PROCEEDINGS OF THE MEETING OF THE CCI OPAG3 EXPERT TEAM ON VERIFICATION

(Tokyo, Japan, 2-4 February 2005)

WORLD METEOROLOGICAL ORGANIZATION
February 2005

1. OPENING AND WELCOMING REMARKS

Dr K. Nagasaka, Director General of the Japan Meteorological Agency (JMA) and Permanent Representative for Japan with the World Meteorological Organization (WMO) opened the session and welcomed all participants to the meeting, and to the JMA. In his address, Dr Nagasaka referred to the many natural disasters and heavy socio-economic impacts felt this past year in Japan and around the world. The JMA helps citizens through provision of appropriate weather and climate information and plays a leading role in international activities through provision of global long-range prediction, training and technical cooperation support. The JMA consistently aims to improve the accuracy and reliability of seasonal to interannual prediction products, in collaboration with the research community. The work of this Expert Team (ET) on verification is particularly relevant to this effort. Dr Nagasaka urged the meeting to consider what kinds of verification are needed to make forecasts useful and reliable for the users, and wished participants a pleasant stay in Tokyo. The full text of his address is attached as **Annex III**.

On behalf of the Secretary General of WMO, Mr Michel Jarraud, Mrs L. Malone welcomed the participants to the meeting and noted the objectives for verification from the various perspectives of the Commission for Basic Systems (CBS) and the Commission for Climatology (CCl). The CCl supports monitoring and improvement of the technical accuracy and skill of the forecasts as required by the Core Standardized Verification System (the SVS) and provision of information on the inherent skill of the forecast system to the National Meteorological Services and (eventually) Regional Climate Centres (RCCs). The CCl also aims to ensure that the needs of the users (the public, the media, decision-makers in various sectors, etc.) are met, and that forecast products are provided that meet their needs and are of value to them. It was noted that this ET on Verification will work in collaboration with other pertinent CCl ETs focused on development and delivery of products and information for end-users, as well as with experts from other technical Commissions on verification issues in meeting the activities in its Terms of Reference (ToRs). The meeting was urged to ensure that its results are presented to the CCl by May 2005, so that results and recommendations can be incorporated into the documentation for the upcoming fourteenth session of the CCl, to be held in China in November 2005.

Dr Nagasaka and the JMA were warmly thanked for the excellent arrangements and support for the meeting, and for their ongoing valuable support to the WMO in many areas of responsibility.

2. ORGANIZATION OF THE MEETING

2.1 Approval of the Agenda

The agenda for the meeting was adopted with minor amendments. The final agenda is attached as **Annex I**.

2.2 Other organization matters

Dr S. Mason, chairman of the meeting, introduced the participants, including members of the ET and invited experts. Prof. S. Walker represented the CCl Open Area

Programme Group on Climate Applications, Information and Prediction Services (OPAG3), and Dr M. Fischer represented the CCI OPAG 3 ET on End-user Liaison. L. Malone acted as rapporteur for the meeting. The list of participants is attached as **Annex II**.

The working hours and work plan for the meeting were agreed. Lunch breaks would be an hour to an hour and a half, and tea breaks 20 minutes, at appropriate points in the agenda. The sessions would begin at 9 AM and end at 5.30 to 6 PM daily. At the kind invitation of the Japan Meteorological Agency, participants were pleased to attend a reception on the evening of 2 February.

3. REVIEW OF ACTIVITIES RELATED TO THE TERMS OF REFERENCE OF THE CCI EXPERT TEAM ON VERIFICATION

3.1 Brief review on establishment of CCI OPAG3 ETs

WMO provided an overview of the structure and activities of the Commission for Climatology (CCI) including the three Open Area Programme Groups (OPAGs). This ET on Verification is part of the OPAG on Climate Applications, Information and Prediction Services (OPAG 3), and its work is closely aligned with activities in the CCI OPAG 3 ETs on End-user Liaison and CLIPS Operations. As well, there are strong linkages with ETs and activities in the Technical Commission for Basic Systems (CBS), and the Commission for Agricultural Meteorology (CAgM).

3.2 Overview of the activities on different Terms of Reference of the CCI ET on Verification

3.2.1 Overall objectives of the ET, and challenges for the meeting

The chairman noted that the ultimate objectives of the ET are to facilitate the communication of forecast quality to the users of forecasts, and to produce a report detailing recommendations on how to improve the language of such communication. To achieve this goal, the following questions need to be addressed at this meeting, and in its ongoing work:

- What do forecasters need to know about how users interpret forecasts and user's understanding of forecast uncertainty?
- What attributes of forecasts are of prime interest to the user?
- To what extent does the existing SVS for LRF successfully communicate these attributes?
- What additional / alternative verification procedures could be used to communicate these attributes most effectively?

3.2.2 Terminology

The chairman reviewed the definitions of some of the key terminology related to verification, to be used during the meeting and in the further work of this ET. Dr Mason noted that there are four attributes to forecasts that distinguish "good" forecasts from "bad" forecasts, namely:

- consistency (whether the forecasts correspond with the forecaster's judgment, i.e. is what the user understands from the forecast consistent with what we think we are communicating?);
- quality (whether the forecasts correspond with the observations, i.e., does what we forecast correspond with what actually happened?);
- relevance (whether what is forecasted is of concern to the user; i.e. is what is in the forecast of interest/use to the user?); and
- value (whether the forecasts are/can be beneficial when used, i.e. do the forecasts contribute to better decisions?). This last point is not the direct concern of this ET.

Other important terms include:
- accuracy (Does what happened match with what was forecasted?);
- skill (Were the forecasts better than an alternative set of forecasts?);
- unconditional bias (Do the forecasts typically differ from the outcome in a specific direction?) (NB: Bias is one aspect of accuracy);
- reliability (Do the forecasts typically differ from the outcome in a specific direction?) (NB: in effect, this is 'conditional bias');
- resolution (Does the outcome differ when the forecast differs?) (NB: This is often confused with reliability) (cf. discrimination);
- discrimination, or the ability to distinguish different outcomes (Does the forecast differ when the outcome differs?) (cf. resolution);
- sharpness (Do the forecasts indicate a high level of confidence? Do they differ from climatological values?); and
- uncertainty (Is there much inherent doubt in what the outcome will be?) (NB: the term uncertainty is used differently in decision models, and the verification glossary should reflect when there are different meanings for terms in various fields/disciplines and address any ambiguities).

The following terms are the attributes of scores, not forecasts:
- equitable (Do all naïve forecasting strategies get the same score?);
- proper (Does the score encourage the forecaster to hedge?) (cf. consistency)

3.2.3 Progress in meeting the goals in the ToRs for the ET:

The Chairman reviewed the Terms of Reference (ToRs) of the ET, and results to date in meeting these goals:

3.2.3.1 ToR (A): To identify requirements of regional climate centres, and NMHSs for verification information on intraseasonal, seasonal and interannual predictions from the perspectives of both deterministic and probabilistic forecasts, including its presentation.

A draft survey has been constructed, with assistance from Prof. Krantz, to be sent to the regional climate centres and NMHSs. The survey requests information on the type of verification metrics and measures that are given to end users, and will be posted on the World Climate Programme CLIPS Focal Point web page. The CLIPS Focal Points will be contacted by WMO by email or by facsimile (for either Internet or hard copy response) once the survey is available for use. A previous survey by Yoshihiro Kimura focuses primarily on determining who forecasts what, and how, so there will be little opportunity for direct comparison of the results of the 2 efforts. Examples of experiences in various countries in using SVS and in communication of forecast quality to end-users are provided under agenda item 4.

Decisions and recommendations:

The meeting proposed that a deadline of two weeks be requested for completion of the survey (using a calendar date) to be followed up with a reminder where necessary. Participants further endorsed the use of other language versions of the survey, and were informed that French, Spanish, and possibly Russian translations would likely be possible. The meeting recommended that the CLIPS Focal Points be assured that their individual responses would be considered confidential, and that a summary report would be available to them shortly after the results were compiled. The results of the survey will likely reveal areas where training and other capacity-building activities are needed to support development and use of appropriate verification techniques.

3.2.3.2 ToR (B): To identify requirements of end-users in different applications sectors for verification information on products they receive.

Verification requirements by specific users and problems in communicating forecasts and forecast quality are described under agenda item 5.

3.2.3.3 ToR (C): To produce a critical review of the methods of forecast verification currently used in both deterministic and probabilistic intraseasonal, seasonal and interannual prediction, and an appraisal of the information content of these methods from the perspective of applications.

An extensive critique of the SVS is provided under agenda item 6. Various publications in the meteorological literature on current methods of forecast verification have been written as part of the ET's work, including:

Mason, S. J., and N. E. Graham, 1999: *Wea. Forecasting*, 14, 713–725.
Mason, S. J., and N. E. Graham, 2002: *Q. J. Royal Meteorol. Soc.*, 128, 2145–2166.
Mason, S. J., 2004: *Monthly Weather Review*, 132, 1891–1895.

3.2.3.4 ToR (D): To identify and develop as necessary appropriate verification techniques for seasonal to interannual predictions, as well as methods for their presentation in order to satisfy the user requirements.

Some recommendations are provided under agenda item 6. Ongoing activities are detailed under agenda item 8. Further research and discussions are necessary.

3.2.3.5 ToR (E): To promote the use of both standardized and recommended techniques.

The chairman noted that training materials on verification have been prepared for the CLIPS Curriculum and used at a number of the CLIPS Training Workshops. The materials include a course on forecast quality, a self-teaching module on ROC, and practical verification exercises. Apart from the CLIPS Training Workshops, the ROC training module has been used by the ET on End-user Liaison, and has been distributed to a number of interested parties from various NMHSs. All materials have been used and distributed at numerous other workshops. The meeting considered that this ToR is also the responsibility of the OPAG 3 ET on End-user Liaison.

3.2.3.6 ToR (F): To develop definitions of terminology used in verification in order to facilitate end-user understanding of these terms.

It was noted by the chairman that a comprehensive glossary pertinent to forecast verification has recently been published by David Stephenson in Forecast Verification: A Practitioner's Guide in Atmospheric Science, I. T. Jolliffe and D. B. Stephenson (2004), Eds., Wiley. Under agenda item 7 the ET plans to develop a first draft of a less technical version for end-users. The ET will work closely with the OPAG 3 ETs on End-user Liaison and CLIPS Operations in development of this glossary, and will consider participation in a joint effort to develop a comprehensive glossary on all aspects of seasonal to interannual forecasting.

3.2.3.7 ToR (G): To advise the implementation/Co-ordination Group, and submit reports in accordance with timetables established by the C-OPAG and/or Management Group.

The meeting was informed that the ET lead has regularly submitted reports when requested, as required.

3.2.3.8 (H) To maintain close links with CBS, CAS, CHy, CAgM, WCRP and with other CCI ETs on the issues involved.

The ET lead, Dr. Simon Mason, has been closely involved with the CBS Expert Team on Verification, and attended a recent CBS ET meeting in Geneva in November 2004. Dr. Mason has interacted closely with the CCI ET on End-User Liaison, and a representative of that ET was invited to work with the ET on Verification at this session. Close ties with CAgM are met through the participation and efforts of Dr Meinke and Prof. Walker, both active in related activities of that Commission. Some collaboration has begun with the OPAG 3 ET on CLIPS Operations, particularly on development of the glossaries under development by both ETs.

3.2.3.9 ToR (I): To prepare a review of methods for the assessment of forecast quality and to recommend methods for verification of intraseasonal, seasonal and interannual forecasts.

This activity is amongst those pursued in the interactions with the various other relevant commissions and ETs. It was noted by the Chairman that this activity may spawn some material for journal publication. Within WMO, there could be a report (perhaps a technical note) developed. Later, guidelines for users in NMHSs could be established.

3.3 Recommendations

The meeting recommends that the CCI MG propose to CCI Members that this ET be re-established with its current membership at CCI-XIV (November 2005), in order to complete the work under ToRs (d) and (e) in particular, and to fully develop the needed liaisons with the CCI ETs on End-user Liaison and CLIPS Operations.

This ET supports the development of new forecast products as requested by the end users and recommends that the ET on End-User Liaison be urged to inform this ET on such developments, so that appropriate verification procedures can be defined to quantify the usefulness and inherent uncertainty in those products.

The meeting recommends that close collaboration be maintained with the CBS ET on Verification (and other relevant groups) on future developments of the SVS. Further, the ET recommends that collaboration be further developed with groups such as the CAgM ET on Impacts of Climate Change and Variability on Medium-to-long range Predictions for Agriculture.

4. EXPERIENCES IN COMMUNICATING VERIFICATION TO END USERS

4.1 Use of the CBS SVS by the Japan Meteorological Agency (JMA)

Mr Maeda introduced the experience in implementing the WMO CBS SVS by the JMA. The verification indicated that products of the JMA's dynamical ensemble seasonal prediction system are generally informative in the probabilistic way, but not in the deterministic way. In order to make more reliable tercile probability predictions from the JMA GPV, a calibration method is being developed in the JMA. In the research, it is found that the skill is crucially dependent on scores used for the evaluation. This experience highlights the need for multiple verification scores.

4.2 Verification of forecasts within West Africa

Seasonal forecasting in West Africa was started in 1998 during the climate outlook forum at Abidjan in Ivory Coast organized by ACMAD and other prediction centers. Seasonal forecasting in West Africa is based on linear regression. SSTs are used as predictors, specifically the NINO3 index, SST averages in the Equatorial Atlantic, and the third principal component of the variation of the global oceans.

Efforts have been made to evaluate the West Africa regional forecasts as seen in the example of Chad. The first way is to compare the forecasted rainfall with the observed rainfall at a regional level. The second way is to compare the observed rainfall at individual stations with the forecasted rainfall for the respective region after transforming the forecast to the station's climatology.

An important lesson is that the verification procedure needs to be understandable. One suggestion is to develop verification games in which the end users are given previous forecasts, asked to make decisions in a role-playing exercise, and then shown the verification.

4.3 Methods of deterministic forecast verification and their practical use in the Russian Federation

The different end users prefer forecasts of different meteorological parameters. Some prefer raw values of meteorological elements (heat and electric station, building, transport), but more useful are special complex meteorological indexes such as forest-fire index, drought index, and biometeorological index. Forecasts may be universal for all users or tailored for different end users. Universal categorical forecasts should include meteorological value, the normal value, standard deviation, comparison of forecast value to the normal value (near normal, above normal or below normal). It is very useful for the end-user to provide an observation of the value of the past year and to show the tendency of the value compared to the value of last year (increase or decrease).

Forecasters currently use measures of forecast quality that include percent correct, success index, hit rate, correct rejection rate, correctness of anomaly sign, mean error, mean square error, etc.

Tailored forecasts with estimation of forecast quality for the end users can be more effective. Errors in forecasts have different costs for different end users. User-specific cost loss tables could be constructed. This type of client-specific product will be of value to many users, but at present some are not yet ready to use them.

4.4 Communicating forecast quality to the energy sector

Communicating forecast results highly depends on the experience and the mathematical/statistical background of the end-user. In the field of energy demand forecasts end users do not pay much attention to the verification of meteorological forecast products. For energy demand predictions the mean error is not really important as long as it remains within acceptable limits. However, the maximum forecast errors are very important for end users. It is most relevant for them, to know how often the forecast error exceeds certain thresholds.

The transfer function between meteorological variables and energy demand is highly non-linear. It is therefore not possible to transform a primary error or probability distribution function into a corresponding secondary probability function. In order to do so the transfer model has to be run with individual meteorological forecasts (ensemble members) to create an energy demand forecast ensemble. Verification is applied directly to the energy demand forecast ensemble.

5. REQUIREMENTS OF END USERS IN DIFFERENT APPLICATIONS SECTORS FOR VERIFICATION INFORMATION ON PRODUCTS THEY RECEIVE

5.1 Probabilistic forecasts

There was consensus amongst ET members that deterministic climate forecasts are misleading and should neither be considered nor endorsed.

> *… although we still have much to learn about the underlying physical processes we now appreciate that climate has many chaotic and non-deterministic features, which will prevent us from ever achieving complete certainty in climate forecasting. Any deterministic forecasting system is therefore either wrong or misleading and should not be endorsed….* (sic, Meinke and Stone, 2005)

In the early days of seasonal climate predictions, and as a consequence of how weather forecasts were delivered, prediction information was issued as so-called deterministic forecasts (often also inappropriately referred to as categorical forecasts), where one category of rainfall (for example, much below average, below average, or average) was predicted to occur. Based on a ‘daily verification in real time’, recipients of such weather forecasts intuitively interpret such forecasts as probabilistic, i.e., they accept that there will be error associated with these predictions.

Such experience cannot easily be developed for seasonal forecasts, where the outcome is only realized once or twice a year. Proponents of deterministic forecasts point to the perceived advantage in providing a definitive yes–no statement. However, the perceived difficulties in communicating probabilistic information must be balanced against the possibility of creating a misleading impression of confidence via an absolute (deterministic) statement and the probability of being dismissed as irrelevant after issuing a series of definitive statements that proved to be false. Uncertainties that are inherent in seasonal predictions must be reflected in forecasts. Murphy (1983) states that the widespread practice of ignoring uncertainties when formulating and communicating forecasts represents an extreme form of inconsistency and generally results in the largest possible reductions in quality and value of the forecast.

There is an increasing awareness regarding the necessity to communicate the known level of uncertainty via probabilistic statements. This is not only logical from a user's risk management perspective, it also protects against any potential liability (corporate risk) that organizations expose themselves to through the provision of inadequate or misleading advice.

There are several reasons why probabilities are an integral part of risk management:

- scientists know that chaos plays a large role in climate systems; in fact, the atmosphere frequently acts like a random number generator such that absolute statements justifiably cannot be made;
- scientists have a responsibility to communicate their degree of knowledge as well as their ignorance;
- scientists are not the decision makers; all they can provide is discussion support. The ultimate decision rests with the practitioner. Rather than providing a recipe, scientists should present the choices, the chances and the consequences (Hayman, 2000);
- 'dumbing down' the message can lead to poorer risk management where farmers plan for a single most likely outcome rather than a range of possible outcomes;
- assuming that probabilistic concepts are too difficult to understand is insulting to users who are already good risk managers who have managed effectively under uncertainty all their lives;
- deterministic forecasts create a sharp and artificial boundary between areas where there is and there is not sufficient scope to make a prediction, thus placing decision makers operating near these boundaries in a quandary.

5.2 Saliency

Saliency, as defined by Cash and Buizer, is the need for relevant forecast products (called 'salience' by Cash and Buizer). To understand the user's needs deeply, it is useful to consider some general principles of decision-making. Forecasts can be used either in <u>selection</u> of plans, or in their <u>implementation</u>.

Plan selection generally makes use of judgments about the chances of one or several critical events. A very simple example would be a choice between two plans: (A) planting rice in a certain field versus (B) planting maize in that field. The critical event is ample seasonal rainfall (which might be enough for rice, and excessive for maize). For this plan selection, a probabilistic forecast about total seasonal rainfall might be seen as directly relevant. A different example would be the choice between (A) planting cotton early versus (B) waiting until the rains are well established. A probabilistic forecast about the timing of steady rains might be seen as directly relevant.

For more complex plans, which involve elaborate system design, the critical events that are judged might be more abstract and/or more complex. Staying in the domain of agricultural examples, Plan (A) involving consideration of how much to plant of different crop types or seed types, when to plant, how much fertilizer to use, etc., might be selected if the user judges that the chance of achieving each of several important goals is high. Achievements of various overall goals are the critical events considered. In this scenario, Plan (A) might include use of one or more continuous forecast variables to adjust quantities. Here one sees how climate forecasts might be used in the implementation of a plan, rather than in plan selection. Verification information about the forecast method could be important in leading the user to select Plan (A).

The differences between the critical events that are judged in the rice/maize decision and in the cotton planting decision illustrate the need for relevant forecast products. A forecast about total seasonal rainfall might be relevant to one plan selection but irrelevant or misleading for the other. A highly credible but irrelevant forecast can put users in a difficult situation: thus forecasts need to be accompanied by explanations concerning their potential use.

The examples also illustrate that both probabilistic categorical forecasts and continuous forecasts can be useful.

5.3 Artificial skill

A further concern is the issue of ‘artificial skill’, which arises particularly when a multitude of possible forecasting schemes can be employed and a number of ‘best’ performing schemes are selected on the basis of some test procedure or statistics, rather than first principles. This issue is complex and goes beyond the ToRs of this ET. However, we feel that we must at least flag it as a problem that needs to be addressed.

Artificial skill is the apparent hindcast skill in forecasting schemes arising from chance. It is the skill that does not survive when the forecasting scheme is applied in real time to new or independent data (the need for cross validation including definitive guidelines on how cross validations should be performed is discussed below). It can arise from many sources such as when models are over-parameterized using a number of cross-correlated predictors or when a multitude of possible, statistical forecasting schemes can be employed and ‘best’ performing schemes are selected on the basis of some test statistics, rather than first principles. For instance, countless seasonal forecasting schemes based on multiple regression techniques have been developed. Many of them do not attempt to capture the underlying dynamics and are hence prone to capitalizing on random fluctuations in samples of limited size. However, the issue is equally important for dynamic prediction schemes where artificial skill can arise for a multitude of reasons (e.g., cross-correlated parameters; inappropriate use of statistics; inadequate or over-parameterization and many more).

Some of the traps leading to artificial skill in forecasting systems include:

- the smaller the sample size, the greater the chance of spuriously exceeding a given skill level;
- the greater the number of potential predictors, the greater the chance of artificial skill;
- the more combinations used, the more spurious relationships found;
- that the atmosphere acts in part as a random number generator.

6. THE WMO CBS SVS

6.1 Overview of the WMO CBS SVS

Dr S. Mason presented a description of the Standardized Verification System (SVS) for Long-Range Forecasts, drafted under the auspices of the Commission for Basic Systems (CBS). Dr S. Mason outlined the original objectives in drawing up the SVS, and noted that these objectives had focused on the interests of the modelers and the forecasters. The parameters to be verified were described, and there was some discussion about the need for verification at the station-level. It was explained that cross-validation is required for all statistical forecasts, but it was agreed that more detailed guidelines on how to perform the cross-validation would be desirable. There are detailed descriptions of the various bootstrapping and significance testing procedures that should be conducted. The individual recommended scores were then discussed.

6.2 The limitation of CBS SVS for end-user needs

The SVS provides a detailed set of recommendations for the exchange of verification scores. It is primarily targeted for communicating forecast quality between forecast modelers and forecast practitioners, and so does not explicitly address the specific interests of the end-users of long-range forecasts. Identified weaknesses in the SVS from the perspective of the end-user were discussed. Questions pertaining to the pragmatics of the communication of forecast quality to the users are not addressed (these are covered under the auspices of the ETs on End-User Liaison and CLIPS Operations), but only the content of such communication.

Brier and Allen (1951) and Jolliffe and Stephenson (2004) provide extensive discussions on the purposes of forecast verification. Of these, only a few are of direct concern to the SVS, but it is important to identify these verification objectives from within the specific context of the CBS, and to compare them with the objectives of verification within the context of the concerns of the CCI ET on Verification. Important objectives in the CBS's SVS include the scientific objectives of improving the forecast models, understanding their strengths and weaknesses, comparing models, and communicating to the forecaster information to assist in making an intelligent combination of various forecast products. From the CCI ET's perspective, these objectives create two problems: the needs for information on forecast quality by the end-users of the forecasts are not explicitly addressed; the various verification scores and procedures are not necessarily presented in formats that are user-friendly. These two problems are discussed in more detail below.

Ultimately, the end-users want to know whether long-range forecasts can be used to economic or societal benefit, which raises questions of whether the forecasts have value. The focus of this ET is not to address the actual value of the forecasts themselves, but rather to focus on a prerequisite for forecast value: do the forecasts, as they stand, contain any useful information that might, at least in certain decision-making contexts, effect some benefit. Forecast quality is a sine qua non of forecast value, and the focus of this ET is on forecast quality. The subject of verification is thus the actual forecasts of meteorological parameters, not the outputs of application models that use meteorological forecasts as inputs. The important objective in the context of our ET is: to communicate the appropriate level of confidence that the user can place in the forecasts. For such communication to be effective, forecasters need to understand what attributes of forecasts are important to end-users, so that the user's questions are adequately addressed. Forecasters also need to identify how various verification metrics and procedures can be used to address the questions directly.

The second problem with the SVS from this ET's perspective, as stated above, concerns the format of the presentation of verification information. While many of the recommended scores do address questions of interest to the users, it is not always obvious how they do so. Interpretations of the verification information need to be provided. However, one difficulty is that there is a very wide range of user requirements and sophistication of understanding. Questions range from as simple as "How often are the forecasts correct?" to "What information about the uncertainty in the forecast does the ensemble distribution provide?" To meet the user requirements for verification information adequately, this full range of interests needs to be addressed. Thus the objectives of obtaining the user's initial interest in the forecasts, through building a degree of confidence from the user, to providing a comprehensive assessment of forecast quality, are all considered appropriate for this ET's work.

6.2.1 Parameters to be verified

The SVS recommends that verification be performed on the following parameters:

- surface air-temperature anomalies;
- precipitation anomalies;
- sea-surface temperature anomalies.

It is implicit that these are taken as temporally (typically 3-month) averaged values. Most users of long-range forecasts request additional or tailored information, and it is important that recommendations and considerations be addressed for *all* forecasts available to the end users. In cases where the forecast target represents frequencies some of the SVS recommendations may be inappropriate. The CCI ETs on End-User Liaison and on CLIPS Operations including Product Generation should be consulted to determine what type of forecast targets are being proposed. Ideally the appropriate verification procedures should be in place before new forecast products are made available to the forecast users. It may be necessary to recommend that this ET be reconstituted after CCI-XIV to ensure that such verification standards are in place.

6.2.2 Aggregated verification

Attention has to be paid when comparing general circulation model (GCM) output with observations. Model variables do not necessarily correspond to observed variables, even if they represent the same physical quantity. General circulation models approximate the dynamics of the atmosphere or the ocean with a limited spatial resolution. Model variables are computed on discrete spatial points, which has important consequences for the meaning of modeling results:

- model variables like temperature or precipitation do not represent a local physical quantity at a certain point but instead an average over the area of a grid box;
- due to limited spatial resolution small-scale orography cannot be represented and hence small-scale orographic effects cannot be resolved by a GCM.

For model verification it is therefore important to ensure that only quantities are compared that correspond to each other. Either observations should be transformed, such that they correspond to model variables, or the GCM output must be mapped onto observed quantities. The SVS recommends the assessment of model performance using surface air temperature anomalies, precipitation anomalies and sea surface temperature anomalies. The following points should be kept in mind.

- surface air temperature anomalies
    - is the observed temperature influenced by small-scale orography?
    - does the observed temperature represent a sufficiently large area or is it possible to average observations over an area that corresponds to the model resolution?
    - does the altitude in the model correspond to the altitude of the observational point?
    - are the temperature measurements influenced by human activities?

- precipitation anomalies
    - model precipitation represents the average over a grid box, not local precipitation
    - depending on the model resolution extreme events, like strong cold fronts, thunderstorms, cyclones and the related rainfall cannot be simulated

- sea surface temperature anomalies
    - are measurements influenced by costal effects (costal upwelling, rivers, etc)?
    - is it possible to average sea surface temperature observations over an area that corresponds to the model resolution?

Verification of large-scale averaged climate anomalies is of minimal direct use to most users of forecasts. While grid-point averages are of much more interest than the large-scale averages, ideally the verification should be performed at the scale of specific locations, which is more consistent with the scale of most forecast applications. If the application of long-range forecasts is to be promoted, the quality of the forecasts at the scale at which they are to be used does need to be communicated honestly, rather than being effectively exaggerated through being considered at a spatial scale that is relatively easy to forecast accurately, but otherwise of little relevance. Similar arguments can be made for temporal downscaling of forecasts.

6.2.3 Verification Strategy

Any conditions under which forecast quality is likely to alter (positively or negatively) should be communicated to the user. One attribute of forecast quality that is not addressed adequately by the SVS is the prediction of forecast accuracy. This topic is discussed in greater detail below.

Interpretations of all verification scores and procedures should be provided. For most users it is not obvious how many of the scores address their own questions about forecast quality.

6.2.4 Verification scores

The SVS discusses a number of verification scores that are based on evaluating a "deterministic" (meaning single scenario, hereafter point) forecast. The relevance of such scores to users, modelers or theoreticians needs to be better understood; or, more pertinently, the relevance of such *forecasts* to users, modelers or theoreticians needs to be better understood. Jolliffe and Stephenson say it is more realistic to interpret such forecasts as "probabilistic forecasts in which the uncertainty information is not provided." In the case of probabilistic forecasts discussed here, where uncertainty information *is* provided, one might rightly question why it should ever be suppressed by the verification procedure.

This is not to suggest, of course, that some model-run based statistic corresponding, say, to a model-run ensemble mean might not be useful in computing the probability forecast; rather the point is that there is no reliable method to determine either the skill or the value of a probabilistic forecast from the analysis of *any* point forecast extracted from it, without a host of additional assumptions regarding both the probability forecast itself and the underlying physical process being forecast (as well as, for that matter, the user application and utility). In short, we recommend that probability forecasts be evaluated as such, and that the current focus on the verification of the ensemble mean be discouraged.

Comments are made on the specific scores recommended in the SVS. In all cases verification scores should not be presented as a set of statistical metrics that are supposed to measure forecast quality, but should rather be presented as direct answers to questions raised by users about forecast quality.

6.2.4.1. Mean-square skill score (MSSS)

This scores attempts to measure forecast accuracy and skill at the same time. For the end-users it would be better to have a score that specifically measures accuracy only, and to have a corresponding score for the reference strategy so that skill can be inferred rather than defined explicitly. Thus it is more useful to have answers to the two questions "How good are the forecasts?" and "How good are the reference forecasts?" rather than an answer to the single question, "How much better are the forecasts compared to the reference forecasts?" By providing the score and the score for the reference strategy, the units of the original forecast are retained, whereas the skill score is a dimensionless number, which is more difficult to interpret.

It is not obvious that the squaring of the score is appropriate, since it implies an exponentially increasing cost at a rate that is essentially arbitrary. Some users prefer a logarithmic function. The mean absolute error (MAE) would be a more intuitive score to use, and a simple difference between the MAE and the MAE for the reference forecasts has an obvious interpretation.

A primary problem with the MSSS and MSE is that inferences about the frequency of large errors cannot be made unless precise distributional assumptions are met. These assumptions invariably are not met in seasonal climate forecasting anyway. Further, in the case of the distribution of forecast errors being skewed, it is not clear that the MSE or MSSS is an appropriate measure of accuracy. While the mean absolute error may be a more intuitive score, ultimately the user is interested in how often errors of certain magnitudes can be expected. A score or set of scores that directly address the latter question would be useful.

6.2.4.2. Contingency tables

While contingency tables are easy to understand, they are problematic because they retain the deterministic language of forecasts, and encourage a deterministic interpretation. The construction of contingency tables by considering the category with the highest probability as the forecast category in a deterministic forecast system should be discouraged for these reasons. Another weakness is that the forecast errors are defined only after the categorization, and so errors within a category may be larger than between two categories. (In addition, since the two outer categories are unbounded, it is even possible for an error that spans these two categories to be larger than an error within one of the categories). Tables showing the frequencies of errors of certain magnitudes would thus form a useful supplement. For skewed data, it would be beneficial to have the tables that maintain the sign of the errors.

6.2.4.3 Gerrity score

This score has useful statistical attributes and seems to be one of the most suitable of the summary measures of the contingency table. However, in addition to the general criticisms of the contingency table given above, it suffers from a lack of intuitive and interpretative appeal. How does one define what a Gerrity score of 40% means? Some guidelines on what the score means should be provided, even if this can only be done so by making only general statements about what combination of attributes the score captures. This lack of interpretability ultimately stems from the attempt by the score to measure too many attributes at once. Effectively it is trying to answer the ultimately vague, or poorly defined question, “How good are the forecasts?”

It is recommended that, when addressing questions of forecast quality, the individual attributes of forecast quality be considered separately, and that summary measures only be presented afterwards. Apart from promoting the realization that forecast quality is a multi-faceted problem, such an approach would succeed in addressing well-defined user questions upfront, and would provide a sounder basis for presenting a less interpretable, but more encompassing over-all measure of quality.

6.2.4.4. Hit and false-alarm rates

Again, as in the discussions with the contingency tables and Gerrity scores, the calculation of hit and false-alarm rates should be discouraged, except in the context of the ROC, discussed below.

6.2.4.5. Relative (or Receiver) Operating Characteristic (ROC)

While the ROC does complement decision-making models such as the cost-loss model, forecasters should be under no illusions about the extent to which such models are used in real life: they are vastly oversimplified. In addition, the cost-loss model itself involves an inappropriate indication of risk management. However, the ROC can be used to address the question of "How often are the forecasts correct?" in the context of probabilistic forecasts. While this question is ultimately poorly formed a suggested approach is detailed below, rather than dismissing the question out-of-hand.

In order to be more useful to the forecast user, the ROC needs to be presented in a much more user-friendly manner. For one, its name is distinctly unhelpful in most contexts, and so should be downplayed. Secondly, in the context of probabilistic forecasts, the terms, hit, miss, false alarm, and correct rejection are inappropriate and misleading, and alternatives should be found. After making a standard response to the impossibility of probabilistic forecasts being right or wrong (when the probabilities are not exactly zero or one), the question "Was the forecast correct if a probability of above *p*% is treated as a (deterministic) forecast (or warning) for that category?" can be answered. The ROC proceeds simply by changing *p*. Thus, for example: "If I treat a probability of greater than 80% as a warning of above-normal temperatures, how often is such a forecast correct?" And: "How often is it correct when I assume that a probability of 60% or more constitutes a warning?" etc. By spelling out the questions for each point on the ROC graph, it could be a more effective tool for communicating forecast quality. It would also be useful to indicate on the table accompanying the graph how many warnings are effectively issued at each threshold. To provide such information on the graph itself would clutter it unnecessarily.

These scores directly address a number of user questions, and it would be helpful to make these questions explicit. For example, in the context of forecasts for above-normal temperatures, the hit rate defines the number of cases of above-normal temperatures that were forecasted successfully. Mason and Graham (1999) provide a more detailed discussion. It may be worth pointing out the distinction between the hit rate and the correct alarm ratio (the latter defines how many times forecasts of above normal temperatures were correct). The analogue to the false-alarm rate is the miss ratio, although the false-alarm ratio should also be considered; it may be helpful to draw the distinction between these two scores and the false-alarm rate.

Further details of interpretation of the ROC are provided by Mason and Graham (2002). It is demonstrated that the area beneath the curve constitutes a measure of discrimination. While this measure is of interest to the forecaster, it is of less interest to the user because of the highly artificial nature of the two-alternative forced choice question.

It has been suggested that ROC diagrams could be constructed for deterministic forecasts, but that they would have only one point. It is true that only one point could be drawn if the forecasts themselves are for a binary discrete outcome, but multiple points could be drawn if the forecasts are themselves on a continuous scale (or on an ordinal discrete scale). Take as an example a situation where above-normal temperatures are defined as anomalies exceeding +1°C. Now assume two forecasts on a continuous scale, one of which is for an anomaly of +0.5°C, and the other for –0.5°C. Neither of these forecasts are for above-normal temperatures, but in the absence of any other information about the degree of confidence in these two forecasts, it seems quite reasonable to assume that above-normal temperatures seems a more unlikely outcome given the forecast of –0.5°C than given the forecast of +0.5°C. An ROC diagram could then be drawn in which the thresholds used to define warnings are not %s but actual anomalies in °C. More generally, recommendations for verification procedures when the forecast is on a continuous scale but the verification is discrete need to be made.

A primary problem with the ROC is the fact that it fails to address forecast reliability. This limitation should be pointed out to the user. In fact, if there are a sufficiently large number of forecasts, the reliability diagram should be used in preference to the ROC. (See further comments on the reliability diagram below).

6.2.4.6. Reliability diagrams

Reliability diagrams can indicate a full range of forecast attributes, and the focus on reliability is an advantage since it indicates whether the probabilities correspond with user's intuitive interpretation. However, they should be explained carefully since the axis labels are not always easily translatable into user questions, and when frequency histograms are included (which they should be) the diagrams can at first appear cluttered. Reliability diagrams must be presented with error bars, and there must be a standard set of procedures for drawing the curve to avoid artificially optimising the impression of reliability.

It is recommended that the reliability diagram rather than the attributes diagram be drawn (Hsu and Murphy, 1985). The attributes diagram includes a diagonal line half way between the perfect reliability diagonal and the horizontal no-resolution line. This diagonal is labelled the "no skill" line, and it is implied that a reliability curve below the no-skill line provides no useful information to the forecast user. The no-skill line is an artificial construct related to the requirement that the resolution term in the Murphy (1973) decomposition of the Brier score be larger than the reliability term. Such a requirement is misleading, and relates to the lack of equitability in the Brier score (Mason, 2004). Since forecasts with a non-zero resolution term are potentially usable, any positively sloping reliability curve indicates "good" forecasts.

6.3 Other considerations

The SVS provides a set of recommendations for verifying forecasts of discrete outcomes that are forecasted either on a discrete scale or on a probabilistic scale. It also provides recommendations for verification of deterministic forecasts on a continuous scale. However, no recommendations are made on how to verify probabilistic forecast on a continuous scale. The verification of ensembles of forecasts is thus inadequately addressed.

Given probabilistic forecasts of continuous variables, of two specific questions of interest, namely "How accurate is the location of the centre of the ensemble?" and "Is there any information in the ensemble distribution?" only the first can be addressed by the SVS (the relevant recommended score is the MSSS, and limitations of this score have been discussed above). In the context of the first question, it seems reasonable to use the ensemble mean, although a case for preferring the ensemble median could be made, especially in the context of skewed data distributions.

The second question, which relates to the information in the ensemble distribution, has received minimal attention, even in the more technical meteorological literature. So called Talagrand diagrams, or binned probability histograms, provide some information, but are too dominated by the information in the central tendency to provide much insight on the information in the ensemble distribution. Forecast accuracy – ensemble spread relationships address the question more directly, and typically take the approach of correlating the ensemble mean error with the ensemble variance. The standard objections in the context of skewed data can be raised. In addition, the ensemble variance may be a function of the ensemble mean, especially when data values are bounded (as in the case of precipitation), and so a correlation with accuracy may reflect the skill from the ensemble mean. Some recent published work addresses some of these concerns and the CCI ET on Verification is currently investigating a wide range of possible accuracy – spread metrics.

More generally, the need to verify forecasts as generated by the model need to be recommended. The majority of discussions on verification tasks start from the assumption that the forecast had been issued in some mathematically convenient format: a collection of N probabilities assigned to set of N distinct outcomes, a pdf on a continuous variable, and so on. The translation of a model forecasts into these mathematically convenient formats is well understood for common univariate statistical models in the limit of large sample sizes, all of which are largely irrelevant for seasonal forecasting.

For dynamical models, casting the forecast from a (set of) model simulation(s) to some such standard form can destroy much of the information in the simulations; since it is not clear how to best make this translation, there is justifiable interest in forecast verification scores more closely related to the raw scenario simulations and users needs. *A better understanding of such methods is needed*. One example from seasonal ensemble forecasting is the Bounding Box Capture Rate, which reflects the frequency with which this verifying observation falls within the range of the set of forecasts. This can be interpreted directly as a measure of relevance (of the forecast) to a given application by users. These methods usually have the advantage of leaving the verification procedure and forecast in the language of conceptually meaningful units (degrees, millimetres, etc). It has the disadvantage that the model simulations must be cast into an empirical form ("bias corrected"), and the at the statistical properties of these more immediately relevant measures are not as well understood; the corresponding statistical properties which apply under assumptions known not to hold are well understood. The former may prove of greater value to both the user and the model developer, while the later is often of more interest and relevance to the academic statistician.

A forecast stated as a probability distribution function (pdf) on the continuous (empirical) target variable(s) can be evaluated directly; as well as cast into other formats (say, the probability of various ranges or categories) evaluated in that form, with (of course) a loss of information that may prove critical to the end-user.

Verification scores for pdf forecasts of the continuous variable x are functions of the forecast pdf, and the observations (verification). A given score will be either proper or improper, either local or nonlocal. The functional form of the score determines both of these properties. In expectation, a proper score will always assign the pdf the best score under the assumption that the pdf is accurate. A score is local if and only if it depends only on the forecast probability assigned to the verifying observation. There is only one score for continuous variables that is both local and proper, this score is was introduced by Good (1952), and deserves further consideration by the ET.

7. WORKING GROUPS TO DRAFT RECOMMENDATIONS FOR THE CCL SVS FOR END USER NEEDS, AND TO DEVELOP EASY-TO-UNDERSTAND GLOSSARY

Various members of the team made extensive comments on the technical glossary developed by David Stephenson. Various draft definitions were prepared at the meeting. These are to be collated, developed, and discussed by the Team with a deadline of May 2005 for completion. Additional comments are provided under section 6.3.

8. PREPARE A REVIEW OF METHODS FOR THE ASSESSMENT OF FORECAST QUALITY AND TO RECOMMEND METHODS FOR VERIFYING INTRASEASONAL, SEASONAL AND INTERANNUAL FORECASTS

A detailed assessment of the SVS was provided under agenda item 6, and some suggestions for alternative and additional recommended scores were provided. Further research and discussion is required before a firm set of recommendations can be made. Collaboration with the ET on End-User Liaison will be required so that appropriate verification procedures can be defined to quantify the usefulness and inherent uncertainty in new forecast products.

9. ANY OTHER BUSINESS

No other business was recorded.

10. MEETING REPORT

It was agreed that the meeting chairman, ET lead Dr S. Mason, and the rapporteur, L. Malone, would assemble the contributions of the participants before and during the meeting, and develop the first draft of the report by 11 February. This would be distributed to participants for review and comment shortly thereafter.

11. SUMMARY OF RECOMMENDATIONS

Recommendations from the meeting, compiled, approved by and shared with the participants before the close of the session, were:

11.1 This ET recommends that the web survey developed under ToR (A) be made available in four languages: English, French, Spanish, and Russian, to ensure the best possible response.

11.2 This ET recommends to CCL MG that this ET be re-established at CCI-XIV (November 2005), in order to complete the work under ToRs (d) and (e) in particular, and to fully develop the needed liaisons with the CCI ETs on End-user Liaison and CLIPS Operations.

11.3 This ET supports the development of new forecast products to meet the needs of the end users and urges the ET on End-user Liaison to inform this ET on such developments, so that appropriate verification procedures can be defined to quantify the usefulness and inherent uncertainty in those products.

11.4 This ET recommends that close collaboration be maintained with the CBS ET on Verification (and other relevant groups) on future developments of the SVS. Further, the ET recommends that enhanced collaboration be developed with groups such as the CAgM ET on Impacts of Climate Change and Variability on Medium- to-long range Predictions for Agriculture.

11.5 Members of this ET recommend that all forecast products be presented in probabilistic format. The ET also recommends that the CCI MG ensure that the information in these forecasts is effectively conveyed to both climate service specialists and end-users. This may involve capacity-building activities, and the ET urges WMO support to meet any required training or information needs.

11.6 In appreciation for the excellent logistical and organizational support for the meeting, WMO will send a letter of thanks to the DG JMA.

11.7 The ET noted that support for the work of Expert Teams represents a considerable contribution to WMO, and that some ET members are not from National Meteorological and Hydrological Services. The ET therefore recommends that, as appropriate, such in-kind support be acknowledged to the home organization of the ET participants, to encourage continuation of that support to meet future needs of the Commission.

12. CLOSURE OF THE MEETING

Dr. Mason and Mrs. Malone, on behalf of the WMO, warmly thanked the participants for their contributions, including those specifically for and at the meeting, and also expressed sincere appreciation to the hosts for the session, the Japan Meteorological Agency, for the excellent logistical arrangements and support throughout the planning and implementation of the meeting. The meeting was closed at 3:30 PM on 4 February.

Annex I

MEETING OF CCI OPAG3 EXPERT TEAM ON VERIFICATION
Japan Meteorological Agency
Tokyo, Japan
2-4 February 2005

Agenda

1. OPENING AND WELCOMING REMARKS

2. ORGANIZATION OF THE MEETING

2.1 Approval of the Agenda
2.2 Other organization matters

3. REVIEW OF ACTIVITIES RELATED TO THE TERMS OF REFERENCE OF THE CCI EXPERT TEAM ON VERIFICATION

3.1 Brief review on establishment of CCI OPAG3 ETs
3.2 Overview of the activities on different Terms of Reference of the CCI ET on Verification

4. EXPERIENCES IN COMMUNICATING VERIFICATION INFORMATION TO END USERS

5. REQUIREMENT OF END USERS IN DIFFERENT APPLICATIONS SECTORS FOR VERIFICATION INFORMATION ON PRODUCTS THEY RECEIVE

6. The WMO CBS SVS

6.1 Overview of the WMO CBS SVS
6.2 The limitation of CBS SVS for end-user needs.

7. WORKING GROUPS TO DRAFT RECOMMENDATIONS FOR THE CCL SVS FOR END USER NEEDS, AND TO DEVELOP EASY-TO-UNDERSTAND GLOSSARY.

8. PREPARE A REVIEW OF METHODS FOR THE ASSESSMENT OF FORECAST QUALITY AND TO RECOMMEND METHODS FOR VERIFYING INTRASEASONAL, SEASONAL AND INTERANNUAL FORECASTS.

9. ANY OTHER BUSINESS

10. MEETING REPORT

11. SUMMARY OF RECOMMENDATIONS

12. CLOSURE OF THE MEETING

Annex II

WORLD METEOROLOGICAL ORGANIZATION
MEETING OF THE EXPERT TEAM ON VERIFICATION
TOKYO, JAPAN,
2 – 4 FEBRUARY 2005

LIST OF PARTICIPANTS

Australia

Dr Holger MEINKE
Department of Primary Industries and Fisheries (DPI & F)
P.O. Box 102
Toowoomba QLD 4350
Tel: +61 7 4688 1378 (Office)
Fax: +61 7 4688 1193 (Office)
Email: holger.meinke@dpi.qld.gov.au

Brunei Darussalam

Ms Rokiah H. ANGAS
Brunei Meteorological Service
Department of Civil Aviation
Ministry of Communications
BB2513
Brunei International Airport
BANDAR SERI BEGAWAN
Tel: +673 2 332 735
Fax: +673 2 332 735
Email: rokiah_angas@civil-aviation.gov.bn
meteobru@brunet.bn

Chad

Mr Samuel MBAINAYEL
Direction des Ressources en eau et de la Météorologie
P.B. 429
N'DJAMENA
Tel: +235 52 3081/235 52 6001
Fax: +235 523 043
Email: s_mbainayel@hotmail.com
Mbainayel10@yahoo.fr

Japan

Mr Shuhei MAEDA
Japan Meteorological Agency
Climate Prediction Division
1-3-4 Otemachi
Chiyoda-Ku
TOKYO, 100-8122
Tel: +81 3 3211 8406
Fax: +81 3 3211 2032
Email: smaeda@met.kishou.go.jp

New Caledonia

Dr Martin FISCHER
MIKAN
27 rue de Sébastopol
BP 12270
98802 NOUMÉA CEDEX
Tel: +687 862 593
Fax: +687 262 824
Email: mfischer@mikan.nc

Russian Federation

Dr Boris SHERSTYUKOV
Russian Research Institute for Hydrometeorological Information World Data Center (RIHMI-WDC)
6, Korolyov str.
OBNINSK 249035
Tel: +7 08439 74690
Fax: +7 095 255 2225
Email: Boris@meteo.ru

South Africa

Prof. Sue WALKER
University of the Free State
Department of Soil, Crop and Climate Sciences
Box 339 (int.60)
BLOEMFOENTEIN, 9300
Tel: +27 51 401 2222
Fax: +27 51 401 2212
Email: WalkerS.SCI@mail.uovs.ac.za

United Kingdom

Prof. Leonard A. SMITH
London School of Economics
Centre for the Analysis of Time Series
Houghton Street
LONDON WC2A 2AE
United Kingdom
Tel: +44 207 955 7626
Fax: +44 207 955 7416
Email: lenny@maths.ox.ac.uk

United States

Prof. David H. KRANTZ
The Earth Institute / Center for Research on Environmental Decisions
Department of Psychology
Mail Code 5501
Columbia University
NEW YORK, NEW YORK 10027
Tel: +1 212 854 4534
Fax: +1 212 854 3609
Email: dhk@columbia.edu
URL: http://cred.columbia.edu

Dr Simon J. MASON
International Research Institute
for Climate Prediction (IRI)
Columbia University
61 Route 9W
P.O. Box 1000
Palisades, NEW YORK 10964-8000
Tel: +1 845 680 4514
Fax: +1 845 680 4865
Email: simon@iri.columbia.edu
URL: http://iri.columbia.edu/cgi-bin/staff?smason

WMO SECRETARIAT
7bis, Avenue de la Paix
CP 2300
CH-1211 GENEVA 2
Telefax:(41 22) 730 8042

Mrs L. MALONE
Scientific Officer
World Climate Applications and
CLIPS Division
World Climate Programme Department
Telephone: (41 22) 730 8220
E-mail: LMalone@wmo.int

Observers

Dr Shoji KUSUNOKI
Meteorological Research Institute
Climate Research Department
4th Laboratory (Global Warming Projection)
1-1 Nagamine, Tsukuba, Ibaraki
Japan
Tel: +81 29 853 8608
Fax: +81 29 855 2552
Email: skusunok@mri-jma.go.jp

Japan Meteorological Agency
Climate Prediction Division
1-3-4 Otemachi
Chiyoda-Ku
TOKYO, Japan 100-8122
Tel : +81 3 3211 8406
Fax : +81 3 3211 8406

Dr Tomoaki OSE
Email : t-ose@met.kishou.go.jp

Mr Fumio WATANABE
Email :
fumio.watanabe@met.kishou.go.jp

Mr I. YOSHIKAWA
Email : yoshikawa@met.kishou.go.jp

Mr S. YABU
Email : yabu@met.kishou.go.jp

Annex III

Meeting of the Expert Team on Verification
Commission for Climatology (CCl)
of the World Meteorological Organization (WMO)

Tokyo, 2 February 2005

Welcome Address
by
Mr. Koichi NAGASAKA
Director-General of the Japan Meteorological Agency

Good morning,
Distinguished participants,
Ladies and gentlemen,

It is my great pleasure and deep privilege to deliver the welcome address at the beginning of the Meeting of the Expert Team on Verification of the Commission for Climatology of the World Meteorological Organization (WMO). First of all, on behalf of the Japan Meteorological Agency and in my own capacity as the Director-General of the Agency, I would like to extend my warm and coordinal welcome to all of the distinguished participants who gathered here from various parts of the world in the meeting of the Expert Team on Verification of the Commission for Climatology.

In recent years, the world community has been seriously suffering from various natural disasters having extremely heavy impacts on our society, including those by severe weathers, with increasing frequency. It is also the case for Japan. Last year, Japan was hit by ten typhoons, which nearly doubled the previous record of six for the annual total of landfall typhoons. Floods and landslides caused by the typhoons and active Bai-u front claimed more than three hundred lives and seriously damaged the socio-economic activities across our country. Furthermore, heat wave in the summer brought record-breaking high temperatures to Eastern Japan with significant impact on the civil life in this area. Even this week, we are suffering from heavy snowfall in wide areas of our country.

These days, modern societies are becoming increasingly sensitive and more vulnerable to meteorological hazards and impacts, as they rapidly develop and diversify. Accordingly, not only disaster prevention authorities, but also the general public as well as a variety of industries demand meteorological information which are more detailed, informative, accurate and readily comprehensible.

It is one of the most important responsibilities of the National Meteorological and Hydrological Services (NMHSs) to contribute for the reduction of the disastrous damage, and of impact of weather and climate extreme and for the achievement of the sustainable development of the socio-economic activities in the world through providing appropriate information on weather and climate in a timely manner.

Ladies and gentlemen,

International cooperation in sharing scientific and technological knowledge on meteorology and climatology is indispensable for all the NMHSs to provide the public with qualified long-range forecasts. In this connection, JMA is playing a leading role and taking the strong initiatives in international frameworks of not only daily weather services but also long-range and climate forecasting services. Since October 2002, the Tokyo Climate Centre of JMA has been providing NMHSs with fundamental climate information including global long-range forecast and conducting training and technical cooperation for capacity building.

Unfortunately, accuracy and reliability of long-range predictions such as seasonal and inter-annual forecasts still lag behind the demands from the public as compared with those of daily weather forecasts. Intensive efforts are therefore continued to improve accuracy and reliability of the seasonal to inter-annual predictions by many National Meteorological Services and various research communities in the world.

In the meantime, we have to establish an appropriate scheme to evaluate the long-range forecasts to contribute to the development of useful and valuable seasonal and inter-annual prediction services. Among the nine terms of reference of your Expert Team on Verification, probabilistic information on the forecast is particularly relevant, from the technical point of view.

Establishment of objective and users' requirements oriented VERIFICATION methods for long-range forecasts to support for improvement of the forecast skills is fully expected through your intensive deliberation.

It is my anticipation that the individual distinguished experts from all over the world gathered in this room will widely review and intensively discuss what kind of verifications are required for making long-range forecasts more useful and effective for users by discharging your advanced knowledge, comprehensive experience and wide-ranged expertise. Also, I strongly hope that this meeting will contribute to more user-friendly and more useful long-range forecasts through your ideas and outcomes on the verification methods.

Our staff will spare no efforts to support the meeting and assist all of the participants in the coming three days. Finally, I hope, in spite of the strong cold surge dominating over our country this week, you will enjoy comfortable stay in Tokyo under the clear and bright sky associated with the cold surge.

Thank you very much for your attention.