallel to the original system for sometime in order to guarantee a smooth transition. During this year the future operational changes were presented and discussed at the annual meeting with users of medium range and extended range products and at the 10th Workshop on Meteorological Operational Systems. Documentation of the operational system will be updated accordingly when the new system will be implemented.

In addition products from the EUROSIP (European Seasonal to Inter-annual Predictions) multi-model system will be published on the web. EUROSIP is a real-time multi-model ensemble forecasting and is part of the operational seasonal forecast suite at ECMWF. At the moment ECMWF, Met Office and Météo-France coupled systems are part of this multi-model mix but it is possible that other models may be included at a later stage. The Met Office coupled system (GloSea) is initialised using an ensemble of ocean analyses and it runs on the ECMWF computing facility in parallel configuration with the ECMWF system seasonal prediction model. Météo-France is running their system remotely using the ECMWF computer. Météo-France system is initialised using the Mercator ocean analyses. Météo-France hind-cast set is limited by the availability of Mercator ocean analysis that extend back to 1993. Also, the ensemble generation strategy is different, and is based on a lagged average approach with cross-matching between ocean and atmosphere dates.

The initial multi-model product will be based on a simple average of the 3 probabilities associated with each individual system. At a later stage Bayesian techniques, based on work that has been done at ECMWF, will be extended to the multi-model. Raw (non calibrated) data will still be accessible. Estimates of the multi-model skill levels and of the skill of each individual component will be made available. Such estimates will be evaluated using methods that have been agreed by the SVS for LRF. Documentation of the multi-model products and of each individual component will be also available.

ANNEX TO PARAGRAPH 8.1.2

New Attachment II-8 to the *Manual on the GDPFS* (WMO-No. 485), Volume I

Standardised Verification System (SVS) for Long-Range Forecasts (LRF)

STANDARDISED VERIFICATION SYSTEM (SVS) FOR LONG-RANGE FORECASTS (LRF)

Table of contents

Standardised Verification System (SVS)
for Long-Range Forecasts (LRF)

Executive Summary

1. Formulation

The SVS is formulated in five parts:

1.1 Diagnostics. Information required incorporates derived diagnostic measures and contingency tables. Estimates of the statistical significance of the scores achieved are also required. Additional diagnostic measures are suggested but are not incorporated into the Core SVS as yet. Use of the additional diagnostics is optional.

1.2 Parameters. Key variables and regions are proposed. However producers are not limited to these key parameters, thus all producers can contribute regardless of the structure of individual forecast systems. The parameters to be verified are defined on three levels:

Level 1: Diagnostic measures aggregated over regions and for indices
Level 2: Diagnostic measures evaluated at individual grid-points
Level 3: Contingency tables provided for individual grid-points.

The SVS makes provision for a staged implementation of the three levels of information and the inclusion of estimates of skill significance over a two year period.

1.3 Verification data sets. Key data sets of observations against which forecasts may be verified are proposed.

1.4 System details. Details of forecast systems employed.

1.5 Exchange of verification information

The SVSLRF verification results are made available through a web site maintained by the Lead Centre. The functions of the Lead Centre for SVSLRF include creating and maintaining coordinated Web sites for the LRF verification information so that potential users would benefit from a consistent presentation of the results. The address of the web site is http://www.bom.gov.au/wmo/lrfvs/.

2. Diagnostics

Three diagnostic measures are incorporated in the Core SVS - Relative Operating Characteristics, reliability diagrams and accompanying measure of sharpness and Mean Square Skill Scores with associated decomposition. Estimates of the statistical significance in the diagnostic scores are also included in the Core SVS. The three diagnostics permit direct intercomparison of results across different predicted variables, geographical regions, forecast ranges, etc. They may be applied in verification of most forecasts and it is proposed that, except where inappropriate, all three diagnostics are used on all occasions. Tabulated information at grid-point resolution is also part of the core SVS. The tabulated information will allow reconstruction of scores for user defined areas and calculation of other diagnostic measures such as economic value.

2.1 Relative Operating Characteristics. To be used for verification of probability forecasts. For Level 1 information (measures aggregated over regions) the ROC curve and the standardized area under the curve (such that perfect forecasts, give an area of 1 and a curve lying along the diagonal gives 0.5) should be provided. For Level 2 information (gridded values) the standardized area under the ROC curve should be provided.

2.2 Reliability diagrams and frequency histograms. To be used in assessment of probability forecasts. They are required as part of the Level 1 information only.

2.3 Mean Square Skill Score and decomposition. To be used in verification of deterministic forecasts. For Level 1, an overall bulk MSSS value is required and will provide a comparison of forecast performance relative to "forecasts" of climatology. The three terms of the MSSS decomposition provide valuable information on phase errors (through forecast/observation correlation), amplitude errors (through the ratio of the forecast to observed variances) and overall bias. For Level 2, quantities pertaining to the three decomposition terms should be provided. Additional terms relating to MSSS are required as part of the Level 3 information.

2.4 Contingency tables. In addition to the derived diagnostic measures contingency table information provided at grid-points for both probability and categorical deterministic forecasts form part of the core SVS. This information constitutes Level 3 of the exchange and will allow RCCs and NMHSs (and in some cases end-users) to derive ROC, reliability, other probability based diagnostics and scores for categorical deterministic forecasts for user defined geographical areas.

A number of recommended contingency table-based diagnostics are listed. The Hanssen-Kuipers score is the deterministic equivalent to the area under the ROC curve, and thus provides a useful measure for comparing probabilistic and deterministic skill. The Gerrity score is one recommended score for overall assessment of forecasts using two or more categories.

3. Parameters

The key list of parameters in the Core SVS is provided below. Any verification for these key parameters should be assessed using the Core SVS techniques wherever possible. Many long-range forecasts are produced which do not include parameters in the key list (for example, there are numerous empirical systems that predict seasonal rainfall over part of/or over an entire, country). The Core SVS diagnostics should be used to assess these forecasts also, but full details of the predictions will need to be provided.

Forecast can be made using different levels of post-processing typically no-post-processing (raw or uncalibrated), simple correction of systematic errors (calibrated, i.e. calibration of mean and of variance) and more complex correction using hindcast skill (recalibrated, e.g. Model Output Statistics or perfect prog approaches). Most centres are currently issuing forecasts resulting from a simple calibration and so for sake of comparison on the Lead Centre web site scores for forecasts that were raw or calibrated (as specified in respective skill score section) are to be submitted. At the moment the team prefer to exclude forecast that were recalibrated, but GPCs are encouraged to apply the SVSLRF methodology and to display the results on their recalibrated forecasts on their web site.

3.1 Level 1: Diagrams and scores to be produced for regions

Diagrams (e.g. ROC and reliability curves) are to be supplied in digital format as specified on the Lead Centre website.

3.1.1 Atmospheric parameters. Predictions for:

T2m Screen Temperature anomalies with standard regions:
- Tropics 20°N to 20°S
- Northern Extratropics >=20°N
- Southern Extratropics <=20°S

Precipitation anomalies with standard regions:
Tropics 20°N to 20°S
Northern Extratropics >=20°N
Southern Extratropics <=20°S

3.1.2 Scores and diagrams to be produced for probabilistic forecasts
Reliability diagram and frequency histograms
The ROC curve and the standardised area under the curve.
Estimations of error (significance) in the scores.

The above scores and diagrams to be produced for equi-probable tercile categories.

3.1.3 Scores to be used for deterministic forecasts
Mean Square Skill Score (MSSS) with climatology as standard reference forecast.

3.1.4 Stratification by season
Four conventional seasons MAM, JJA, SON, DJF

3.1.5 Lead-time
Preferred minimum: 2 lead-times, one preferably to be 2-weeks or greater, with lead-time not greater than 4 months.

3.2 Level 2: Grid point data for mapping

3.2.1 Grid point verification data to be produced for each of the following variables. Verification should be provided on a 2.5°x2.5° grid.
T2m
Precipitation
SST

3.2.2 Verification parameters to be produced for deterministic verification

The necessary parameters for reconstructing the MSSS decomposition, the number of forecast/observation pairs, the MSE of the forecasts and of climatology and the MSSS are all part of the core SVS. Significance estimates for the correlation, variance, bias, MSE and MSSS terms should also be supplied.

3.2.3 Verification to be provided for probability forecasts

ROC area for three tercile categories. Significance of the ROC scores should also be provided.

3.2.4 Stratification by season

If available twelve rolling 3-month periods (e.g. MAM, AMJ, MJJ). Otherwise four conventional seasons (MAM, JJA, SON, DJF).

3.2.5 Lead-time

Preferred minimum: 2 lead-times, one preferably to be 2-weeks or greater, with lead-time not greater than 4 months.

3.2.6 Stratification according to the state of ENSO.

Stratification by the state of ENSO should be provided if sufficient ENSO events are contained within the hindcast period used. Scores should be provided for each of three categories:

- All hindcast seasons
- Seasons with El Niño active
- Seasons with La Niña active

3.3 Level 3: Tabulated information to be exchanged

Tabular information to be provided for grid points of a 2.5x2.5 grid.

3.3.1 Contingency tables

Contingency tables to be produced for verifying forecasts of tercile categories in each of the following variables

T2m
Precipitation
SST

3.3.2 Tables to be produced for probabilistic forecast verification

The number of forecasts hits and false alarms to be recorded against each ensemble member or probability bin for each of three equi-probable categories (terciles). It is recommended that the number of bins remain between 10 and 20. The forecast providers can bin according to percentage probability or by individual ensemble members as deemed necessary. No latitude weighting of the numbers of hits and false alarms is to be applied in the contingency tables.

The user is encouraged to aggregate the tables over grid-points for the region of interest and to apply methods of assessing statistical significance of the aggregated tables.

3.3.3 Tables to be produced for deterministic forecasts

3x3 contingency tables comparing the forecast tercile with the observed tercile, over the hindcast period.

3.3.4 Stratification by season

If available twelve rolling 3-month periods (e.g. MAM, AMJ, MJJ). Otherwise four conventional seasons (MAM, JJA, SON, DJF).

3.3.5 Lead-time

Preferred minimum: 2 lead-times, one preferably to be 2-weeks or greater, with lead-time not greater than 4 months.

3.3.6 Stratification according to the state of ENSO

Stratification by the state of ENSO should be provided if sufficient ENSO events are contained within the hindcast period used. Scores should be provided for each of three categories:

All hindcast seasons

Seasons with El Nino active
Seasons with La Nina active

3.4 Verification for indices (Level 1)

3.4.1 Indices to be verified

Niño3.4 region SST anomalies. Other indices may be added in due course.

3.4.2 Scores to be calculated for probabilistic forecasts

ROC area for 3 tercile categories. Where dynamical forecast models are used the ROC scores should be calculated for the grid-point averaged SST anomaly over the Niño3.4 region. It is recommended that significance of the ROC scores should also be calculated.

3.4.3 Scores to be calculated for deterministic forecasts

The three terms of the Murphy decomposition of MSSS, produced with climatology as standard reference forecast. As a second, optional, control it is recommended that damped persistence be used. Significance estimates should accompany each of the three terms.

Where dynamical models are used the MSSS decomposition should be calculated for the grid-point averaged Niño3.4 anomaly.

3.4.4 Stratification by month

Verification should be provided for each calendar month.

3.4.5 Lead-time

Verification for each month should be provided for 6 lead times. Namely zero-lead and leads of 1-month, 2-months, 3-months, 4-months and 5-months. Additional lead times are encouraged if available.

4. Staged implementation

In order to ease implementation, producers may stage the provision of the elements of the Core SVS according to the following recommendation.

a) Verification at levels 1 and 2 in the first year of implementation
b) Verification at level 3 by the middle of the year following implementation of levels 1 and 2
c) Level of significance by the end of the year following implementation of levels 1 and 2.

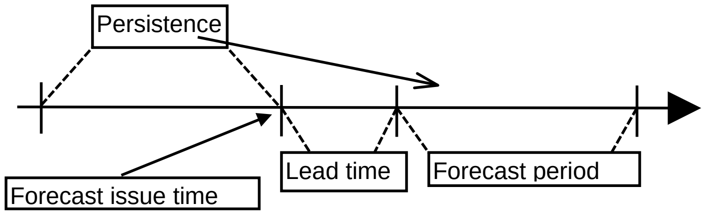


Figure 1: ***Definition of forecast period, lead time and persistence as applied in a forecast verification framework.***

Standardised Verification System (SVS)
for Long-Range Forecasts (LRF)

1. Introduction

The following sections present the detailed specifications for the development of a Standardised Verification System (SVS) for Long-Range Forecasts (LRF) within the framework of a WMO exchange of verification scores. The SVS for LRF described herein constitutes the basis for long-range forecast evaluation and validation, and for exchange of verification scores. It will grow as more requirements are adopted.

2. Definitions

2.1. Long-Range Forecasts

LRF extend from thirty (30) days up to two (2) years and are defined in Table 1.

Table 1: Definition of long-range forecasts.

Monthly outlook:	Description of averaged weather parameters expressed as departures from climate values for that month.
Three-month or 90-day 'rolling season' outlook:	Description of averaged weather parameters expressed as departures from climate values for that three-month or 90-day period.
Seasonal outlook:	Description of averaged weather parameters expressed as departures from climate values for that season.

Seasons have been loosely defined in the Northern Hemisphere as December-January-February (DJF) for winter (summer in the Southern Hemisphere), March-April-May (MAM) for spring (Fall in the Southern Hemisphere), June-July-August (JJA) for summer (winter in the Southern Hemisphere) and September-October-November (SON) for Fall (spring in the Southern Hemisphere). Twelve rolling seasons are also defined e.g. MAM, AMJ, MJJ. In the Tropical areas, seasons may have different definitions. Outlooks over longer periods such as multi-seasonal outlooks or tropical rainy season outlooks may be provided.

It is recognised that in some countries long-range forecasts are considered to be climate products.

This attachment is mostly concerned with the three-month or 90-day outlooks and the seasonal outlooks.

2.2. Deterministic Long-Range Forecasts

Deterministic LRF provide a single expected value for the forecast variable. The forecast may be presented in terms of an expected category (referred to as categorical forecasts, e.g. equiprobable terciles) or may take predictions of the continuous variable (non-categorical forecasts). Deterministic LRF can be produced from a single run of a Numerical Weather Prediction (NWP) model or a General Circulation Model (GCM), or can be produced from the grand mean of the members of an Ensemble Prediction System (EPS), or can be based on an empirical model.

The forecasts are either objective numerical values such as departure from normal of a given parameter or expected occurrences (or non-occurrences) of events classified into categories (above/below normal or above/near/below normal for example). Although equi-probable categories are preferred for consistency, other classifications can be used in a similar fashion.

2.3. Probabilistic Long-Range Forecasts

Probabilistic LRF provide probabilities of occurrences or non-occurrences of an event or a set of fully inclusive events. Probabilistic LRF can be generated from an empirical model, or produced from an Ensemble Prediction System (EPS).

The events can be classified into categories (above/below normal or above/near/below normal for example). Although equi-probable categories are preferred for consistency, other classifications can be used in a similar fashion.

2.4. Terminology

There is no universally accepted definition of forecast period and forecast lead time. However, the definition in Table 2 will be used here.

Table 2: Definitions of forecast period and lead time.

Forecast period:	Forecast period is the validity period of a forecast. For example, long-range forecasts may be valid for a 90-day period or a season.
Lead time:	Lead time refers to the period of time between the issue time of the forecast and the beginning of the forecast validity period. Long-range forecasts based on all data up to the beginning of the forecast validity period are said to be of lead zero. The period of time between the issue time and the beginning of the validity period will categorise the lead. For example, a winter seasonal forecast issued at the end of the preceding summer season is said to be of one season lead. A seasonal forecast issued one month before the beginning of the validity period is said to be of one month lead.

Figure 1 presents the definitions of Table 2 in graphical format.

Forecast range determines how far into the future LRF are provided. Forecast range is thus the summation of lead time and forecast period.

Persistence, for a given parameter, stands for persisting the anomaly, which has been observed over the period of time with the same length as the forecast period and immediately prior to the LRF issue time (see Figure 1). It is important to realise that only the anomaly of any given parameter can be persisted. The persisted anomaly is added to the background climatology to retrieve the persisted parameter. Climatology is equivalent to persisting a uniform anomaly of zero.

3. SVS for Long-Range Forecasts

Forecast can be made using different levels of post-processing typically no-post-processing (raw or uncalibrated), simple correction of systematic errors (calibrated, i.e. calibration of mean and of variance) and more complex correction using hindcast skill (recalibrated, e.g. Model Output Statistics or perfect prog approaches). Most centres are currently issuing forecasts resulting from a simple calibration and so for sake of comparison on the Lead Centre web site scores for forecasts that were raw or calibrated (as specified in respective skill score section) are to be submitted. At the moment the team prefer to exclude forecast that were recalibrated, but GPCs are encouraged to apply the SVSLRF methodology and to display the results on their recalibrated forecasts on their web site.

3.1 Parameters to be verified

The following parameters are to be verified:

a) Surface air temperature (T2m) anomaly at screen level
b) Precipitation anomaly
c) Sea surface temperature (SST) anomaly.

In addition to these three parameters, the Niño3.4 Index, defined as the mean SST anomaly over the Niño-3.4 region from 170°W to 120°W and from 5°S to 5°N all inclusive is also to be verified.

It is recommended that three levels of verification be done:

a) level 1: large scale aggregated overall measures of forecast performance (see section 3.1.1).
b) level 2: verification at grid points (see section 3.1.2).
c) level 3: grid point by grid point contingency tables for more extensive verification (see section 3.1.3).

Both deterministic and probabilistic forecasts are verified if available. Level 1 is applicable to T2m anomaly, Precipitation anomaly and Niño3.4 Index. Levels 2 and 3 are applicable to T2m anomaly, Precipitation anomaly and SST anomaly.

3.1.1 Aggregated verification (level 1)

Large scale verification statistics are required in order to evaluate the overall skill of the models and ultimately for assessing their improvements. These are bulk numbers calculated by aggregating verification over all grid points within large regions; they will not necessarily reflect skill for any sub-region. This aggregated verification is performed over three regions:

a) Tropics: from 20°S to 20°N all inclusive.
b) Northern Extra-Tropics: from 20°N to 90°N, all inclusive.
c) Southern Extra-Tropics: from 20°S to 90°S, all inclusive.

The verification of Niño3.4 Index is also part of level 1 verification.

3.1.2 Grid point verification (level 2)

The grid point verification is recommended for a regionalised assessment of the skill of the model. The verification latitude/longitude grid is recommended as being 2.5° by 2.5°, with origin at 0°N, 0°E. Verification should be supplied to the Lead Centre for visual rendering. The formats for supplying derived verification are specified on the Lead Centre website.

3.1.3 Contingency tables (level 3)

Contingency tables allow users to perform more detailed verifications and generate statistics that are relevant for localised regions. The content and structure of the contingency tables is defined in sections 3.3.2 and 3.3.3. Data formats for supplying the contingency tables are specified on the Lead Centre website

3.1.4 Summary of the Core SVS

The following gives a summary of parameters, validation regions and diagnostics that form the core SVS. The required periods, lead-times and stratification against the state of ENSO are given in section 3.2.

Level 1			
Parameters	**Verification regions**	**Deterministic forecasts**	**Probabilistic forecasts**
T2m anomaly Precipitation anomaly	Tropics Northern Extra-Tropics Southern Extra-Tropics (section 3.1.1)	MSSS (bulk number) (section 3.3.1)	ROC curves ROC areas Reliability diagrams Frequency histograms (sections 3.3.3 and 3.3.4)
Niño3.4 Index	N/A	MSSS (bulk number) (section 3.3.1)	ROC curves ROC areas Reliability diagrams Frequency histograms (sections 3.3.3 and 3.3.4)
Level 2			
Parameters	**Verification regions**	**Deterministic forecasts**	**Probabilistic forecasts**
T2m anomaly Precipitation anomaly SST anomaly	grid point verification on a 2.5° by 2.5° grid (section 3.1.2)	MSSS and its three term decomposition at each grid point (section 3.3.1)	ROC areas at each grid point (section 3.3.3)
Level 3			
Parameters	**Verification regions**	**Deterministic forecasts**	**Probabilistic forecasts**
T2m anomaly Precipitation anomaly SST anomaly	grid point verification on a 2.5° by 2.5° grid (section 3.1.2)	3 by 3 contingency tables at each grid point (section 3.3.2)	ROC/reliability tables at each grid point (section 3.3.3)

The number of realisations of LRF is far smaller than in the case of short term numerical weather prediction forecasts. Consequently it is essential as part of the core SVS, to calculate and report error bars and level of significance (see section 3.3.5).

In order to ease implementation, participating LRF producers may stage the introduction of the core SVS by prioritizing implementation of verification at levels 1 and 2.

Other parameters and indices to be verified as well as other verification scores can be added to the core SVS in future versions.

3.2 Verification strategy

LRF verification should be done on a global latitude/longitude grid with areas as defined in section 3.1.1. Verification can also be done at individual stations or groups of stations. Verification on a latitude/longitude grid is performed separately from the one done at stations.

The verification latitude/longitude grid is recommended as being 2.5° by 2.5°, with origin at 0°N, 0°E. Both forecasts and the gridded verifying data sets are to be interpolated onto the same 2.5° by 2.5° grid.

In order to handle spatial forecasts, predictions for each point within the verification grid should be treated as individual forecasts but with all results combined into the final outcome. The same approach is applied when verification is done at stations. Categorical forecast verification can be performed for each category separately.

Similarly, all forecasts are treated as independent and combined together into the final outcome, when verification is done over a long period of time (several years for example).

Stratification of the verification data is based on forecast period, lead time and verification area. Stratification by forecast period should, for T2m and precipitation, be by 4 conventional seasons for Level 1. For Levels 2&3 stratification should be on 12 rolling seasons (section 2.1) if available, otherwise 4 conventional seasons should be used. Verification results for different seasons should not be mixed. Stratification by lead-time should include a minimum of two leadtimes, with lead-time not greater than 4 months. Forecasts with different lead times are similarly to be verified separately. Stratification according to the state of ENSO (where there are sufficient cases) should be as follows:

a) all hindcast seasons

b) seasons with El Niño active

c) seasons with La Niña active

For Niño3.4 SST anomaly verification should be stratified according to each calendar month and lead-time. Six lead-times should be provided, ranging from zero to 5-month lead.

3.3 Verification scores

The following verification scores are to be used:

- Mean Square Skill Score (MSSS)
- Relative Operating Characteristics (ROC).

MSSS is applicable to deterministic forecasts only, while ROC is applicable to both deterministic and probabilistic forecasts. MSSS is applicable to non-categorical forecasts (forecasts of continuous variables), while ROC is applicable to categorical forecasts either deterministic or probabilistic in nature.

Verification methodology using ROC, is derived from signal detection theory. This methodology is intended to provide information on the characteristics of systems upon which management decisions can be taken. In the case of weather/climate forecasts, the decision might relate to the most appropriate manner in which to use a forecast system for a given purpose. ROC is applicable to both deterministic and probabilistic categorical forecasts and is useful in contrasting characteristics of deterministic and probabilistic systems. The derivation of ROC is based on contingency tables giving the hit rate and false alarm rate for deterministic or probabilistic forecasts. The events are defined as binary, which means that only two outcomes are possible, an occurrence or a non-occurrence. It is recognised that ROC as applied to deterministic forecasts is equivalent to the Hanssen and Kuipers score (see section 3.3.2).

The binary event can be defined as the occurrence of one of two possible categories when the outcome of the LRF system is in two categories. When the outcome of the LRF system is in three (or more) categories, the binary event is defined in terms of occurrences of one category against the remaining ones. In those circumstances, ROC has to be calculated for each possible category.

3.3.1 MSSS for non-categorical deterministic forecasts

Let x_{ij} and f_{ij} (i=1,…,n) denote time series of observations and continuous deterministic forecasts respectively for a grid point or station j over the period of verification (POV). Then, their averages for the POV, $\bar{x}_j$ and $\bar{f}_j$ and their sample variances s_{xj}^2 and s_{fj}^2 are given by:

$$\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}, \; \bar{f}_j = \frac{1}{n}\sum_{i=1}^{n} f_{ij}$$

$$s_{xj}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_{ij} - \bar{x}_j\right)^2, \; s_{fj}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(f_{ij} - \bar{f}_j\right)^2$$

The mean squared error of the forecasts is:

$$MSE_j = \frac{1}{n}\sum_{i=1}^{n}\left(f_{ij} - x_{ij}\right)^2$$

For the case of cross-validated (see section 3.4) POV climatology forecasts where forecast/observation pairs are reasonably temporally independent of each other (so that only one year at a time is withheld), the mean squared error of 'climatology' forecasts (Murphy, 1988) is:

$$MSE_{cj} = \left(\frac{n}{n-1}\right)^2 s_{xj}^2$$

The *Mean Squared Skill Score* (MSSS) for j is defined as one minus the ratio of the squared error of the forecasts to the squared error for forecasts of 'climatology':

$$MSSS_j = 1 - \frac{MSE_j}{MSE_{cj}}$$

For the three domains described in Sec. 3.1.1 it is recommended that an overall MSSS be provided. This is computed as:

$$MSSS = 1 - \frac{\sum_j w_j MSE_j}{\sum_j w_j MSE_{cj}}$$

where w_j is unity for verifications at stations and is equal to $\cos(\theta_j)$, where θ_j is the latitude at grid point j on latitude-longitude grids.

For either $MSSS_j$ or MSSS a corresponding *Root Mean Squared Skill Score* (RMSSS) can be obtained easily from

$$RMSSS = 1 - (1 - MSSS)^{1/2}$$

MSSSj for forecasts fully cross-validated (with one year at a time withheld) can be expanded (Murphy, 1988) as

$$MSSS_j = \left\{ 2\frac{s_{fj}}{s_{xj}} r_{fxj} - \left(\frac{s_{fj}}{s_{xj}}\right)^2 - \left(\frac{\left[\bar{f}_j - \bar{x}_j\right]}{s_{xj}}\right)^2 + \frac{2n-1}{(n-1)^2} \right\} \Bigg/ \left\{ 1 + \frac{2n-1}{(n-1)^2} \right\}$$

where r_{fxj} is the product moment correlation of the forecasts and observations at point or station j.

$$r_{fxj} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(f_{ij} - \bar{f}_j\right)\left(x_{ij} - \bar{x}_j\right)}{s_{fj}\, s_{xj}}$$

The first three terms of the decomposition of $MSSS_j$ are related to phase errors (through the correlation), amplitude errors (through the ratio of the forecast to observed variances) and overall bias error, respectively, of the forecasts. These terms provide the opportunity for those wishing to use the forecasts for input into regional and local forecasts to adjust or weight the forecasts as they deem appropriate. The last term takes into account the fact that the 'climatology' forecasts are cross-validated as well.

Note that for forecasts with the same amplitude as that of observations (second term unity) and no overall bias (third term zero), $MSSS_j$ will not exceed zero (i.e. the forecasts squared error will not be less than for 'climatology') unless r_{fxj} exceeds approximately 0.5.

The core SVSLRF requires grid-point values of the correlation, the ratio of the square roots of the variances, and the overall bias i.e.

$$r_{fxj}, \frac{s_{fj}}{s_{xj}}, \left[\bar{f}_j - \bar{x}_j\right].$$

In addition it is recommended that grid-point (j) values of the following quantities are provided:

$$n, \bar{f}_j, \bar{x}_j, s_{fj}, s_{xj}, r_{fxj}, MSE_j, MSE_{cj}, MSSS_j$$

As an additional standard against which to measure forecast performance, cross-validated *damped persistence* (defined below) should be considered for certain forecast sets. A forecast of *ordinary persistence*, for a given parameter and target period, stands for the persisted anomaly (departure from cross-validated climatology) from a period immediately preceding the start of the lead time for the forecast period (see Figure 1). This period must have the same length as the forecast period. For example, the ordinary persistence forecast for a 90-day period made 15 days in advance would be the anomaly of the 90-day period beginning 105 days before the target forecast period and ending 16 days before. Ordinary persistence forecasts are never recommended as a standard against which to measure other forecasts if the performance or skill measures are based on squared error, like herein. This is because persistence is easy to beat in this framework.

Damped persistence is the optimal persistence forecast in a least squared error sense. Even *damped persistence should not be used in the case of extratropical seasonal forecasts*, because the nature of the interannual variability of seasonal means changes considerably from one season to the next in the extratropics. For all other cases damped persistence forecasts can be made in a cross-validated mode (Section 3.4) and the skill and performance diagnostics based on the squared error described above (bulk measures, grid-point values, and tables) can be computed and presented for these forecasts.

Damped persistence is the ordinary persistence anomaly $x_{ij}(t-\Delta t)-\overline{x}_{ij}^{m}(t-\Delta t)$ damped (multiplied) towards climatology by the cross-validated, lagged product moment correlation between the period being persisted and the target forecast period.

Damped persistence forecast: $r_{\Delta,j}^{m}\left[x_{ij}(t-\Delta t)-\overline{x}_{ij}^{m}(t-\Delta t)\right]$

$$r_{\Delta,j}^{m}=\frac{\frac{1}{m}\sum_{m}\left[x_{ij}(t-\Delta t)-\overline{x}_{ij}^{m}(t-\Delta t)\right]\left[x_{ij}(t)-\overline{x}_{ij}^{m}(t)\right]}{s_{xj}^{m}(t-\Delta t)\,s_{xj}^{m}(t)}$$

where t is the target forecast period, t-Δt the persisted period (preceding the lead time), and m denotes summation (for $r_{\Delta,j}^{m}$, $\overline{x}_{ij}^{m}$, s_{xj}^{m}) at each stage of the cross-validation over all i except those being currently withheld (Section 3.4).

⇒ MSSS, provided as a single bulk number, is mandatory for level 1 verification in the core SVS. MSSS together with its three term decomposition are also mandatory for level 2 verification in the core SVS. For the exchange of scores via the Lead Centre web site the MSSS and its decomposition term should be calculated using the raw forecasts and preferably not the calibrated ones.

3.3.2 Contingency tables and scores for categorical deterministic forecasts

For two- or three-category deterministic forecasts the core SVSLRF includes full contingency tables, because it is recognized that they constitute the most informative way to evaluate the performance of the forecasts. These contingency tables then form the basis for several skill scores that are useful for comparisons between different deterministic categorical forecast sets (Gerrity, 1992) and between deterministic and probabilistic categorical forecast sets (Hanssen and Kuipers, 1965) respectively.

The contingency tables should be provided for every combination of parameter, lead time, target month or season, and ENSO stratification (when appropriate) at every verification point for both the forecasts and (when appropriate) damped persistence. The definition of ENSO events is provided on the Lead Centre web site. If x_i and f_i now denote an observation and corresponding forecast of category i (i = 1,…,3), let n_{ij} be the count of those instances with forecast category i and observed category j. The full contingency table is defined as the nine n_{ij}. Graphically the nine cell counts are usually arranged with the forecasts defining the table rows and the observations the table columns:

Table 3: General three by three contingency table.

		Observations			
		Below Normal	Near Normal	Above Normal	
Forecasts	Below Normal	n_{11}	n_{12}	n_{13}	$n_{1\bullet}$
	Near Normal	n_{21}	n_{22}	n_{23}	$n_{2\bullet}$
	Above Normal	n_{31}	n_{32}	n_{33}	$n_{3\bullet}$
		$n_{\bullet 1}$	$n_{\bullet 2}$	$n_{\bullet 3}$	T

In Table 3, $n_{i\bullet}$ and $n_{\bullet i}$ represents the sum of the rows and columns respectively; T is the total number of cases. Generally about at least 90 forecast/observation pairs are required to properly estimate a three by three contingency table. Thus it is recommended that the provided tables be aggregated by users over windows of target periods, like several adjacent months or overlapping three-month periods, or over verification points. In the case of the latter the weights W_i should be used in summing n_{ij} over different points i (see discussion on Table 4). W_i is defined as:

$W_i = 1$ when verification is done at stations or at single grid points within a limited geographical region.

$W_i = \cos(\theta_i)$ at grid point i, when verification is done on a grid.

$\theta_i =$ the latitude at grid point i.

On a 2.5 degree latitude-longitude grid the minimally acceptable sample is easily attained even with a record as short as n = 10 by aggregating over all grid points with a 10 degree box. Or alternatively in this case, an adequate sample can be achieved by aggregation over three adjacent months or overlapping three-month periods and within a 5 degree box. Regardless, scores derived from any contingency table should be accompanied by error bars, confidence intervals or level of significance.

Contingency tables such as the one in Table 3 are mandatory for level 3 verification in the core SVS.

The *relative sample frequencies* p_{ij} are defined as the ratios of the cell counts to the total number of forecast/observation pairs N (n is reserved to denote the length of the POV):

$$p_{ij} = n_{ij} / N$$

The sample probability distributions of forecasts and observations respectively then become

$$p(f_i) = \sum_{j=1}^{3} p_{ij} = \hat{p}_i \, ; i = 1, \ldots, 3$$

$$p(x_i) = \sum_{j=1}^{3} p_{ji} = p_i \, ; i = 1, \ldots, 3$$

A recommended skill score for the three by three table, which has many desirable properties and is easy to compute is the *Gerrity Skill Score*, GSS. The definition of the score uses a scoring matrix s_{ij} (i = 1,…,3), which is a tabulation of the reward or penalty every forecast/observation outcome (represented by the contingency table) will be accorded:

$$GSS = \sum_{i=1}^{3}\sum_{j=1}^{3} p_{ij}\, s_{ij}$$

The scoring matrix is given by

$$s_{ii} = \frac{1}{2}\left(\sum_{r=1}^{i-1} a_r^{-1} + \sum_{r=i}^{2} a_r\right)$$

$$s_{ij} = \frac{1}{2}\left[\sum_{r=1}^{i-1} a_r^{-1} - (j-i) + \sum_{r=j}^{2} a_r\right]; 1 \le i < 3, i < j \le 3$$

where

$$a_i = \frac{1 - \sum_{r=1}^{i} p_r}{\sum_{r=1}^{i} p_r}$$

Note that GSS is computed using the sample probabilities, not those on which the original categorisations were based (i.e. 0.33, 0.33, 0.33).

The GSS can be alternatively computed by the numerical average of two of the three possible two-category, unscaled Hanssen and Kuipers scores (introduced below) that can be computed from the three by three table. The two are computed from the two two-category contingency tables formed by combining categories on either side of the partitions between consecutive categories: (1) above normal and a combined near and below normal category and (2) below normal and a combined near and above normal category.

The GSS's ease of construction ensures its consistency from categorization to categorization and with underlying linear correlations. The score is likewise equitable, does not depend on the forecast distribution, does not reward conservatism, utilizes off diagonal information in the contingency table, and penalizes larger errors more. For a limited subset of forecast situations it can be manipulated by a forecaster to his/her advantage (Mason and Mimmack, 2002), but this is not a problem for objective forecast models that have not been trained to take advantage of this weakness. For all these reasons it is the recommended score.

An alternative score to the GSS for consideration is LEPSCAT (Potts et al., 1996)

Table 4 shows the general form for the three possible two by two contingency tables referred to above (the third is the table for the near normal category and the combined above and below normal category). In Table 4, T is the grand sum of all the proper weights applied on each occurrence and non-occurrence of the events.

Table 4: General ROC contingency table for deterministic forecasts.

		Observations		
		occurrences	**non-occurrences**	
forecasts	**occurrences**	O_1	NO_1	$O_1 + NO_1$
	non-occurrences	O_2	NO_2	$O_2 + NO_2$
		$O_1 + O_2$	$NO_1 + NO_2$	T

The 2X2 table in Table 4 may be constructed from the 3X3 table described in Table 3 by summing the appropriate rows and columns.

In Table 4, O_1 represents the correct forecasts or hits:

$$O_1 = \sum W_i (OF)_i$$

(OF) being 1 when the event occurrence is observed and forecast; 0 otherwise. The summation is over all grid points or stations.

NO_1 represents the false alarms:

$$NO_1 = \sum W_i (NOF)_i$$

(NOF) being 1 when the event occurrence is not observed but was forecast; 0 otherwise. The summation is over all grid points or stations.

O_2 represents the misses:

$$O_2 = \sum W_i (ONF)_i$$

(ONF) being 1 when the event occurrence is observed but not forecast; 0 otherwise. The summation is over all grid points or stations.

NO_2 represents the correct rejections:

$$NO_2 = \sum W_i (NONF)_i$$

(NONF) being 1 when the event occurrence is not observed and not forecast; 0 otherwise. The summation is over all grid points or stations.

$W_i = 1$ when verification is done at stations or at single grid points.

$W_i = \cos(\theta_i)$ at grid point i, when verification is done on a grid.

$\theta_i =$ the latitude at grid point i.

When verification is done at stations, the weighting factor is one. Consequently, the number of occurrences and non-occurrences of the event are entered in the contingency table of Table 4.

However, when verification is done on a grid, the weighting factor is $\cos(\theta_i)$, where θ_i is the latitude at grid point i. Consequently, each number entered in the contingency table of Table 5, is, in fact, a summation of the weights properly assigned.

Using stratification by observations (rather than by forecast), the Hit Rate (HR) is defined as (referring to Table 4):

$$HR = O_1 / (O_1 + O_2)$$

The range of values for HR goes from 0 to 1, the latter value being desirable. An HR of one means that all occurrences of the event were correctly forecast.

The False Alarm Rate (FAR) is defined as:

$$FAR = NO_1 / (NO_1 + NO_2)$$

The range of values for FAR goes from 0 to 1, the former value being desirable. A FAR of zero means that in the verification sample, no non-occurrences of the event were forecast to occur.

Hanssen and Kuipers score (see Hanssen and Kuipers, 1965 and Stanski et al, 1989) is calculated for deterministic forecasts. Hanssen and Kuipers score (KS) is defined as:

$$\begin{aligned} KS &= HR - FAR \\ &= \frac{O_1 NO_2 - O_2 NO_1}{(O_1 + O_2)(NO_1 + NO_2)} \end{aligned}$$

The range of KS goes from -1 to +1, the latter value corresponding to perfect forecasts (HR being 1 and FAR being 0). KS can be scaled so that the range of possible values goes from 0 to 1 (1 being for perfect forecasts):

$$KS_{scaled} = \frac{KS + 1}{2}$$

The advantage of scaling KS is that it becomes comparable to the area under the ROC curve for probabilistic forecasts (see section 3.33) where a perfect forecast system has an area of one and a forecast system with no information has an area of 0.5 (HR being equal to FAR).

⇒ Contingency tables for deterministic categorical forecasts (such as in Table 3) are mandatory for level 3 verification in the core SVS. These contingency tables can provide the basis for the calculation of several scores and indices such as the Gerrity Skill Score, the LEPSCAT or the scaled Hanssen and Kuipers score and others.

3.3.3 ROC for probabilistic forecasts

Tables 5 and 6 show contingency tables (similar to Table 4) that can be built for probabilistic forecasts of binary events.

Table 5: General ROC contingency table for probabilistic forecasts of binary events with definitions of the different parameters. This contingency table applies when probability thresholds are used to define the different probability bins.

bin number	**forecast probabilities**	**observed occurrences**	**observed non-occurrences**
1	0-P_2 (%)	O_1	NO_1
2	P_2-P_3 (%)	O_2	NO_2
3	P_3-P_4 (%)	O_3	NO_3
•••	•••	•••	•••
n	P_n-P_{n+1} (%)	O_n	NO_n
•••	•••	•••	•••
N	P_N-100 (%)	O_N	NO_N

In Table 5,

n = number of the n^{th} probability interval or bin n; n goes from 1 to N.
P_n = lower probability limit for bin n.
P_{n+1} = upper probability limit for bin n.
N = number of probability intervals or bins.

$$O_n = \sum W_i (O)_i$$

(O) being 1 when an event corresponding to a forecast in bin n, is observed as an occurrence; 0 otherwise. The summation is over all forecasts in bin n, at all grid points or stations.

$$NO_n = \sum W_i (NO)_i$$

(NO) being 1 when an event corresponding to a forecast in bin n, is not observed; 0 otherwise. The summation is over all forecasts in bin n, at all grid points i or stations i

$W_i = 1$ when verification is done at stations or at single grid points.

$W_i = \cos(\theta_i)$ at grid point i, when verification is done on a grid.

$\theta_i =$ the latitude at grid point i.

Table 6: General ROC contingency table for probabilistic forecasts of binary events with definitions of the different parameters. This contingency table applies when the different probability bins are defined as function of the number of members in the ensemble.

bin number	member distribution	observed occurrences	observed non-occurrences
1	F=0, NF=M	O_1	NO_1
2	F=1, NF=M-1	O_2	NO_2
3	F=2, NF=M-2	O_3	NO_3
●●●		●●●	●●●
n	F=n-1, NF=M-n+1	O_n	NO_n
●●●		●●●	●●●
N	F=M, NF=0	O_N	NO_N

In Table 6,

M = number of members in the ensemble

n = number of the n^{th} bin; n goes from 1 toN=M+1.
F = the number of members forecasting occurrence of the event.
NF = the number of members forecasting non occurrence of the event.

The bins may be aggregated.

$$O_n = \sum W_i (O)_i$$

(O) being 1 when an event corresponding to a forecast in bin n, is observed as an occurrence; 0 otherwise. The summation is over all forecasts in bin n, at all grid points i or stations i.

$$NO_n = \sum W_i (NO)_i$$

(NO) being 1 when an event corresponding to a forecast in bin n, is not observed; 0 otherwise. The summation is over all forecasts in bin n, at all grid points i or stations i.

$W_i = 1$ when verification is done at stations or at single grid points.

$W_i = \cos(\theta_i)$ at grid point i, when verification is done on a grid.

$\theta_i =$ the latitude at grid point i.

To build the contingency table in Table 5, probability forecasts of the binary event are grouped in categories or bins in ascending order, from 1 to N, with probabilities in bin n-1 lower than those in bin n (n goes from 1 to N). The lower probability limit for bin n is P_n and the upper limit is P_{n+1}. The lower probability limit for bin 1 is 0%, while the upper limit in bin N is 100%. The summation of the weights on the observed occurrences and non-occurrences of the event corresponding to each forecast in a given probability interval (bin n for example) is entered in the contingency table.

Tables 5 and 6 outline typical contingency tables. It is recommended that the number of probability bins remain between 10 and 20. The forecast providers can bin according to percent thresholds

(Table 5) or ensemble members (Table 6) as deemed necessary. Table 6 gives an example of a table based on ensemble members.

Hit rate and false alarm rate are calculated for each probability threshold P_n (see Tables 5 and 6). The hit rate for probability threshold P_n (HR_n) is defined as (referring to Tables 5 and 6):

$$HR_n = \sum_{i=n}^{N} O_i \Big/ \sum_{i=1}^{N} O_i$$

and the false alarm rate (FAR_n) is defined as:

$$FAR_n = \sum_{i=n}^{N} NO_i \Big/ \sum_{i=1}^{N} NO_i$$

where n goes from 1 to N. The range of values for HR_n goes from 0 to 1, the latter value being desirable. The range of values for FAR_n goes from 0 to 1, zero being desirable. Frequent practice is for probability intervals of 10% (10 bins, or N=10) to be used. However the number of bins (N) should be consistent with the number of members in the ensemble prediction system (EPS) used to calculate the forecast probabilities. For example, intervals of 33% for a nine-member ensemble system could be more appropriate.

Hit rate (HR) and false alarm rate (FAR) are calculated for each probability threshold P_n, giving N points on a graph of HR (vertical axis) against FAR (horizontal axis) to form the Relative Operating Characteristics (ROC) curve. This curve, by definition, must pass through the points (0,0) and (1,1) (for events being predicted only with $>$100% probabilities (never occurs) and for all probabilities exceeding 0% respectively). No-skill forecasts are indicated by a diagonal line (where HR=FAR); the further the curve lies towards the upper left-hand corner (where HR=1 and FAR=0) the better

The area under the ROC curve is a commonly used summary statistics representing the skill of the forecast system. The area is standardised against the total area of the figure such that a perfect forecast system has an area of one and a curve lying along the diagonal (no information) has an area of 0.5. The normalised ROC area has become known as the ROC score. Not only can the areas be used to contrast different curves, but they are also a basis for Monte Carlo significance tests. It is proposed that Monte Carlo testing should be done within the forecast data set itself. For the core SVSLRF the area under the ROC curve should be calculated using the Trapezium rule (Other techniques are available to calculate the ROC score (see Mason, 1982).)

⇒ Contingency tables for probabilistic forecasts (such as in Tables 5 and 6) are mandatory for level 3 verification in the core SVS. ROC curves and ROC areas are mandatory for level 1 verification in the core SVS while ROC areas only are mandatory for level 2 verification in the core SVS.

3.3.4 Reliability diagrams and frequency histograms for probabilistic forecasts

It is recommended that the construction of reliability curves (including frequency histograms to provide indications of sharpness) be done for the large-sampled probability forecasts aggregated over the tropics and, separately, the two extratropical hemispheres. Given frequency histograms, the reliability curves are sufficient for the ROC curve, and have the advantage of indicating the reliability of the forecasts, which is a deficiency of the ROC. It is acknowledged that the ROC curve is frequently the more appropriate measure of forecast quality than the reliability diagram in the context of verification of long-range forecasts because of the sensitivity of the reliability diagram to small sample sizes. However, because measures of forecast reliability are important for modellers, forecasters, and end-users, it is recommended that in the exceptional cases of the forecasts being spatially aggregated over the tropics and over the two extratropical hemispheres, reliability diagrams be constructed in addition to ROC curves.

The technique for constructing the reliability diagram is somewhat similar to that for the ROC. Instead of plotting the hit rate against the false alarm rate for the accumulated probability bins, the hit rate is calculated only from the sets of forecasts for each probability bin separately, and is plotted against the corresponding forecast probabilities. The hit rate for each probability bin (HR_n) is defined as:

$$HR_n = \frac{O_n}{O_n + NO_n}$$

This equation should be contrasted with the hit rate used in constructing the ROC diagram.

Frequency histograms are constructed similarly from the same contingency tables as those used to produce reliability diagrams. Frequency histograms show the frequency of forecasts as a function of the probability bin. The frequency of forecasts (F_n) for probability bin n is defined as:

$$F_n = \frac{O_n + NO_n}{T}$$

where T is the total number of forecasts (and $T = \sum_{n=1}^{N} (O_n + NO_n)$).

⇒ Reliability diagrams and frequency histograms are mandatory for level 1 verification in the core SVS.

3.3.5 Level of significance

Because of the increasing uncertainty in verification statistics with decreasing sample size, significance levels and error bars should be calculated for all verification statistics. Recommended procedures for estimating these uncertainties are detailed below.

ROC area

In certain special cases the statistical significance of the ROC area can be obtained from its relationship to the Mann–Whitney U-statistic. The distribution properties of the U-statistic can be used only if the samples are independent. This assumption of independence will be invalid when the ROC is constructed from forecasts sampled in space because of the strong spatial (cross) correlation between forecasts (and observations) at nearby grid-points or stations. However, because of the weakness of serial correlation of seasonal climate anomalies from one year to the next, an assumption of sequential independence may frequently be valid for long-range forecasts, and so Mann–Whitney U-statistic may be used for calculating the significance of the ROC area for a set of forecasts from a single point in space. An additional assumption for using the Mann–Whitney U-test is that the variance of the forecast probabilities (not that of the individual ensemble predictions per se) for when non-events occurred is the same as those for when events occurred. The Mann–Whitney U-test is, however, reasonably robust to violations of homoscedasticity which means that the variance of the error term is constant across the range of the variable, and so significance tests in cases of unequal variance are likely to be only slightly conservative.

If the assumptions for the Mann–Whitney U-test cannot be held, the significance of the ROC area should be calculated using randomisation procedures. Because the assumptions of permutation procedures are the same as those of the Mann–Whitney U-test, and because standard bootstrap procedures assume independence of samples, alternative procedures such as moving block bootstrap procedures (Wilks, 1997) should be conducted to ensure that the cross- and/or serial-correlation structure of the data is retained.

ROC curves

Confidence bands for the ROC curve should be indicated, and can be obtained either by appropriate bootstrap procedures, as discussed above, or, if the assumption of independent forecasts is valid, from confidence bands derived from a two-sample Kolmogorov-Smirnov test comparing the empirical ROC with the diagonal.

MSSS

Appropriate significance tests for the MSSS and the individual components of the decomposition again depend upon the validity of the assumption of independent forecasts. If the assumption is valid, significance tests could be conducted using standard procedures (namely the F-ratio for the correlation and for the variance ratio, and the t-test for the difference in means), otherwise bootstrap procedures are recommended.

⇒ Level of significance will be mandatory in the core SVS once guidelines for calculation have been established for the complete suite of scores. A phased in introduction of level of significance in the SVS may be used (see section 3.1.4).

3.4 Hindcasts

In contrast to short- and medium-range dynamical Numerical Weather Prediction (NWP) forecasts, LRF are produced relatively few times a year (for example, one forecast for each season or one forecast for the following 90-day period, issued every month). Therefore the verification sampling for LRF may be limited, possibly to the point where the validity and significance of the verification results may be questionable. Providing verification for a few seasons or even over a few years only may be misleading and may not give a fair assessment of the skill of any LRF system. LRF systems should be verified over as long a period as possible in hindcast mode. Although there are limitations on the availability of verification data sets and in spite of the fact that validating numerical forecast systems in hindcast mode requires large computer resources, the hindcast period should be as long as possible. The recommended period for the exchange of scores is advertised on the Lead Centre web site (http://www.bom.gov.au/wmo/lrfvs/).

Verification in hindcast mode should be achieved in a form as close as possible to the real time operating mode in terms of resolution, ensemble size and parameters. In particular dynamical/empirical models must not make any use of future data. Validation of empirical models, dynamical models with postprocessors (including bias corrections), and calculation of period of verification means, standard deviations, class limits, etc. must be done in a cross-validation framework. Cross-validation allows the entire sample to be used for validation (assessing performance, developing confidence intervals, etc.) and almost the entire sample for model and post-processor building and for estimation of period of verification climatology. Cross-validation proceeds as follows:

1. Delete 1, 3, 5, or more years from the complete sample;
2. Build the statistical model or compute the climatology;
3. Apply the model (e.g. make statistical forecasts or postprocess the dynamical forecasts) or the climatology for one (usually the middle) year of those deleted and verify;
4. Replace the deleted years and repeat 1-3 for a different group of years;
5. Repeat 4 until the hindcast verification sample is exhausted.

Ground rules for cross–validation are that every detail of the statistical calculations be repeated, including redefinition of climatology and anomalies, and that the forecast year predictors and predictands are not serially correlated with their counterparts in the years reserved for model building. For example, if adjacent years are correlated but every other year is effectively not, three years must be set aside and forecasts made only on the middle year (see Livezey, 1999, for estimation of the reserved window width).

The hindcast verification statistics should be updated once a year based on accumulated forecasts.

⇒ Verification results over the hindcast period are mandatory for the exchange of LRF verification scores. Producing centres have to send new hindcast verification results as soon as their forecast system is changed.

3.5 Real-time monitoring of forecasts

It is recommended that there be regular monitoring of the real time long range forecasts. It is acknowledged that this real-time monitoring is neither as rigorous nor as sophisticated as the hindcast verification; nevertheless it is necessary for forecast production and dissemination. It is also acknowledged that the sample size for this real-time monitoring may be too small to assess the overall skill of the models. However, it is recommended that the forecast and the observed verification for the previous forecast period be presented in visual format to the extent possible given the restrictions on availability of verification data.

Real-time monitoring of forecast performance is an activity for the GPCs rather than the Lead Centre. GPCs are free to choose the format and content of real-time monitoring information.

4. VERIFICATION DATA SETS

The same data should be used to generate both climatology and verification data sets, although the forecast issuing Centres/Institutes own analyses or reanalyses and subsequent operational analyses may be used when other data are not available.

Many LRF are produced that are applicable to limited or local areas. It may not be possible to use the data in either the recommended climatology or verification data sets for validation or verification purposes in these cases. Appropriate data sets should then be used with full details provided.

1. Verification should be done using the recommended data sets as listed on the Lead Centre web site (http://www.bom.gov.au/wmo/lrfvs/).

5. SYSTEM DETAILS

Information must be provided on the system being verified. This information should include (but is not restricted to):

1. Whether the system numerical, empirical or hybrid.
2. Whether the system is deterministic or probabilistic
3. Model type and resolution.
4. Ensemble size.
5. Boundary conditions specifications.
6. List of parameters being assessed.
7. List of regions for each parameter.
8. List of forecast ranges (lead times) and periods for each parameter.
9. Period of verification.
10. The number of hindcasts or predictions incorporated in the assessment and the dates of these hindcasts or predictions.

11. Details of climatological and verification data sets used (with details on quality control when these are not published).
12. If appropriate, resolution of fields used for climatologies and verification.

Verification data for the aggregated statistics and the grid point data should be provided on the web. The contingency tables should be made available by the web or anonymous FTP. Real-time monitoring should be done as soon as possible and made available on the web.

6. LEAD CENTRES FOR SVSLRF

The WMO Fourteenth Congress endorsed the designation by CBS (Ext. 02) of WMC Melbourne and the Canadian Meteorological Centre Montreal as Co-Lead Centres for verification of long-range and SI forecast activities Congress. The co-lead centre functions include creating and maintaining coordinated Web sites for the LRF verification information, so that potential users would benefit from a consistent presentation of the results. The goal is to help the RCCs and NMHSs to have a tool for improving the long-range forecasts delivered to the public. Congress urged all Members to actively participate in that activity as either users or producers of LRF verification information to assure the use of the best available products.

6.1 Role of lead centre

6.1.1 Create, develop and maintain web-site (the "SVSLRF web site") to provide access to the LRF verification information. **The address of the web site is** http://www.bom.gov.au/wmo/lrfvs/. The web-site will:

(i) Provide access to standardized software for calculating scoring information (ROC curves, areas, contingency table scores, hit rates, …).

(ii) provide consistent graphical displays of the verification results from participating centres through processing of digital versions of the results;

(iii) contain relevant documentation and links to the web sites of global-scale producing centres;

(iv) provide some means for the collection of feedback from NMHSs and RCCs on the usefulness of the verification information;

(v) Contain information and, preferably, provide access to available verification data sets;

6.1.2 The centre will also:

(i) Produce monthly verification data sets in common format on 2.5° x 2.5° grid where appropriate;

(ii) liaise with other groups involved in verification (e.g. WGSIP, CCl, etc.) on the effectiveness of the current standardised verification system (SVS) and identify areas for future development and improvement;

(iii) provide periodic reports to CBS and other relevant Commissions assessing the effectiveness of the SVS.

(iv) facilitate the availability of information to assess the skill of long-range forecasts but not to provide a direct inter-comparison between the GPCs' models.

6.1.3 Detailed tasks of the "lead centre":

6.1.3.1 The Lead Centre will provide access to verification datasets on the SVSLRF web site. The verification datasets will be in GRIB1 format. They will be translated to GRIB2 format when the encoder/decoder becomes widely available. The RSMC Montreal will take the responsibility for preparing the verification datasets. These will be updated on the SVSLRF web site on a yearly basis provided that new data is made available. The choice of the verification datasets will be revised as new datasets become available and as recommended by the appropriate CBS expert team.

6.1.3.2 The Lead Centre will develop and provide specifications defining the format of the data to be sent to the Lead Centre for graphics preparation. There is no need to specify standards for graphics to be sent to the SVSLRF web site because all graphics will be generated by the Lead Centre. The WMC Melbourne will develop the infrastructure to generate all graphics posted on the SVSLRF web site.

6.1.3.3 The Lead Centre will have the responsibility to make available the digital verification information as specified at levels 1, 2 and 3 (see section 3.1).

6.1.3.4 The Lead Centre will ensure that clear and concise information explaining the verification scores, graphics and data is available and maintained up-to-date on the SVSLRF web site. The production of this documentation will be shared between the two co-lead centres. Also, links to the participating Global Producing Centres (GPCs) will be listed on the SVSLRF web site. The content of the documentation and information on interpretation and use of the verification data will be determined in consultation with the appropriate CBS expert team.

6.1.3.5 The Lead Centre will consult with the GPCs to make sure that the verification data is correctly displayed before making available their verification results on the SVSLRF web site.

6.1.3.6 The Lead Centre will ensure that the verification results placed on the SVSLRF web site come from global producing centres (officially recognised by CBS) with operational LRF commitments; 6.1.3.7 The Lead Centre will provide and maintain software to calculate the verification scores. The development of the software will be the responsibility of the RSMC Montreal. The software code will be available on the SVSLRF web site. It will be coded in FORTRAN language. However, it is recognised that the use of this software is not mandatory.

6.1.3.8 The Lead Centre will publicise the SVSLRF web site to other organisations involved in verification (such as WGSIP, CCl etc.) and establish contacts in order to receive feedback and facilitate discussion for further development and improvement.

6.1.3.9 Once the SVSLRF web site is operational, the Lead Centre will provide progress reports every two years to CBS, prior to its meetings.

7. REFERENCES

Gerrity, J. P. Jr., 1992: A note on Gandin and Murphy's equitable skill score. *Mon. Wea. Rev*., 120, 2707-2712.

Hanssen A. J. and W. J. Kuipers, 1965: On the relationship between the frequency of rain and various meteorological parameters. *Koninklijk Nederlands Meteorologist Institua Meded*. Verhand, 81-2-15.

Livezey, R. E., 1999: Chapter 9: Field intercomparison. *Analysis of Climate Variability: Applications of Statistical Techniques*, H. von Storch and A. Navarra, Eds, Springer, pps. 176-177.

Mason I., 1982: A model for assessment of weather forecast. *Australian Met. Magazine*, 30, 291-303.

Mason, S. J., and G. M. Mimmack, 2002: Comparison of some statistical methods of probabilistic forecasting of ENSO. *J. Climate*, 15, 8-29.

Murphy, A. H., 1988: Skill scores based on the mean square error and their relationships to the correlation coefficient. *Mon. Wea. Rev.*, **16**, 2417-2424.

New, M., M. Hulme and P. Jones, 1999: Representing twentieth-century space-time climate variability. Part I: Development of a 1961-90 mean monthly terrestrial climatology. *J. Climate*, 12, 829-856.

New, M. G., M. Hulme and P. D. Jones, 2000: Representing twentieth-century space-time climate variability. Part II: Development of 1901-1996 monthly grids of terrestrial surface climate. *J. Climate*, 13, 2217-2238.

Potts J. M., C. K. Folland, I. T, Jolliffe and D. Sexton, 1996: Revised "LEPS" scores for assessing climate model simulations and long-range forecasts, *J. Climate*, 9, 34-53.

Reynolds, R. W. and T. M. Smith, 1994: Improved global sea surface temperature analyses using optimum interpolation. *J. Climate*, 7, 929-948.

Smith M. T., R. W. Reynolds, R. E. Livezey and D. C. Stokes, 1996: Reconstruction of Historical Sea Surface Temperatures Using Empirical Orthogonal Functions, *J. Climate*, 1403-1420.

Stanski H. R., L. J. Wilson and W. R. Burrows, 1989: Survey of common verification methods in meteorology. *World Weather Watch Technical Report No. 8*, WMO/TD 358, 114 pp.

Wilks, D. S., 1997: Resampling hypothesis tests for autocorrelated fields. *J. Climate*, 10, 65-92.