TABLE OF CONTENTS

		Page No.
1.	RELEVANT RECOMMENDATIONS FOR THE DEVELOPMENT OF WGNE/GMPP ACTIVITIES	1
2.	PHYSICAL PARAMETERIZATIONS IN MODELS	1
	2.1 Cloud parameterizations	1
	2.2 Land-surface processes	2
	2.3 Atmospheric boundary layer	3
3.	STUDIES AND COMPARISONS OF ATMOSPHERIC MODEL SIMULATIONS	4
	3.1 General model intercomparisons	4
	3.2 Standard climate model diagnostics	6
	3.3 Developments in numerical approximations	6
	3.4 Model-derived estimates of ocean-atmosphere fluxes and precipitation	7
	3.5 Snow Models Intercomparison Project (SNOWMIP)	8
	3.6 Model stratospheric representation	8
	3.7 Regional climate modelling	9
	3.8 Other climate-related modelling initiatives	10
4.	DATA ASSIMILATION AND ANALYSIS	11
	4.1 Reanalysis projects	11
	4.2 Observing system and observation impact studies	12
	4.3 Co-ordinated Enhanced Observing Period	13
	4.4 Current Issues in Data Assimilation	14
5.	NUMERICAL WEATHER PREDICTION TOPICS	15
	5.1 Short- and medium-range weather prediction	15
	5.2 Ensemble prediction	19
	5.3 Recent developments at operational forecast centres	22
6.	OTHER WGNE ACTIVITIES AND FUTURE EVENTS	29
7.	CLOSURE OF SESSION	29

APPENDICES:

A. List of participants

B. WMO Policy Statement on the “Scientific basis for, and inherent limitations of weather and climate forecasting”.

C. The WGNE survey of verification methods for numerical prediction of weather elements and severe weather events- Philippe Bougeault

The eighteenth session of the CAS/JSC Working Group on Numerical Experimentation (WGNE), held jointly with the sixth session of the GEWEX Modelling and Prediction Panel (GMPP), was kindly hosted by Météo- France, Toulouse, France, from 18-22 November 2002. The session was opened at 0900 hours on 18 November by the Chairman of WGNE, Dr. K. Puri, and of GMPP, Dr. D. Randall. The list of participants in the (joint) session is given in the Appendix A.

Dr. Daniel Cariolle, Directeur du Centre National de Recherche Météorologique (CNRM), could not be present due to official responsibilities. Dr. Philippe Bougeault, Chef du Groupe de Météorologie de Moyenne Echelle and Head, Mesoscale Meteorology Division,Météo-France CNRM/GMME/D, welcomed all participants to the International Conference Center, Météo-France, and spoke of the importance of the agenda to be taken up at the session which should lead to valuable results for meteorological services.

On behalf of all participants, Dr. K. Puri expressed gratitude to Dr. Daniel Cariolle and Météo-France for hosting the joint session of WGNE and GMPP and the excellent arrangements made. He voiced his appreciation to Dr. P. Bougeault, ably assisted by Ms. Isabelle Varin and Sylviane Balland, for the efforts and time they had put into the organization of the session.

1. RELEVANT RECOMMENDATIONS FOR THE DEVELOPMENT OF WGNE/GMPP ACTIVITIES

At its fifty-fourth session, the WMO Executive Council approved the WMO Statement on the scientific basis for, and limitations of, weather and climate forecasting, which was largely prepared by WGNE, WWRP and WGSIP. The Council commended the CAS for its substantial efforts in developing the WMO Statement which presented the issues in a balanced manner and which would provide important guidance for NMHSs in their dealings with government officials, users, the media and the general public. WGNE expressed satisfaction at the completed report which is attached as Appendix B.

The Commission viewed collaboration between WGNE and the World Weather Research Programme (WWRP) as of considerable importance. One key area of such collaboration is THe Observing system Research and Predictability Experiment (THORPEX) being undertaken as a “Research and Development Programme” of WWRP in collaboration with WGNE. The themes proposed (see section 5.1) are of major interest to WGNE, and the studies of predictability and observing system issues being taken up will have benefits throughout the WCRP. The international coordination of THORPEX is under the auspices of the WMO, WWRP and WGNE. The THORPEX International Science Steering Committee (ISSC) defines the core research objectives with guidance from the THORPEX International Core Steering Committee (ICSC) whose members are selected by national permanent representatives to the WMO. WGNE reiterated its support for THORPEX as a collaborative WWRP/WGNE experiment. At the WGNE session, a joint WWRP/WGNE draft resolution concerning the current status and the next steps in the development of THORPEX was reviewed and finalised in consultation with the Chair of the WWRP, Dr R. Carbone. The committees agreed that the essential next step is the development and submission of the detailed THORPEX Science plan for review and consideration by WWRP and WGNE.

WGNE was asked to comment on the document,” Possible banner for WCRP- Predictability assessment of the Climate System”, produced by the JSC Task force. WGNE discussed at length this issue. A detailed submission containing WGNE’s views has been made to the JSC.

2. PHYSICAL PARAMETERIZATIONS IN MODELS – PROCESSES LINKED TO THE WATER CYCLE IN ATMOSPHERIC MODELS

The GEWEX modelling and prediction thrust, with which WGNE works in close association, is devoting efforts to the refinement of the representation of processes linked to the water cycle within atmospheric models, notably those of clouds and radiation, land surface processes and soil moisture, and the atmospheric boundary layer.

2.1 Cloud parameterizations

One of the main activities supporting refinement of model cloud parameterizations is the GEWEX Cloud System Study (GCSS) being conducted as a component of the GEWEX modelling and prediction thrust. The goal of GCSS is to improve the parameterization of cloud systems in atmospheric models through improved physical understanding of cloud system processes. The main tool of GCSS is the cloud-resolving model (CRM), which is a numerical model that resolves cloud-scale (and mesoscale) circulations in either two or three spatial dimensions. The large-eddy simulation (LES) model is closely related to the 3D CRM, but resolves the large turbulent eddies. The primary approach of GCSS is to use single-column models (SCMs), which contain the physics parameterizations of GCMs and NWP models, in conjunction with CRMs, LES models, and observations, to evaluate and improve cloud system parameterization.

Dr. S. Krueger, Chair of the GCSS Science Panel, recalled that GCSS is composed of five working groups, relating to boundary-layer cloud systems, cirrus cloud systems, extratropical layer cloud systems, precipitating convective cloud systems, and polar cloud systems. C. Bretherton and P. Brown began serving as WG chairs during 2002, respectively for the boundary-layer and the cirrus groups.

The GCSS workshop held at Kananaskis in May 2002 reflected the increasing interest of the GCM community in GCSS activities and the increasing interaction of GCSS with the radiation, microphysics, aerosol, and cloud-remote sensing communities. The following scientific advances are expected in the GCSS WGs during the next several years:

- rapid progress on the representation of sub-grid scale cloud overlap and inhomogeneity due to the combination of CRMs, cloud radar observations, and faster methods of calculating radiative fluxes for arbitrary cloud configurations;
- steady progress in the understanding and representation of cloud microphysical, formation, and dissipation processes due to integrated use of LES models, CRMs, SCMs, GCMs, and cloud-scale observations, plus insights from recent and upcoming field experiments; and
- use of super-parameterizations (i.e., CRMs used as parameterizations) in some GCMs will provide more physically realistic representations of cloud processes, to increase knowledge and understanding of interactions between cloud processes and large-scale processes (including cloud feedbacks), and to help improve conventional parameterizations.

EUROpean project on Cloud Systems in climate and NWP models(EUROCS)

Dr. J.L. Redelsperger outlined the scope of EUROCS which aims to improve the treatment of cloud systems in global and regional climate and NWP models. The major issues being focussed are: the representation of strato-cumulus over ocean, diurnal cycle of cumulus and precipitating deep convection over continents, and the lack of sensitivity of deep convection development on the moisture profile. These are issues which lead to major deficiencies in the predicted global and regional climates and have not been sufficiently or not at all addressed in the past. A transversal approach developed during the project concerns the diagnosis of cloud and radiation behaviour of cloud transition (from strato-cumulus to cumulus to cumulonimbus) through a cross-section over the Pacific in full 3D GCM simulations.

The strategy used in EUROCS is based on the use of the hierarchy of models and observations to integrate cloud studies across the full range of scales. Six climate models are closely associated to specific SCM and to CRM/LES. Results presented during the WGNE meeting showed significant improvements for the different selected cases. More information is available on the EUROCS Web site: http://www.cnrm.meteo.fr/gcss/EUROCS/EUROCS.html.

2.2 Land-surface processes

Dr. J. Polcher reported on the GEWEX Global Land-Atmosphere System Study (GLASS) project which is progressing through the various actions which were defined in the implementation plan. Under PILPS (Project for Intercomparison of Land Surface Parameterization Schemes), a set of simulations at the local and regional level was finalised over the Rhône basin (The Rhone-AGGregation project), a new local study including carbon fluxes was initiated over a forested land in the Netherlands (PILPS-C1), and a third off-line intercomparison of land surface models is starting for the first time in a semi-arid region (San Pedro catchment in the southwestern U.S.).

The Global Soil Wetness Project 2 (GSWP-2) will start in early 2003 and first results should be available by the end of the year. Its goals are to:

- Produce state-of-the-art global data sets of surface fluxes, of soil wetness and related hydrologic quantities.
- Develop and test large-scale validation, calibration, and assimilation techniques over land.
- Provide a large-scale validation and quality check of the ISLSCP data sets.
- Compare Land Surface Schemes and conduct sensitivity studies of specific parameterizations which should aid future model development.

A major product of GSWP-2 will be a multi-model land surface analysis for the ISLSCP II period.

In order to assess our knowledge of the role surface moisture and temperature states play in the evolution of weather and the generation of precipitation, a new study called GLACÉ (Global Land-Atmosphere Coupling Experiment) has been launched. It aims to study the role of land-surfaces in the variability of the climate system in a multi-model approach. The study is based on three ensembles of GCM experiments to be performed by all participating atmospheric models coupled to their land-surface schemes. The three ensembles will differ in the freedom that will be given to the land-surface to respond to the atmospheric forcing. It is hoped that the statistical analysis will reveal the contribution of the continental surfaces to the synoptic variability in the different models. The inter-model differences will allow an assessment to be made of the current knowledge of the strength of the coupling.

Another project of note is ELDAS, the European Union project for the assimilation of soil moisture, which in its first year has been mainly devoted to the construction of a demonstration database over Europe and the design of a data assimilation system that extracts information on soil moisture from a blend of observation types, including synops and satellite data.

GSWP and related activities at Météo-France

The Global Soil Wetness Project is a GEWEX initiative to study the feasibility of producing global soil moisture climatologies by driving state-of-the-art land surface models (LSM) with 6-hourly atmospheric reanalyses corrected for their monthly biases in precipitation and surface radiation. Dr. H. Douville presented the studies conducted using the ISBA LSM of Météo-France in GSWP1 that covers the 1987-88 period. ISBA has been coupled to the TRIP river routing model in order to simulate monthly river discharge over large basins. Although very simple, TRIP allowed the detection of some deficiencies in ISBA such as occasional lack of subgrid runoff or overestimated infiltration over frozen soils. A notable result was that all land surface models participating in GSWP underestimated the annual runoff over the Amazon basin. Various sensitivity tests have suggested that this robust bias could be due to deficiencies in the precipitation forcing and land surface parameters derived from the ISLSCP1 dataset, rather than to the models. Such experiments suggest that the experimental design used in GSWP (as well as in the GLDAS project where the precipitation forcing is even more questionable than in GSWP) does not guarantee the production of a realistic soil moisture climatology. Data assimilation is probably necessary for this purpose and it was shown that the assimilation of SYNOP observations of screen-level temperature and humidity has a strong potential but requires an interactive boundary layer. Therefore it cannot be used efficiently in GSWP. Nevertheless, the relevance of realistic soil moisture boundary conditions (derived from GSWP) for atmospheric seasonal simulations was demonstrated through a case study based on years 1987 and 1988. It was shown that the ARPEGE atmospheric model captures much better the interannual variability of the boreal summer climate if the total soil moisture simulated in ISBA is relaxed toward the GSWP climatology. Although the persistence of initial soil moisture anomalies remains an important issue, this result suggests that soil moisture should be initialized carefully in dynamical seasonal forecasts.

African Monsoon Multidisciplinary Analysis (AMMA)

The AMMA project, presented by Dr. J.L. Redelsperger, is an integrated multidisciplinary project that aims at addressing both fundamental scientific questions related to the understanding of the West African monsoon (WAM) variability and practical issues related to prediction and applications. The AMMA project is endorsed by CLIVAR and GEWEX. The project has three overarching aims: (i) To improve our understanding of the WAM and its influence on the physical, chemical and biological environment regionally and globally, (ii) To provide the underpinning science that relates climate variability to issues of health, water resources and food security for West African nations and defining relevant monitoring strategies and iii) To ensure that the multidisciplinary research is effectively integrated with prediction and decision making activity. To achieve these aims a multidisciplinary approach to the study of the WAM is required involving substantial international collaboration. AMMA will link observations, data analysis and modelling on a wide range of space and time scales. The project will address the following interacting science areas: Monsoon dynamics and scale interactions, continental water cycle, aerosols, atmospheric chemistry, food, water and health. AMMA is planned to be a multi-year project and will involve 3 observing periods: the LOP (Long term Observing Period) considering historical observations and supplementary long term observations to document and analyse interannual variability, the EOP (Enhanced Observing Period) designed to document over 2-3 years the annual cycle and the memory effects; the SOP (Special Observing Period) providing detailed observations of specific processes at various key stages of the rainy season during 2005.

2.3 Atmospheric boundary layer

The "GEWEX Atmospheric Boundary Layer Study" (GABLS) has the principal objective of improving the representation of the atmospheric boundary layer in regional and large-scale models, based on advancing the understanding of the relevant physical processes involved. GABLS will also provide a framework in which scientists working on boundary layer research issues at different scales can interact. The first focus of GABLS is on stable boundary layers (SBL) over land. Much of the warming predicted by climate models is during stable conditions over land (either in winter or at night), while at the same time the understanding and parameterization of the SBL is still very poor.

Dr. B. Holtslag summarized the progress so far in GABLS. A GABLS workshop on Stable Boundary Layers was held at the European Center for Medium-Range Weather Forecasting (ECMWF) in Reading, UK, on March 25-27, 2002 with a balanced participation of process modellers, observation specialists and GCM modellers.

Three task groups were defined on the following topics: the analysis of existing observations, in order to provide data sets to validate LES results and to help scope out the parameterization problem, large eddy simulations to help guide and evaluate proposed parameterizations, and GCM studies to provide feedback on updated parameterizations.

3. STUDIES AND COMPARISONS OF ATMOSPHERIC MODEL SIMULATIONS

3.1 General model intercomparisons

A key element in meeting the WGNE basic objective to identify errors in atmospheric models, their causes, and how they may be eliminated or reduced, was a series of model intercomparison exercises. These encompassed a number of fairly general wide-ranging intercomparisons as outlined in this section, as well as more specific efforts, e.g., evaluation of snow models as employed in atmospheric circulation models (see section 3.5), or assessment of stratospheric analyses and predictions (see section 3.6).

Atmospheric Model Intercomparison Project (AMIP)

The most important and far-reaching of the WGNE-sponsored intercomparisons is the Atmospheric Model Intercomparison Project, conducted by the Programme for Climate Model Diagnosis and Intercomparison (PCMDI) at the Lawrence Livermore National Laboratory, USA, with the support of the US Department of Energy. Dr. P. Gleckler reviewed the status and progress of the project, which, based on a community standard control experiment simulating the period 1979 – ‘near present’, is now reaching the end of its second phase (AMIP-II). Approximately twenty-five modelling groups have submitted simulations and much of the data from these runs are available for a wide range of diagnostic sub-projects. In addition to the standard runs, ensembles and runs at varying horizontal resolutions are being archived for specific research sub-projects. Regular updates of the overall status of AMIP, model integrations, and diagnostic subprojects are posted on the AMIP home page http://www-pcmdi.llnl.gov/amip. On the technical side, PCMDI now has a powerful open source software system which enables efficient management of the voluminous AMIP data sets. An automatic system is now in place which can organise simulations, perform extensive quality control, and make the data accessible (via ftp) to interested users. Most importantly, the facility is now able to rapidly provide a detailed diagnostic report on a model simulation.

Following the recommendation of the WGNE, an International AMIP Workshop was held in Toulouse from 12-15 November 2002. The WGNE-appointed AMIP panel served as the Scientific Organising Committee which was chaired by Dr. Peter Gleckler of PCMDI. A key decision made by the committee was to have a focus on innovative diagnostics and have a strong representation from the observational communities. The workshop program and abstracts are available at: http://www.cnrm.meteo.fr/amip2/. Model diagnostic sessions were broken into the General Circulation, Tropical Variability and Monsoons, Fluxes and Cloud-Radiative effects, the Hydrological Cycle, Land Surface Processes and Phenomena and Extra-Tropical Variability. Keynote speakers included: M. Miller (ERA40), B.J. Hoskins (Dynamical approaches), J.-F. Royer (West African Monsoon), T. Koike (Co-ordinated Enhanced Observing Period, CEOP), S. Krueger (GEWEX Cloud Systems Study), R. Koster (GLASS poor man’s LDAS), and J.-J. Morcrette (use of ARM data for model diagnosis). Several discussion forums were devoted to refining the experiment and prioritizing future activities.

Some key conclusions of the Conference included –

- Despite limitations, the idealised AMIP SST experiment is still a powerful diagnostic test,
- A ‘mean AMIP model’ generally outperforms any individual model and is a useful reference,
- Diagnostic Subproject analysis has become an increasingly useful exercise,
- There was an encouraging synergism with the GEWEX modelling projects,
- There was strong support by conference attendees to see AMIP continue in some form.

The conference proceedings will be published as a WCRP report.

The WGNE session immediately following the AMIP Conference discussed future directions for AMIP. The discussions included recommendations from an AMIP panel meeting held in Reading in February 2002 namely, comprehensive diagnostic reports should be made available to modellers soon after they submit a simulation to PCMDI; AMIP should be exploited as a diagnostic tool of the coupled system with WGNE and WGCM working towards integrating AMIP and CMIP; and process studies should become an increasing diagnostic focus. It was also noted that (i) there was an external review of PCMDI in March 2002, which provided encouraging support for the continuation of systematically evaluating AMIP runs; (ii) diagnosis of coupled models is now a higher priority project at PCMDI than AMIP, and (iii) PCMDI will soon have a new director which would have a significant bearing on the future directions for the project. As with previous WGNE discussions on this topic and in line with the Workshop conclusions, WGNE continues to strongly support the continuation of AMIP as an experimental protocol providing an independent evaluation of atmospheric models and facilitating increasingly advanced diagnostic research. The Group recommended that the AMIP panel should meet again to discuss and provide advise on future directions for the project.

"Transpose" AMIP

Dr. D. Williamson outlined the basic concept underlying a "transpose" AMIP as proposed by himself and Dr. M. Miller, and a similar exercise "CCPP-ARM Parameterization Testbed" (CAPT) being undertaken by PCMDI and NCAR. Operational Numerical Weather Prediction has proven to be an excellent platform for examining parameterization methods as it allows direct comparison of the parameterized variables (e.g. clouds, precipitation) with observations early in the forecast while the modeled state is still near that of the atmosphere, but after initial transient computational modes are damped. Forecast centers report that such an approach is very useful in developing and evaluating parameterizations. Climate modeling groups not associated with an NWP centre generally have not been able to take advantage of such an approach because of the large amount of work involved in developing data ingest and assimilation systems. The question is how to obtain the benefits conferred by application of a model operationally in forecasting and assimilation for developing the parameterizations in climate models. The basic idea of a "transpose" AMIP and the companion project CAPT is to apply climate models to forecasts and examine how well the models predict the detailed evolution of the atmosphere at the spatial scales resolved by these models. Comparison with state variables from analyses and reanalyses and with estimates of parameterized variables from field campaigns should yield insight into the errors in parameterizations and lead to improved formulations.

The critical aspect is the initialization of the model for the forecasts. The basic approach is to map the climate scales as represented in analyses onto the climate model grid, eliminating the unresolved scales. The mapping of atmospheric state variables is reasonably straightforward as long as changes in orography and vertical coordinate system are accounted for. The mapping of parameterized atmospheric variables which have a time history (e.g. cloud water) is less obvious, but might be possible by considering the details of the parameterizations in both the climate model and analysis model. However, these variables are often related to fast process so their initialization might be less critical. Land model variables are more problematic because it is difficult to map the discrete/discontinuous land variables between different grids, there may be different dominant land types in the two systems, and there is no uniform definition of land model state variables. One approach currently being tried to obtain appropriate initial land values is to spin-up the land, and possibly atmospheric parameterized variables, over a period of time by having them interact with the atmosphere model constrained to follow the analyses in time by either periodically (e.g. 6-hourly) updating the atmospheric state or by adding a term to the model to force the state to follow the analyses to some degree (nudging). Both approaches may be more successful if poorly predicted atmospheric variables which drive the land, such as precipitation, are replaced by observed estimates as they are exchanged. Alternative approaches will also be considered. These include mapping reanalysis soil moisture profile to climate model by maintaining equivalent soil moisture availabilities, off-line land initialization (as in GSWP) driven with global observations, and inversion of observed surface fluxes.

WGNE is duly developing a project on these lines. Although a number of questions need to be resolved, the work to date is promising. Appropriate contacts will be taken with potential participants in discussing how to proceed. Advantage will also be taken of the experience in the Global Land-Atmosphere System Study (GLASS) where the planning of global scale interactive integrations has faced similar difficulties in the initialization of land surface and soil variables.

International Climate of the Twentieth Century Project (C20C)

The objective of the International Climate of the Twentieth Century Project, developed under the leadership of the Center for Ocean-Land Atmosphere Studies (COLA) and the UK Met Office Hadley Centre for Climate Prediction and Research, is to assess the extent to which climate variations over the past 130 years can be simulated by atmospheric general circulation models given the observed sea surface temperature fields and sea-ice distributions and other relevant forcings such as land-surface conditions, greenhouse gas concentrations and aerosol loadings. The initial experimentation being undertaken has involved carrying out "classic" C20C/extended AMIP-type runs using the observed sea surface temperature and sea ice as the lower boundary conditions (the HadISST 1.1 analyses provided by the Hadley Centre) for the period 1949-1997, with a minimum ensemble size of four members.

A workshop was convened in Calverton, MD, USA in January 2002 jointly by the Hadley Centre and COLA to review the results that had so far been obtained from the C20C model integrations and to plan a more highly structured C20C project. At the workshop the results from ensembles of runs forced with HadISST from the recent informal phase of C20C were summarised. Besides a number of diagnostic methods and new results on simulating 20th century climate, a presentation was made on the question of how limited AGCMs may be in simulating the variance of climate adequately. A specially designed experiment was created whereby the Hadley Centre HadAM3 AGCM was forced in ensemble mode with daily SSTs from part of a very long control run of the CGCM HadCM3. The initial conclusion is that the variance of those quantities looked on seasonal to decadal time scales are not significantly less in the AGCM than in the CGCM. Small differences that did occur were, however, consistent with the notion of excessive thermal damping in AGCM simulations. This supports the general validity of the AGCM approach for many types of climate predictability and trend studies. However an unresolved issue is whether some specific modes are missing in the AGCM that are present in the CGCM due to the lack of coupling. This work is being extended to include another AGCM/CGCM pair. The workshop decided that more emphasis should be placed on including forcings in addition to SST. Because of uncertainties in some forcings, and their tendencies to partially cancel, it was agreed to use (i) data from the Hadley Centre on changes in carbon dioxide since 1871, (ii) volcanic stratospheric forcing from 1950 only, and (iii) changes in tropospheric and stratospheric ozone. Participants will carry out a set of six integrations for 1871-2002 and a further set of 10 from 1950-2002. The HadISST will soon be updated in near real-time to make this possible. Participants will carry out a further 100-year control run with the 1961-1990 climatology of HadISST in order to study the role of naturally occurring modes.

Given that AMIP and C20C have a number of common features, WGNE expressed the view that both projects would gain by closer collaboration. C20C could, for example, follow AMIP in establishing a tighter experimental protocol and adapt some AMIP procedures, while AMIP should consider using the HadISST for any future phases.

3.2 Standard climate model diagnostics

Dr. D. Williamson recalled that WGNE standard diagnostics of mean climate have now been in use for a number of years and, in particular, were the basis for the "quick-look" diagnostics for AMIP simulations computed by PCMDI (see section 3.1). (The list of these standard diagnostics is available at http://www.pcmdi.llnl.gov/amip/OUTPUT/WGNEDIAGS/wgnediags.html.)

The standard diagnostics of mean climate included traditional variance and eddy statistics, but additional diagnostics of large-scale variability are also needed to characterize models. Over the past three years WGNE members have developed a list of standard diagnostics of variability focusing on the troposphere. These diagnostics have been demonstrated to be useful by individual developers and include measures of intraseasonal variability, Madden-Julian Oscillation (MJO), El Nino - Southern Oscillation (ENSO), blocking, seasonal cycle, diurnal cycle, atmospheric angular momentum, and modes of variability. They also include wavenumber-frequency plots, and histograms of precipitation. Examples of these diagnostics calculated from simulations with the NCAR Community Atmosphere Model (CAM2) can be seen at: http://www.cgd.ucar.edu/cms/mstevens/variability/AMWG/variab.html

They will also be included in the PCMDI "quick-look" diagnostics mentioned above. Code for the diagnostics will be available from both centres in the future. It was decided to ask SPARC to provide the same list for the stratosphere.

3.3 Developments in numerical approximations

Dr. D. Williamson led the discussion of this item. The range of approaches being followed in numerical approximations for integrating partial differential equations on a sphere, and the types of grids being tried, were well illustrated by the scope of presentations at the 2001 Workshop on the Solution of Partial Differential Equations on the Sphere in Montreal, Canada, May 2001, the International Workshop on the next generation Climate Model in Tokushima, Japan, March 2002, the Second Hybrid-isentropic Modeling Workshop in Louisville, USA, April 2002, and the 2002 Workshop on the Solution of Partial Differential Equations on the Sphere in Toronto, Canada, August 2002. Examples included, for the shallow water equations, techniques for using icosahedral, cubed sphere, and spherical grids. Likewise for baroclinic systems to which much more attention was now being given, methods using icosahedral, cubed sphere, spherical grids with variable resolution, and adaptive meshes were described. In the vertical, although an example of the application of finite elements was presented, traditional "sigma" co-ordinates are still very much in use. Several new vertical approaches are being developed including the use of cubic spline in the vertical advection with the semi-Lagrangian scheme coupled with cubic finite-element in the vertical at ECMWF, and spectral element vertical and horizontal discretization coupled with semi-lagrangian transport at the Naval Research Laboratory. Additional studies in this area (e.g., to take advantage of isentropic co-ordinates) are now definitely needed.

Specific consideration is also being given to the development of new methods for application in climate models, and for simulation of atmospheric transport (e.g., of aerosols, trace chemicals) where local conservation and preservation of the shape of distributions are essential. Energy conservation in climate models is of particular importance. In practice, conservation of better than 0.1 wm^{-2} is needed, whereas schemes with non-linear intrinsic diffusion (e.g., Lin-Rood, monotonic semi-Lagrangian) can lose energy at a rate of 1.5 wm^{-2}, as can explicit diffusion schemes. This loss should be converted to heat, but this might not be the correct approach. This is still a basic uncertainty in model formulation that must be kept in mind. One possible approach being pursued is to move away from spectral to local grid point based methods with local conservation and shape preservation without polar filters.

The numerical representation of orography and transport modelling remain particular issues which WGNE intends to follow. Another important component of activities in this area is the development of tests of the various numerical schemes/grids in a baroclinic system before introduction into complete models where complex feedbacks can obscure effects of new schemes. Two new related tests were presented at the 2002 Workshop on the Solution of Partial Differential Equations on the Sphere which were based on the growth of baroclinicly unstable modes. These were developed by L.M. Polvani (Princeton University) and R.K. Scott (Columbia University) and by C. Jablonowski (University of Michigan). Due to the nonlinear interactions of the growing modes the true solution is not known and reference solutions were computed with very high resolution dynamical cores. Details will be published in the future.

Aqua-Planet Experiments

In addition to tests of dynamical cores in isolation, the interactions of physics parameterizations with each other and with the dynamics need to be examined. Stripped down versions of atmospheric models with very simplified surface conditions, in particular "aqua-planet" experiments with a basic sea surface temperature distribution, offer a useful vehicle in this regard, with considerable potential to understand the performance and effects of different dynamical cores and different representations of physical processes. For example, at NCAR, aqua-planet simulations with Eulerian and semi-Lagrangian dynamical cores coupled to the CCM3 parameterization suite produced very different zonal average precipitation patterns. Analysis showed that the contrasting structures were caused primarily by the different timestep in each core and the effect on the parameterizations rather than by different truncation errors introduced by the dynamical cores themselves. When the cores were configured to use the same time step, and same three time-level formulation and spectral truncation, similar precipitation fields were produced.

WGNE has recognized that aqua-planet experiments could have wide application in testing basic model numerics and parameterizations in the way described above and has duly endorsed the proposal for an "aqua-planet intercomparison project". This would be led by the University of Reading together with NCAR and PCMDI. The objective would not just be to assess current model behaviour and to identify differences, but to establish a framework to pursue and undertake research into the differences. An experimental design and data to be collected has been developed and a list of diagnostics to be computed and compared was being considered. Details of experimental design are available at: http://www.met.reading.ac.uk/~mike/APE/ape_home.html

3.4 Model-derived estimates of ocean-atmosphere fluxes and precipitation

Evaluation and intercomparison of global surface flux products (over ocean and land) from the operational analyses of a number of the main NWP centres (the "SURFA" project) remains a high priority for WGNE. As well as the increasing concern in NWP centres with improving the treatment of surface fluxes, this activity responded to the request of the joint JSC/SCOR Working Group on Air-Sea Fluxes and the GCOS/GOOS/WCRP Ocean Observations Panel for Climate for a WGNE initiative to collect and intercompare flux products inferred from operational analyses. Moreover, the intercomparison of land-surface fluxes is of importance in the context of GLASS.

The atmospheric and coupled modelling communities and oceanographers have very strong interest in advancing SURFA, which could provide a good opportunity for real progress in estimating and determining surface fluxes. Some NWP fluxes are already being accumulated at PCMDI. Unfortunately, a committed funding source has yet to be identified for SURFA. Given the importance of this effort for a variety of research communities, it is hoped that this issue can be resolved soon.

Air–sea fluxes are directly important for a number of WCRP projects. Therefore, a background paper on ‘WCRP and Fluxes’ has been prepared by JPS/WCRP and WGNE was invited to comment on it. WGNE was also requested to consider the need to setup a ‘WCRP Coordinating Committee on Air-Sea fluxes’, given the very wide and varied requirements for air-sea fluxes within WCRP, and closely related programmes (e.g. GODAE, GCOS). WGNE supported the idea and suggested that the proposed committee should have a nominee from WGNE whose contributions to the new group will be in validation of surface fluxes and through AMIP subprojects.

3.5 Snow Models Intercomparison Project (SNOWMIP)

Dr. E. Martin reported on the progress of SNOWMIP project undertaken by Météo-France (Centre National de Recherches Météorologiques, Centre d'Etudes de la Neige, CNRM/CEN) under the auspices of WGNE and the International Snow and Ice Commission (ICSI) of the International Association of Hydrological Sciences. Liaison is also maintained with the GLASS. The objective of this project is to compare snow models of various complexity at four sites belonging to various climatic regions. A total of 24 models from 18 teams are involved. The models vary from simple models used for hydrology to sophisticated ones for snow physics research. The data for the runs were released in November 2000. After a workshop in July 2001 some teams were allowed to re-submit their results and the analysis began in January 2002. Some models show a good ability to correctly simulate the snow pack features for all of the sites, whereas other models are more adapted to particular conditions. The high alpine site is the best simulated site, because the accumulation and melting periods are distinct. The current analysis shows that when looking at a specific parameterization (e.g. albedo, water retention…) the results are highly variable and some show discrepancies between observations and models. For instance an albedo parameterization based on age only give bad results for the onset of melting at some sites. In 2003, the project intends to submit several papers and begin intercomparisons of detailed snow models. More information is available at http://www.cnrm.meteo.fr/snowmip/.

3.6 Model stratospheric representation

In the past two or three years, there has been growing interest in the representation of, and data assimilation and prediction in, the stratosphere and several major global operational centres have significantly increased the vertical extent and resolution of their models in the stratosphere and into the mesosphere; the associated data assimilation components have also been enhanced. WGNE is thus undertaking a new intercomparison of stratospheric analyses initially, followed subsequently by an intercomparison of model predictive skill in the stratosphere. This work, which is being led on behalf of WGNE by Dr. G. Roff (BMRC), closely complements that carried out in SPARC "GRIPS".

Data from five NWP models (BoM, ECMWF, NCEP, NOGAPS and Met Office) have been received for the northern hemisphere component of this study. The target period for the intercomparison was January - February 2000 which was an active period for the northern hemisphere polar vortex. The analyses were found to be relatively similar though there were distinct differences in the polar night jet magnitude, extent and location as well as the size and shape of the polar vortex low temperature regions between the models. All the available model forecasts were found to provide reasonable forecasts but were also found to have difficulty with certain days associated with large changes in the polar vortex. These days were generally linked to the rapid elongation of the polar vortex. Some models were found to cope with these days better than others. This study has now been extended to the southern hemisphere and similar datasets will be examined for the polar vortex splitting event in September - October 2002.

SPARC Data Assimilation Project(SPARC -DA).

Dr. A. Lorenc reported on the first Stratospheric Processes And their Role in Climate (SPARC) Data Assimilation (DA) workshop held on June 10-12, 2002 in Maryland, USA. Participants represented 16 groups from Europe, USA, Canada, and Japan. Topics presented included stratospheric data assimilation and data intercomparisons, constituent satellite data, PV reconstruction, ozone and chemical assimilation. Other outcomes of the workshop include formation of a Working Group on SPARC DA and planning of yearly workshops and sessions at scientific meetings.

GCM Reality Intercomparison Project for SPARC (GRIPS).

SPARC has been undertaking for a number of years an intercomparison of model stratospheric simulations. The first GRIPS results were published in 2000 (Bulletin of American Meteorological Society and Journal of Geophysical Research). The highlights of 2002 included several events. At a workshop in Tsukuba, Japan, the participants agreed on a further distribution of tasks within GRIPS. There were 12 participants in the project at the time of the presentation. A project web site was established at: http://userpages.umbc.edu/~pawson/html_flies/grips.htm.

Among the tasks of level 1, which aim to answer the question "How well do we simulate the present-day climate system?", studies of sudden stratosphere warming events, atmospheric tides and travelling waves, computation of spatial wave number spectra are now close to completion. The work has focussed on preparation of model documentation, generation of basic climatological data, troposphere-stratosphere connection and exchange, polar vortices, and Southern Hemisphere variability.

The level 2 tasks are aimed at development of parameterizations and studying model sensitivity to them. Simulation experiments on imposed mesospheric forcing have been successfully completed. Progress has been reported in experiments on the evaluation of gravity-wave drag schemes. Evaluation of corresponding radiation schemes and retrieval of gravity-wave drag data in diagnostic mode need attention.

The level 3 tasks are designed to study mechanisms by which various forcing factors control the atmospheric circulation and how they are represented in models. The experiments on intercomparison of model response to the Mount Pinatubo eruption conditions and on response to ozone trends are proceeding satisfactorily. At the same time assessment of ozone related sensitivity faces uncertainties in perturbation levels and even the base state. Therefore one of the GRIPS Level 3 tasks will be to resolve existing uncertainties in time for the next WMO-UNEP and IPCC assessment.

The year 2003 is seen as the right time for completion of GRIPS level 1 and level 2 tasks. It is important to support further studies of GRIPS level 3 tasks, which would require forcing simulation by complex models and would focus on climate - atmospheric chemistry interactions. The experience of model intercomparisons in the WCRP is very significant, particularly under the auspices of the CAS-WCRP Working Group on Numerical Experimentation (WGNE) and the Working Group on Coupled Modelling (WGCM). Relationships of SPARC and GRIPS with these groups need to be strengthened.

Some new results are related to experimentation on the temperature and ozone concentration response to solar forcing. Disparities between model responses were very significant. The effort was aimed at finding regular patterns of the response and suggesting the dominant mechanisms for it. The experimentation highlighted a need for better observational data to compare temperature and ozone variations and additional evaluation of two mechanisms likely to be important in linking ozone and temperature changes to chemical interactions in the lower stratosphere and dynamical response in the tropical and subtropical stratopause.

3.7 Regional climate modelling

The Chairman of the WGNE/WGCM RCM panel, Prof. R. Laprise, reported on the second meeting of the PRUDENCE consortium that took place on 2-4 October 2002 in conjunction with the Second ICTP Conference on ‘Detection and Modelling of Regional Climate Change’, held at the Abdus Salam International Centre for Theoretical Physics (ICTP) in Trieste, Italy. The PRUDENCE activities that relate directly to WGNE and WGCM include the coordinated use of several climate models to assess, in a controlled manner, a number of numerical modelling uncertainties associated with climate-change projections. These include the use of several low resolution coupled GCMs (CGCM), atmosphere only GCMs (AGCM) and nested RCMs. AGCMs are usually run at medium resolutions, as time slices of high resolution uniform resolution models, or as variable resolution AGCMs. These models are driven with sea states based on recent climate analyses (or alternatively GCM control experiments) to which are added the climate change from CGCM simulations.

The higher resolution used in the AGCM experiments can lead to reduced systematic biases (if coupled with relevant adjustments to the model physics) and thus make these models more suitable for driving RCMs.

Experimentation continues at the University of Quebec at Montreal (UQAM) following the so-called ‘Big Brother Experiment’ (BBE) perfect model protocol to assess the ability of nested regional climate models to reproduce with fidelity fine scale features. Earlier work using BBE focussed on the winter season over an eastern North American region where surface forcing is not dominant (Denis et al., 2002 and 2003). Further experiments have been carried out over a western North American region where there is a strong forcing exerted by orography (Antic et al., 2003), and for the summer season when surface processes exert a significant influence (Dimitrijevic et al., 2003). The overall conclusions of these perfect model experiments are as follows. One-way nesting RCMs can simulate quite accurately climate in terms of both large and fine scale components of stationary and transient eddies, when driven by large scale information in midlatitude winter. The results are improved by the presence of strong surface orographic forcing. The RCMs’ ability to reproduce accurately fine scale features is substantially reduced in summer, due to less effective large scale control by lateral boundary nesting. Additional findings of these studies concern the acceptable jump in spatial resolution between the driving and nested models and the acceptable time interval for providing lateral boundary conditions. For a 45-km grid RCM, it appears that a maximum jump of 6 (or possibly 12) is acceptable, which corresponds to an equivalent GCM spectral resolution of T60 (or possibly T30). The maximum acceptable update interval of the lateral boundary conditions for the nesting of a 45-km grid RCM appears to be around 6 hours. It is noteworthy that the maximum acceptable values of resolution jump and boundary update interval are mutually dependent.

In the ensuing discussions some further concerns on RCMs were expressed by WGNE members such as possible problems at boundaries due to the response of the land surface scheme and the ability to simulate the variability of extreme events. It was agreed that relevant WGNE members would provide Prof. Laprise with a write up of their concerns that would then be considered by the RCM panel. Following satisfactory resolution of these concerns, the RCM report will be finalised and submitted for publication in a general journal such as the Bulletin of the American Meteorological Society.

A number of options for the proposed WGNE/WGCM sponsored RCM Workshop have been considered, including holding a special session at an already scheduled RCM-related meeting. The favoured alternative currently being considered is for a joint WGNE/WGCM/IPCC RCM Workshop in early 2004. This is expected to be the recommendation of the RCM panel Chair to WGNE.

3.8 Other climate-related modelling initiatives

WGNE noted with interest reports of developments in climate modelling activities in Japan and Europe.

Japan

Dr.K. Saito reported on the Earth Simulator (ES) program in Japan. This program has been undertaken by the Earth Simulator Center (ESC) towards a comprehensive understanding of global changes, with the collaboration of the Japan Marine Science Technology Center (JAMSTEC) and the Ministry of Education, Science & Technology of Japan. Development of ES started in 1997, and its construction was completed at the end of February 2002. ES is a distributed memory, massively parallel vector computer, consists of 640 processing nodes. The peak performance of each processing node is 64 Gflops, and total peak performance is 40 Tflops. 35.8 Tflops (87% of the peak performance) was achieved in the Linpack benchmark.

ESC has been developing models for Earth Simulator. One of the models is an atmospheric GCM based on the CCSR (Center for Climate System Research, Univ. Tokyo)/NIES (Japanese National Institute for Environmental Studies) AGCM which has been rewritten comprehensively with FORTRAN 90. Balancing among the micro-tasking, MPI and vector operation has been considered to optimize the model for the Earth Simulator. As a result, sustained performance 26.58 Tflops was achieved for the T1279L96 (10 km) global model, which corresponds to 64.9 % of peak performance. Using this model, ESC will perform meso-scale resolving simulations to study the interactions between large-scale circulations and meso-scale phenomena.

The Frontier Research System for Global Changes (FRSGC) has been developing a global nonhydrostatic model using the icosahedra hexagonal grid. The Held & Suarez dynamical core experiment was performed, and after 1200 days time integration, position and intensity of jet stream agreed with those obtained by the DWD Global model (GME) and the ECMWF spectral model (IFS). Computational performance on ES shows good scalability, where the elapse time increases proportionally to the expected time corresponding to the horizontal resolution.

JMA has been participating in the Research Revolution Project since April 2002. This is a research project by the Ministry of Education, Science & Technology to use the ES. In this project, the Meteorological Research Institute of JMA performs global warming experiments with a T_L959 (20 km) AGCM for the IPCC report, and severe weather simulations with 2-5 km nonhydrostatic model, while the Numerical Prediction Division of JMA develops next generation NWP models.

Europe

Dr. S. Valcke reported on the European Network for Earth System Modelling which is set up as a think tank to organize, plan and seek funding for efficient distributed Earth System Modelling in Europe. ENES has the long-term goal of achieving a distributed European facility for Earth System Modelling. Its first realization is PRISM, the Program for Integrated Earth System Modelling. PRISM is an infrastructure project (Dec. 2001- Nov 2004) and has 22 partners comprising of leading climate research institutes and computer vendors. The goals of PRISM are:

- to provide software infrastructure to
  - easily assemble earth system coupled models based on existing state-of-art European components models
  - launch/monitor complex/ensembles earth system simulations
  - access, analyse and share results across wide community
- to define and promote technical and scientific standards for Earth System modelling.

More information is available at http://www.prism.enes.org/

4. DATA ASSIMILATION AND ANALYSIS

4.1 Reanalysis projects

ECMWF

Dr. M. Miller reported that the ambitious and comprehensive 40-year reanalysis project at ECMWF (ERA-40, August 1957 to December 2001), with support from the European Union, is progressing well. The assembly of a merged data set of conventional observations has been carried out in collaboration with NCEP and NCAR. A surprisingly large amount of extra data is available compared to the earlier 15-year reanalysis (ERA-15), with, in particular, a significant increase in the number of radiosonde and pilot wind soundings from the NCEP database. EUMETSAT is also reprocessing wind products from METEOSAT-2 from 1983-1988. The collection of observations and their archive is itself a valuable resource that will be shared with NCEP and JMA for future reanalysis. Many problems with observations have been resolved although others remain, especially biases in radiance data.

The data assimilation is based on the elements of the system that was operational from June 2001 to January 2002 and includes 3D-Var analysis, T_L159 L60 resolution model, direct assimilation of raw radiances, analysis of ozone, a coupled ocean-wave model and enhanced set of post-processed products. Additional products include cloud statistics from TOVS radiance processing, fields from the physical parametrizations to support chemical-transport modelling, comprehensive outputs for selected grid points and catchment basins, vertically-integrated fluxes and data on isentropic and constant-PV surfaces. The reanalysis itself is being undertaken in three streams covering the periods 1987-2001 when TOVS, SSM/I, ERS, ATOVS and CMW data were available, 1972-1988 with VTPR, TOVS and CMW data, and 1957-72 (the pre-satellite era).

Tests of the assimilation of SBUV and TOMS ozone data have proceeded in parallel, and have given satisfactory results. SBUV and TOMS assimilation was thus added to the production system from January 1991 onwards. Ozone analyses for 1989 and 1990 will be produced off-line. In this connection, the ERA-40 experience has been invaluable in the development of operational assimilation of ozone at ECMWF.

A number of assessments of the ERA-40 analyses for the late 1980s and early 1990s have been made by the partners in the project (from ECMWF Member States and NCAR). In almost all respects, the quality of the ERA-40 analyses appeared to be superior to that of the ERA-15 analyses. The validation studies have identified some deficiencies especially with the tropical hydrology, and mixed results for the pre-1979 period.

ERA-40 data will be (i) available to all Member States via direct access (free), (ii) condensed onto CD with limited levels and parameters, (iii) available nationally via specific data centres such as NCAR, MPI, BADC, IPSL, (iv) available to non Member States via ECMWF (with handling charges), and (v) available in small subsets on the ERA-40 website (public).

Comprehensive information on ERA-40, including the current status of production and archiving and monitoring plots can be consulted via http://www.ecmwf.int/research/era.

NCEP

Dr. S. Lord reviewed the status of reanalysis activities at NCEP. The original NCEP/NCAR reanalysis from 1948 is continuing to be carried forward to the present in a quasi-operational manner (two days after data time). The reanalyses distributed through NCAR, CDC and NCDC are readily available either electronically or on CD-ROM. A joint NCEP/DOE reanalysis (NCEP-2) for the period 1979-1999 has also been produced (available electronically). This was based on an updated forecast model and data assimilation with corrections for many of the problems seen in the original NCEP/NCAR reanalysis and also improved diagnostic outputs.

The current initiative is the preparation of a regional reanalysis over the USA for the period 1979-2004 to be continued later in near-real time. This should provide a long-term consistent data set for the North American domain, superior to the global reanalysis in both resolution and accuracy. The regional reanalysis will be based on the Eta model and the Eta data assimilation system with the global reanalysis used as boundary conditions. Model resolution will be 80km with 38 layers in the pilot stage and 32km with 45 layers in the production stage. Important features will be direct assimilation of radiances and assimilation of precipitation (over the USA), as well as recent Eta model developments (refined convective and land-surface parameterizations). Free forecasts will be carried out to 72 hours every 2.5 days. A range of data (including all those used in the global reanalysis, various precipitation data sets, TOVS-1B radiances for certain periods, profiler measurements, and lake surface data) has been assembled and a large number of pilot runs carried out. Considerable improvements are apparent in the precipitation patterns which look very similar to the observed precipitation patterns in summer, especially in runs where precipitation was assimilated. The fit to the upper air temperatures and vector winds (as observed by radio-sondes) and surface temperatures are also notably better than that of the global reanalysis.

The production of the regional reanalysis is now in progress and two streams will be run when the Class VIII machine becomes available. It is planned to complete most of the production by 31 August 2003, the last date that Class VIII machine will be available. A Users' Workshop is planned for 2003.

Japan Meteorological Agency (JMA)

Dr. K. Saito reported on the reanalysis activities in Japan. The Japanese Reanalysis Project, JRA-25 is a five-year joint project of JMA, which is providing the operational data assimilation and forecast system, and the Central Research Institute of Electric Power Industry (CRIEPI), a private foundation providing computer resources. The objective of the project is to provide a comprehensive data set for the period 1979-2004 which will form the basis for a dynamical seasonal prediction project and global warming study, for advanced operational climate monitoring services at JMA, and for various activities in climate system studies. A 3D-Var system (operational since 2001) with a model resolution of T106 and 40 levels in the vertical will be employed. As well as data archived at JMA from 1975 to present, the NCEP/NCAR data used in the NCEP reanalysis and the merged ECMWF/NCEP data sets in ERA40, a range of satellite observations (including reprocessed GMS cloud motion wind data), and 'bogus' wind data surrounding tropical cyclones will be assimilated. The project is expected to be completed by 2005, with the products available to scientific groups contributing to the evaluation of the reanalysis and who provide feedback on improvements that could be made. Some recent developments include provision of TC bogus data by PCMDI/LLNL and two-year sample data of ERA-40 by ECMWF. The first announcement of invitation for evaluation group members was made in October 2002 and the second meeting of the JRA-25 Advisory Committee is planned for February 2003.

In the discussions following the presentations WGNE members reiterated that the JSC needs to seriously consider making reanalysis an ongoing effort, given the importance and strong support for the project. The current situation is unsatisfactory and wasteful because expertise built up for a reanalysis is lost when a phase is completed and then has to be reassembled with a new phase. A further advantage of an ongoing exercise is that it would facilitate research that is relevant to WCRP projects.

4.2 Observing system and observation impact studies

Dr. A. Lorenc reported on the role of various groups in OSEs, OSSEs and specifying observation requirements. Three types of study inform decisions about observation requirements:

- Experiments designed to answer specific design decisions about a network or satellite instrument. Ideally these should be done in close liaison with the agency designing the observing system, so the correct aspects are studied, and the results acted upon.
- Experiments with existing systems and observations, to test the data assimilation system, or validate a new observing system before using it.
- Longer-term, and perforce more idealized, studies of the impact of future observations on future NWP systems.

Not all results get published, and no centre can afford to do all the studies it would like, so there is a need for meetings and workshops to exchange results. Questions about observation requirements usually involve future systems not available for experiments. Moreover it is not possible to get significant results concerning rare extreme events within the resources available for an experiment. So judgement is needed in combining and extrapolating relevant results in order to give an expert opinion. This role is undertaken by groups such as the CBS Expert Team on Observational Data Requirements and Redesign of the Global Observing System.

Dr. J. Pailleux reported on progress in OSE activities. For the optimisation of the Global Observing System (GOS), a set of Observing System Experiments has been defined by the ad hoc Expert Team of the Commission for Basic Systems. Examples of the experiments include impact of hourly surface data (ECMWF), impact of denial of radiosonde data globally above the tropopause (CMC), information content of the Siberian radiosonde network and its changes during last decades (University of Petersburg, NCEP), impact of tropical radiosonde data (UK Met Office, Meteo-France), impact of three LEO AMSU-like sounders (UK Met Office, NCEP, ECMWF). Most of these experiments have now been carried out, at least by one NWP centre. A small sample of these results has been presented to WGNE. The status of this activity can be examined in more detail in the recent meeting report of the CBS Expert Team on Observation Data Requirements and Redesign of the Global Observing System (ET-ODRRGOS; Oxford; July 2002). This report can be found on the WMO web: http://www.wmo.ch/ (search by WMO programmes until you find WWW/GOS).Observing System Experiments are also carried out for a wide range of activities. Some examples related to new or emerging observing systems were shown to WGNE. It is worth collecting regularly as much as possible of the results for an overall assessment of the GOS. This was done in the past through two workshops (1997, 2000). A third workshop is planned for early 2004.

4.3 Co-ordinated Enhanced Observing Period (CEOP)

In the joint discussions with the GEWEX Modelling and Prediction Panel, the status of the planning and steps towards implementation of the GEWEX Co-ordinated Enhanced Observing Period (CEOP) were reviewed. CEOP has requested the WGNE community to provide comprehensive gridded output from global data assimilation systems. This requested output includes not only standard meteorological output but also output allowing study and analysis of water and energy processes in the atmosphere and land surface. In particular, detailed Model Output Location Time Series (MOLTS) have been requested at 41 international reference sites, where there are extensive in situ measurements and where extensive satellite products are being developed. This small data set will be complemented by more comprehensive 3 dimensional globally gridded data. Minimum output will include analysis variables, every 6 hours, as well as variables every 3 hours from a 6 hour forecast made every 6 hours as part of the analysis cycle. Every day at 1200 UTC, a corresponding 36 hour forecast is also requested, since this will provide some measure of how the models are adjusting (spinning up) to the initial state. This data will be archived initially by the individual meteorological centers and then later sent to MPI, which will develop a model output archive. NWP Centers are only being asked for comprehensive analysis and forecast output for the period Jul. 1, 2001-Dec. 31, 2004.

Most of the centres represented on WGNE were in principle ready to assist but raised questions concerning the complexity and long-term nature of the request, how the model data would be used in practice, and how CEOP would be useful for NWP centres. The need was expressed for a clearer exposition of the scientific strategy that would be followed by CEOP to exploit the in situ, remotely-sensed, and model output data to meet CEOP objectives. The point was reiterated that potential benefits of CEOP can be fully exploited by operational centres only if the data collected are available in real time. WGNE members were asked to consider carefully what recommendations could be made to CEOP so that it could better serve NWP centres. These recommendations were later communicated to CEOP.

Dr. J. Roads made a presentation on the current status of CEOP and also responded to concerns that had been previously raised by WGNE and communicated to CEOP namely,

1. *CEOP has not yet fully defined its objectives and modus operandi*
   CEOP is developing a revised implementation plan, which will be made available soon.

2. *Why simultaneous observations?*
   It is problematic to base global observations and modelling on single site observations. CEOP wants to develop the best possible global hydroclimatological 3-dimensional synoptic snapshot to provide the basis for future global hydroclimatological research in a wide variety of climate regimes. CEOP's hope is that these simultaneous observations could be continued beyond the initial CEOP period. It should be stressed that CEOP is a pilot project that could become the basis for an even longer-term experiment.

3. *Will there be a central archive or even a small number of distributed and coordinated archives?*
   University of Tokyo will archive the satellite data, UCAR will archive the in situ data, NASA GLDAS will archive the GLDAS and US LDAS products, as well as pertinent satellite data, MPI will archive the model output. MOLTS will be mirrored at all data centers.

4. *Improved interaction needed with climate modellers, some key people claim they have not heard of CEOP.*
   There is growing community awareness.

5. *Some of the Executive blurb seems to envisage large-scale budget-type studies. However the data exchange is not in place and future long-period analyses would be required. Such studies and the making available of global data sets are a vastly bigger undertaking with nothing in place to make a coordinated program.*
   This is a pilot study. If the data set does not prove to be of value then it can be discontinued. CEOP believes the data sets will be continued and extended.

WGNE members were pleased to see that CEOP had responded to their concerns and that progress had been made. There was still some reservation concerning the request for 3-dimensional fields and some members felt that it might be better for CEOP to concentrate on relevant 2-dimensional fields with possibly higher resolution.

4.4. <u>Current Issues in Data Assimilation</u>

Dr. A. Lorenc made a presentation on the current issues in data assimilation. For synoptic-scale many NWP centres are following a similar approach. Intercomparison of scores from the various centres' global models, and considering changes in scores when new methods are implemented, indicates that variational use of satellite radiances is beneficial; several centres's scores improved when they implemented this. Evidence for the benefit of 4D-Var is less conclusive, but most (but not all) centres have this as a development goal. Others are concentrating effort on making good use of the rapidly increasing volume and diversity of satellite observations.

For mesoscale NWP there is consensus that a priority is to make good use of observations of moisture cloud and precipitation, but there is less agreement as to the methods to use.

Some theoretical issues surround the development of 4D-Var to higher resolutions, where some of the scales resolved may not be predictable for the long time-window needed to make best use of observations at larger scales. One solution is an incremental approach using a perturbation forecast (PF) model and its adjoint, which only represents scales which evolve linearly over the time-window. These models are then not tangent-linear to the full high-resolution forecast model, as in the "classical" 4D-Var approach. An alternative approach is the Ensemble Kalman Filter (EnKF). This has been the subject of much recent research interest. As well as coping with the problems of nonlinearity, it has the advantage of simplicity of implementation. Perhaps the main problem is the ensemble size needed to fit many detailed observations.

Data assimilation is being developed for other purposes than NWP, often with requirements concentrating on reducing bias, rather than the detailed accuracy needed for NWP. Predictions from a modelling system depend on three things: (i) Initial conditions, (ii) Forecast model, (iii) External forcings.

Errors in all of these can accumulate, and in some circumstances grow within the system, to limit the predictability. In NWP (i) and (ii) dominate. This has caused operational centres to concentrate on the development of sophisticated atmospheric data assimilation schemes, built round high-resolution forecast models. The usual understanding of "climate" implies that the detailed initial conditions of the atmosphere are not important. Chaotic growth of initial errors means that the dependence of forecasts on initial conditions cannot be predicted beyond about 10 days. Coupled climate systems have components without this rapid growth, e.g. the ocean, the land surface, etc. Seasonal forecasts do depend on the initial conditions of these components, giving some skill to predictions of the atmospheric climate. Their initial conditions are best determined by coupled data assimilation systems, with an atmospheric component. For longer period forecasts, e.g. the anthropogenic climate change problem, predictive skill has traditionally come more from the response of the forecast model to changes in external forcing, than from the initial conditions. Even here, data assimilation is important for model validation and development.

Dr. A. Joly made a presentation on the work being done at Meteo-France on data assimilation and in particular on adaptive observation and adaptive assimilation relating to mid-latitude cyclogenesis. During FASTEX, there was a 60% increase (on average) in the number of in-situ vertical profiles of the troposphere along the North-Atlantic storm-track during a period of two-months. The impact of these additional data in a series of variational assimilations has been documented by Desroziers, G., G. Hello, and J.N. Thépaut (2002). The extra-observations have a strong positive impact on the analyses but the impact diminishes rapidly in the forecasts. This result is consistent with standard evaluation in NWP where all days are mixed for a given area of interest. Adaptive observations, on the other hand, can potentially improve forecasts for individual cases for a given area of interest. In order to realise this potential, studies are being conducted which include extension of sensitivity for some forecasts to initial conditions back to the observations, e.g. finding which satellite radiance sounders are most important in presence of other sources of data; development of improved adaptive algorithms which can provide an estimate of the reduction of the variance of forecast errors for some forecast proprerty for each planned deployment of extra observations; 'adaptive assimilation' which takes into account that data assimilation systems are not optimal and may be improved locally on a day-by-day basis.

5. NUMERICAL WEATHER PREDICTION TOPICS

5.1 Short- and medium-range weather prediction

At its fifty-fourth session, the WMO Executive Council approved the WMO Statement on the scientific basis for, and limitations of, weather and climate forecasting which was largely prepared by WGNE, WWRP and WGSIP. The Council commended the CAS for its substantial efforts in developing the WMO Statement which presented the issues in a balanced manner and which would provide important guidance for NMHSs in their dealings with government officials, users, the media and the general public. WGNE expressed satisfaction at the completed report which is attached as Appendix B.

The World Weather Research Programme

THORPEX: A Global Atmospheric Research Programme

At the invitation of WGNE, Professor A. Thorpe made a comprehensive presentation on THORPEX: A Global Atmospheric Research Programme. A key change in the past year has been the change in focus from a hemispheric to global experiment. Prof. Thorpe described THORPEX as a ten-year international research programme designed to accelerate improvements in short-range (up to 3 days), medium-range (3 to 7 days) and extended-range (week two) weather predictions, and in the societal and economic value of advanced forecast products. The programme builds upon ongoing advances within the basic research and operational forecasting communities and it will make progress by enhancing collaboration between these communities. THORPEX core scientific objectives are to:

- advance basic knowledge of global-to-regional influences on the evolution and predictability of high-impact weather;
- contribute to the development of dynamically-interactive forecast systems, which will include the concept of targeting;
- develop and apply new methods for assessing the economic and societal value of weather information;
- carry out THORPEX Observing-Systems Tests (TOSTs) and THORPEX Regional Campaigns (TRCs);
- demonstrate the full potential of THORPEX research results for improved operational forecasts of predictable high-impact weather events on time scales out to two weeks and beyond. This demonstration, the THORPEX Prediction Experiment, will last for up to one year.

The themes proposed are of major interest to WGNE, and the studies of predictability and observing system issues being taken up will have benefits throughout the WCRP. The international coordination of THORPEX is under the auspices of the WMO, WWRP and WGNE. The THORPEX International Science Steering Committee (ISSC) defines the core research objectives with guidance from the THORPEX International Core Steering Committee (ICSC) whose members are selected by national permanent representatives to the WMO.

WGNE reiterated its support for THORPEX as a collaborative WWRP/WGNE experiment. At the WGNE session, a joint WWRP/WGNE draft resolution concerning the current status and the next steps in the development of THORPEX was reviewed and finalised in consultation with the Chair of the WWRP, Dr R. Carbone. The committees agreed that the essential next step is the development and submission of the detailed THORPEX Science plan for review and consideration by WWRP and WGNE.

Performance of the main global operational forecasting models

As is usual at its sessions, WGNE reviewed the changes in skill of daily forecasts produced by a number of the main operational centres over the past year as presented by Dr.M.Miller. Examples of the twelve-month running means of verification scores (root mean square error) for 500 hPa geopotential in the northern and southern hemisphere at lead-times of two, four and six days, are shown respectively in Figures 1 and 2. For most centres, a marked increase in skill (as indicated by the verification scores of root mean square error of 500 hPa geopotential in the northern and southern hemisphere at various lead times out to seven days) was again apparent; this increase has now been sustained since the first part of 1999. Improvements were particularly notable in the case of ECMWF, NCEP and the Met Office. At all time ranges, the advance in skill of ECMWF forecasts was outstanding. In the southern hemisphere too, there were distinct increases in skill in forecasts from several centres, with levels sometimes approaching those seen in the northern hemisphere. WGNE ascribed this to the increasing capability of using variational data assimilation schemes and an incremental improvement in the exploitation of observational data in the southern hemisphere.

Verification techniques for mesoscale models

Whilst rms errors, anomaly correlations, skill scores etc. are objective indicators of large-scale model performance, consideration needs to be given to providing measures for the much higher resolution and/or mesoscale models now increasingly employed and for verifying predictions of weather elements and severe events. Work is now being undertaken in this area for parameters such as quantitative precipitation forecasts, two-metre temperature and humidity, ten-metre winds, cloudiness etc. For verification purposes, the basic observational data used are SYNOPs, with data from automatic and climate network stations also increasingly important. Additionally, radar data and high-resolution satellite observations have significant potential in this area. There is general consensus that new methods are needed for the verification of mesoscale models, that there should be enhanced international exchange of the relevant data, and that intercomparison of model scores can be useful if done thoroughly and consistently. The issue has been actively discussed at the past two WGNE sessions.

In its 13th session, the WMO/Commission for Atmospheric Science tasked WGNE to prepare a position paper on high-resolution model verifications, oriented towards weather elements and severe weather events (item 5.3.10 of the abridged final report, document WMO-N°941). This recognizes the specific difficulty of traditional verification methods in providing a useful measure of model performance at high resolution and for intense events. First, the verification of mesoscale events is limited by the insufficient density and quality of the observing networks. Second, the related weather elements may be on the edge of predictability, or entirely stochastic from the perspective of current NWP models. As such, the traditional verification methods based on instantaneous comparison of analyzed and predicted fields may not yield useful information, and new methods are needed. Third, there is a great expectation that mesoscale models will deliver products of direct relevance to end-users, and consequently much work is done on the development of user-oriented verifications, but the needs are not the same for user-oriented and developers-oriented verifications.

The verification of numerical models against observations has several purposes. For instance: (i) provide a measure of the progress of the forecast skill over the years; (ii) compare the merits of two versions of a forecasting system in order to decide which is the best for operations; (iii) understand where the problems are and what aspects of the system need refinements; (iv) compare the relative value of two different systems for a specific category of users. No single verification system can be optimal for all of these tasks and there is a need to issue guidance on what methods are good for what purpose.

A position paper on verification has now been prepared by Dr. P. Bougeault on behalf of WGNE and is included in Appendix C. The purpose of the position paper is to report on a survey of methods currently in use or under development in many operational NWP centres, and to provide guidance on desirable features for verification methods, based on shared experience.

In recognition of the importance of verification in general, there is now a proposal to form a joint WGNE/WWRP Working Group on Verification.

Performance of models in high latitudes

Dr. Kattsov informed WGNE on the on-going international activity in the field of high-latitude climate modelling. In particular, he reported on findings of a Workshop on modelling of the Arctic atmosphere (Madison, Wisconsin, 20-22 May 2002) sponsored by the International Arctic Research Center (IARC) through its Community Arctic Modeling Program (CAMP: http://www.iarc.uaf.edu/camp.html). The workshop's objectives were to (1) identify systematic errors in simulations of the Arctic atmosphere by global and regional models, (2) explore reasons for differences in the Arctic simulations by different models, and (3) identify priorities for reducing model errors in the Arctic. Dr. Kattsov summarized the workshop assessment of Arctic simulations by a wide variety of models with an emphasis made on global AGCMs, as well as associated major challenges and needs facing the modelling community.

Intercomparison of typhoon track forecasts - Dr. K.Saito

Dr. K. Saito reported on the Intercomparison of Tropical Cyclone Track Forecasts for 2001. This model intercomparison started in 1991 for the western North Pacific area and was extended to include the north Atlantic region (1999) and north Pacific (2000). The modelling centres currently participating in the intercomparison are ECMWF, UKMO, JMA, CMC and DWD. Data used for the intercomparison are the mean sea level pressure predicted by the global models of the five forecast centers. The initial time is 12 UTC except CMC (00 UTC). For the best track data, TC positions by JMA are used for the western North Pacific area, while positions by NOAA are used for other areas. UKMO uses bogus data for all regions, and JMA uses bogus data in the western North Pacific area. In the mean positional errors of the 72 hours forecast, ECMWF and UKMO show the best performance in the North Atlantic area. ECMWF and JMA show the best performance in the eastern North Pacific area. Similar results are obtained in the western North Pacific area.

Time series of the position errors of 72-hour forecasts in the western North Pacific area show that forecast errors of JMA were dramatically reduced after 1996 when the Arakawa-Schubert cumulus convective parameterization was implemented. The position errors by CMC has also been significantly decreasing in these a few years. Southwestern mean bias errors are commonly seen in all forecast centers after the recurvatures. In case of JMA, negative bias of westerly wind in the mid-latitude of Northern Hemisphere may affect the mean bias errors.

A multi-model ensemble forecast was tested using the TC track forests of the three forecast centers (ECMWF, UKMO and JMA) of 1991-2001. After 1996, the ensemble mean forecast shows the best performance among the forecasts, which suggests the usefulness of the multi-model ensemble forecast.

Verification and intercomparison of precipitation forecasts

As a principal contribution to WGNE activities in this area, NCEP, DWD and BMRC have been verifying twenty-four and forty-eight hour quantitative precipitation forecasts from eleven operational NWP models for a six-year period against rain gauge observations over the USA, Germany, and Australia in order to assess the skill in predicting the occurrence and amount of daily precipitation. It has been found that quantitative precipitation forecasts have greater skill in mid-latitudes than the tropics where the performance was only marginally better than persistence. The best agreement among models, as well as the greatest ability in discriminating rain areas, occurred for a low rain threshold of 1-2 mm/day. In contrast, the skill for forecasting rain amounts greater than 20 mm/day was generally low, pointing to the difficulty in predicting precisely where and when heavy precipitation may occur. In spite of the impressive progress made in numerical weather prediction, quantitative precipitation forecasts have only shown marginal improvement over the five to six year period examined. A paper documenting this work has been accepted for publication in the Bulletin of the American Meteorological Society.

The validation of precipitation forecasts has become an increasingly important activity. Accordingly this WGNE project has expanded significantly and the Met Office, Meteo-France, JMA and CMA have also started verifying precipitation forecasts in their regions. Of particular interest is the Met Office study which will attempt to verify precipitation in 3-h periods. This should shed light on model performance during the spin-up period and diurnal variation of precipitation, in addition to the daily rainfall amounts. WGNE was prominently involved in the organization of the International Conference on Quantitative Precipitation Forecasts that was held in Reading, UK in September 2002.

Performance of operational models during floods over Europe in August 2002

There is now increasing interest in the prediction of severe weather events and Dr. D. Majewski reported on the performance of operational models in the prediction of one such event. Severe flooding hit the regions at the river Danube and Elbe during the period 1 to 12 August. Up to 400% of monthly normals were measured at many stations in the catchment areas of these rivers. The first heavy precipitation event, on Aug. 6 and 7, led to flooding of river Danube. The second one, on Aug. 11 and 12, was connected to a low pressure system moving slowly northeastward from the Mediterranean Sea (Gulf of Genoa) to Poland. More than 350 mm of precipitation was recorded at station Fichtelberg in eastern Germany during this period. From climatological records available, one can assume that this case represents an event which happens about once in every hundred years. The total damage due to the flooding in the eastern parts of Germany is estimated to exceed $(US) 9 billion. For an accurate precipitation forecast on Aug. 11 and 12, the models had to forecast the actual track of the low pressure system correctly. The four global models considered (from DWD, Météo-France, UKMO and NCEP) were unable to give the proper track and rainfall distribution three days ahead of the event. Only in the time range 30 to 54h and 6 to 30h did the models pick up the signal, but the positions of precipitation extremes differed by more than 180 km between the four models. Higher resolution regional models with mesh sizes between 7 and 20 km, providing short-range (up to 48h) forecasts were able to improve over the position and amount given by the global models.

Review on status of mesoscale numerical weather prediction

Dr. J. Côté reviewed the current status of mesoscale numerical weather prediction. Mesoscale models have grid spacings of around 50 km, most around 10 km, and a few special purpose models have smaller grids. A limit on resolution of mesoscale models is given by the resolution of global uniform models, some of which are being run at 40 km (eg. ECMWF). Mesoscale models are being used for operational forecasting, urban air quality, dynamical adaptation, quantitative precipitation forecasting, ensemble prediction, aviation, research and development etc. There are two types of lateral boundary conditions: variable resolution and limited-area. Variable resolution is well posed, two-way interactions are allowed but it is more expensive and slower. Limited-area models use a variety of boundary conditions: the most commonly used are Perkey-Kreisberg and Davies which allow one-way interactions only but are cheaper and faster. Research is being carried on more transparent boundary conditions for limited-area models. This is motivated by the fact that as the grid spacings get smaller so do integration domains and the sensitivity to boundary conditions gets larger. The principles for model design can vary but an example is given by DWD/COSMO model which uses non-hydrostatic dynamical equations, efficient numerical method of solution, comprehensive physics package, flexible choice of initial and boundary conditions, mesh-refinement techniques, ability to focus on regions of interest, handles multi-scale phenomena, and uses high-resolution data sets for external parameters. Some commonly used basic design options that are being used for the dynamics include Eulerian / semi-Lagrangian advection, grid-point / spectral discretization, latitude-longitude / Lambert grid.

Some key issues in mesoscale modelling that are the subject of current and future work include –

- What is the relation between resolution and grid spacing?
- What is the appropriate physics for a given grid spacing?
- Should physics be chosen with resolved wave or grid spacing?
- What is the robustness or sharpness of a given physics parameterization?
- How far can we go with a given physical parameterization?
- Should stationary forcings be filtered and by what amount?
- What is the limit of integration time for a useful forecast with one-way nesting and a given domain?
- Is ensemble forecasting preferable to increased resolution?
- Should dynamics and physics timesteps be the same?
- Is increased vertical resolution needed?
- Where should the top of mesoscale models be?
- Should special care be taken about formal conservation?

The AROME mesoscale NWP project

Dr. F. Bouttier reported on the new NWP System called AROME (Applications of Research to Operations at Mesoscale) being developed at Météo-France. It will comprise a limited-area model and a matching data assimilation facility at convection-resolving scales (2 to 3km horizontal grid). The prime objective is the operational production of short-range forecasts of convective storms and floods. A second objective is the unification of research and operational activities at Météo-France by merging the software and expertise on the existing mesoscale communities, Meso-NH (research model) and Aladin (operational dynamical adaptation model). International collaborations are being sought. The main development will be a non-hydrostatic limited-area model with relevant physics, notably: detailed prognostic cloud microphysics, three-dimensional mixing using prognostic turbulent kinetic energy, and interactive coupling with a state-of-the-art land surface model (ISBA scheme, a soil moisture analysis and a hydrological model). The analysis will be three-dimensional variational (3D-Var) using Meteosat geostationary radiances and cloud products, precipitation radar data and mesoscale conventional observing networks.

The current activities are: non-hydrostatic dynamical core intercomparisons (testing Meso-NH, Aladin-NH and Hirlam-NH on trapped lee waves cases), the calibration of flow-dependent humidity structure functions for 3D-Var, the assimilation of MSG radiances, limited-area predictability studies, verification of precipitation using radar data, and studies on software standardization following the PRISM, Hirlam and WRF initiatives.The short-term plans are: coupling some Méso-NH physical parameterizations into the Aladin-NH dynamics, experiments with a 2-km resolution 3D-Var data assimilation using Méso-NH as assimilating model, studies on physical initialization, cycling and lateral boundary coupling, the development of preprocessing radar reflectivities for their assimilation, and studies on linking mesoscale NWP with nowcasting.

5.2 Ensemble prediction

Recent developments in the use of ensemble prediction techniques were presented by several participants. Dr. A. Lorenc reported on the developments in the use of “Poor man’s Ensemble Prediction System (PEPS) ” at the Met Office in UK. Following encouraging preliminary results, a higher resolution version of PEPS was set up by collaboration with most of the global NWP centres in the world. Forecasts of 6 fields are collected daily from 15 models and model versions spanning 10 centres on a common 1.5° grid, and verified in comparison with the ECMWF EPS. The performance of the PEPS is very similar to that of the EPS at T+24. (If the EPS forecasts are penalised by a time lag on the assumption that they are available later, the PEPS scores better). Results at other lead-times are similar, although interestingly at some lead-times around 4 days the PEPS does have a small advantage over the EPS. This is surprising as it was expected that the singular vector perturbations of the EPS would have a distinct advantage at longer lead-times. Full analysis of these results is in progress, and further experiments have been set up to include bias corrections of all the models during the winter of 2002/03.

A post-processing system has been developed to assess the probability of severe weather events from the ECMWF operational EPS in support of the National Severe Weather Warnings Service (NSWWS). Since the model resolution is insufficient to represent the full intensity of true severe weather events, the system was tuned to optimise the probabilistic performance over the winter of 2000/01. Subsequent verification over the winter of 2001/02 has shown that the system is capable of providing reliable probabilities of severe weather events and of discriminating those occasions when severe weather is more likely to occur. However, as anticipated, on most occasions severe weather can only be predicted at low probability levels, and it is only quite rarely that the system will generate warnings exceeding the 60% probability threshold required for issue to the NSWWS. Perhaps the most significant result of this work is that the performance of the system is best for forecasts at 4-days lead-time - better than at 3 days, and much better than at 1 or 2 days when the system has virtually no probabilistic skill or resolution. This result is extremely robust, being seen for warnings of heavy rainfall and snowfall, as well as gales, and being quite independent of the tuning of the post-processing. No other verification system has shown similar behaviour in the EPS, although it is well known that it performs poorly at day 1 when the singular vector perturbations are in their rapid growth phase. Most systems verify less severe events, however, so it appears that the peak performance at day 4 may be a property of the ability of the EPS to predict severe events. This may be because the singular vector strategy deliberately seeks out the most extreme developments over the first 2 days, with the result that it cannot provide a quasi-random sampling of the pdf until around day 4, by which time the dynamic modes represented by the perturbations have undergone significant non-linear interactions. For less severe events, more normally verified, this effect may be much less noticeable as we are sampling in the bulk of the pdf rather than in the tails of the distribution.

Dr. M. Miller described the ongoing work on EPS and medium-range predictability at the ECMWF. The current operational EPS consists of 51 members run once a day, from 12 UTC. A more frequent updating of probability forecasts in the early medium range may be beneficial, particularly for severe weather events. To investigate this, an additional real-time experimental run of the EPS has been made each day initialised at 00 UTC. Combining the ensembles from 12 and 00 UTC may also be helpful in providing better sampling of the forecast probabilities. An alternative configuration of equal cost would be to run a single EPS initialised at 12 UTC but containing 100 perturbed members. This arrangement also has potential benefits for extreme event forecasting: using 50 instead of the current 25 singular vectors allows sampling from more initial growing structures. A 100- member EPS generated in such a way may identify potential severe events not captured by a 50-member system. It is not clear a priori whether more frequent updating or including additional growing structures will bring the greater benefit. To assess the relative impacts, a set of 100-member 12 UTC ensembles has been run over the last year, providing a sample of 20-25 forecasts for each season. Conclusions, based on probability forecasts of the extreme event ' 500 hPa height more than two standard deviations away from normal' over the northern hemisphere, are consistent across seasons. Performance has been assessed using the Brier skill score (BSS), ROC area and potential economic value (based on the cost- loss model). To some extent the conclusions vary depending on the score under consideration, but the results are representative. It is found to be certainly beneficial to increase ensemble size from 50 to 100, including more singular vectors. Compared to the current operational 50-member EPS, the 100-member EPS gives gains in predictability of up to 6 hours for the BSS, over 12 hours for the ROC area, and from 0 to 24 hours for value, depending on forecast lead time and, for value, on the user. The effect of combining the operational 12 UTC EPS with the 00 UTC EPS initialised 12 hours earlier is mixed. There is a 12-hour gain in predictability for ROC area, but no improvement for BSS. Overall the greatest benefit is gained by having as many members as possible from the most recent analysis. An operational EPS initialised every 12 hours may be most useful for extreme events, provided that it is available in a timely manner and that users can update their decisions more than once a day as new information arrives. An optimally-weighted combination of the two most recent ensembles will provide additional benefits.

ECMWF is now routinely running forecasts using the ECMWF model initialised with the 12 UTC analyses from DWD, Meteo-France, NCEP and the Met Office. A 10-day forecast is run from each analysis and from a 'consensus' analysis generated as a simple average over all the available analyses, including the ECMWF analysis. The forecasts are run at T_L255L40 resolution using the same configuration of the ECMWF model as is used for the EPS. Study has been made of the average 500 hPa height anomaly correlation for spring 2002 over the Northern Hemisphere for each member of this multi-analysis (MA) system and for some of the corresponding operational forecasts from the centre providing the analysis. There are clear differences in overall performance between the MA forecasts started from different analyses. Over this set of cases the predictions from the DWD and Meteo-France analyses have the lowest scores while the consensus analyses provides the most skilful forecasts. The performance of the MA-forecast from the DWD analysis is similar to that of the DWD operational forecast and similar correspondence is found for the forecasts from the Met Office. The general similarity in performance between forecasts from the same analysis (but using different models) is also apparent in the daily scores - indicating that analysis differences are more important than model differences (during the forecast) in explaining forecast differences between operational centres. The set of 6 MA forecasts (including the EPS control) can also be considered as an ensemble. The ensemble mean skill is similar to that of the operational EPS until day 5/6; beyond this the operational EPS is better. However, the MA system has substantially less spread than the EPS, and the MA spread is too small when compared with the ensemble-mean error. The effect of this reduced spread, together with the small ensemble size, is most apparent in the probability scores where there is a clear advantage for the operational EPS in the medium range, especially for the more extreme events.

Dr. J. Nicolau described the work on short-range ensemble forecasts at Météo-France in the PEACE project (Prévision d'Ensemble A Courte Echéance). This ensemble is devoted to detect rare severe events such as storms in the short range (24-48h). Emphasis is put on assessing skill of predicting strong mean sea level pressure gradient probabilities. Because of heavy computational cost, the ensemble is limited to 11 members (10 perturbed + 1 control). It is based on the global spectral ARPEGE model which has a stretched grid. Initial perturbations used in the ensemble are generated by the singular vectors (SV) technique. One particular feature is the singular vectors optimization over a limited area including the Western Europe and the northern part of the Atlantic Ocean. By this way, perturbations are believed to be efficient in the area of interest. No physics is used in the singular vectors computation (apart a simplified physics including diffusion). The total energy norm is used both at initial and final time with a T63 spectral truncation (the stretching is not used in SV computation). Time optimization is fixed to 12 hours. Five perturbations are built by combining the first 16 targetted singular vectors. Perturbed initial conditions are created by adding and subtracting these perturbations to the unperturbed analysis. Then these 11 initial states are integrated up to 48h.This ensemble has been tested over a sample of 68 independent cases of observed or/and forecast storms between December 1998 and April 2002. It appears that although the

PEACE ensemble lacks in spread, it provides an acceptable set and outperforms the ECMWF EPS in MSLP gradients detection. This system is now running routinely once a day at 00UTC up to 48h. Different products will be developed and tested by operational forecasters during the next winter season. The development of this short-range ensemble is in its first stage. Improvements will be necessary in different ways:

- Inclusion of past errors in the initial state uncertainties sampling. This could be done by the implementation of a data assimilation ensemble.
- Enhancement of model perturbations by tuning physical parameterizations and diffusion.

A limited area model (ALADIN) will be used for the detection of heavy precipitation.

Dr.K.Puri reported on the BMRC ensemble prediction system which has been developed for both the global and limited area systems. The two systems follow rather different procedures in generating the initial perturbations and in allowing for model uncertainties. The medium-range Ensemble Prediction System (BoM-EPS) based on the global model consists of a 33-member ensemble of 10-day forecasts from the BoM global assimilation and prediction system. The perturbation strategy used in generating ensemble members follows the singular vector approach pioneered by ECMWF. Perturbations were scaled linear combinations of the sixteen fastest growing 48-hour T42L19 adiabatic singular vectors, localized polewards of 20^o latitude. The model uses a resolution of T_L119L19 with semi-Lagrangian advection and the suite is run twice daily (00Z, 12Z). The system is currently undergoing trials prior to operational implementation. An intercomparison of the performance of the ECMWF and BoM ensemble systems for the southern hemisphere over the five-month period April to August 2001 has been carried out. The ECMWF products appeared to have an advantage in skill of 12-36 hours compared to those of BoM in the medium-range, although the overall characteristics of the two sets of ensembles were similar.

The regional Ensemble Prediction System (LAPS-EPS) uses assimilation of randomly perturbed observations during data assimilation to generate initial perturbations. Model uncertainties are accounted for by using two sets of convective parametrizations – Tiedtke mass flux and Kuo cumulus convection, and stochastic physics formulation as developed at ECMWF. Lateral boundary uncertainties are allowed for by using individual members from the global EPS. Another feature of the LAPS-EPS is the use of perturbed tropical cyclone bogus data which allows the system to provide estimates of TC track uncertainties. The LAPS-EPS uses a resolution of 75km with 29 vertical levels, has 16 members and the system is run out to 3 days from the 12Z base times. The system is being run routinely once a day at 12UTC out to 72h.

5.3 Recent developments at operational forecast centres, including development of long-range and seasonal forecasting systems

Further to the information on progress in ensemble prediction systems presented in section 5.2, reports were given by participants in the session from the main operational forecasting centres on recent developments/extensions/improvements in systems. As usual, constructive discussions on problems of mutual interest took place. A summary of the status of models (global and regional) now in use, and those foreseen in the next three to five years, as well as computing resources is shown in Table 1.

ECMWF(Dr.M.Miller)

The Centre's forecasts attained new levels of quality in 2001-2002 based on recent research developments, improvements in satellite data and on growth of the Centre's computer capabilities. Over the past year, progress has been made in research and development in key areas. Cycle 24r3 of the IFS, implemented on 22 January 2002, included the introduction of finite elements in the vertical discretisation of the dynamics, resulting in major reductions in stratospheric noise and much improved stratospheric scores. The cycle included the pre-conditioning of the 4D-Var minimisation, with marked benefits in the stability and efficiency of the calculation. The numerics of the two-way wind-wave coupling were improved, ameliorating the problem of occasional over-intense secondary lows. Changes in data usage included introduction of QuikSCAT data, more extensive use of ATOVS data and revised bias corrections for SSM/I data. In the ensemble prediction system (EPS), perturbations based on diabatic singular vectors were introduced in ocean basins in which tropical cyclones are found. This improved the spread of tropical cyclone tracks in the EPS and improved the quality of forecast products such as strike probabilities.

Cycle 25r1 of the IFS, implemented on 9 April 2002 included both model and assimilation changes. The revised shortwave radiation scheme uses six bands rather than four, to improve accuracy both in the lower troposphere and in the upper stratosphere. The TESSEL land surface scheme was re-tuned to reduce biases, and the physics of the ocean-wave model was improved. The Cycle included two important milestones in the use of satellite data: the assimilation of ozone data from SBUV and GOME, and the assimilation of Meteosat- 7 clear-sky water-vapour radiances.

Migration of the model, assimilation and EPS software to the IBM is proceeding as planned. On completion of the migration exercise in Autumn 2002, the next major change to the operational system (Cycle 25r3) will include model and assimilation changes. Amongst the model changes will be an improved convection scheme, improved numerics in the cloud scheme. The assimilation changes will include an inner-outer loop algorithm for 4D-Var, where non-linear processes such as data quality control and scatterometer de-aliasing are performed in the outer loop, so that the inner loop presents a quadratic minimization problem, leading to improved convergence. For improved accuracy, the high-resolution trajectory will be interpolated to the grid of the low-resolution model in the inner loop. Two inner loops will be used with resolutions of T_L95 and T_L159. With the current operational linearised physics, there is no benefit in using a T_L255 resolution in the third inner loop. A more comprehensive and accurate set of linearised physics is in preparation, and will be used in further experimentation on the resolution of the inner loops. The next cycle will include changes in data usage including assimilation of radiances from NOAA-I 7, sky radiances from SSM/I, additional HIRS radiances, water vapour radiances from GOES and SAR wave spectra from ERS-2.

Work is proceeding on the development of the severe weather forecast system. Within the research department considerable effort has been devoted to comparing the merits of a 100-member ensemble delivered once-daily and a 50-member ensemble delivered twice-daily. For applications in severe weather forecasting, the twice-daily system appears to be better.

System 1 of the seasonal forecast system was prepared in 1995-1996. It has been running routinely since then and performed well on the major ENSO event of 1997-1998. System 2 of the seasonal forecast software includes a number of important scientific changes, provides much more technical flexibility and has been implemented for operational running.

The UK Met Office has implemented its seasonal forecast model at ECMWF. Seasonal forecasts from the Met Office system will be generated with the same forecasting protocol (start-dates etc.), archived in the same format and verified and displayed with the same tools as the ECMWF system. It is expected that the Met Office system will become operational at ECMWF in the first half of 2003, probably after commissioning of the new IBM High Performance Computing Facility.

An experimental monthly ensemble forecast system has been developed and is now run twice monthly to explore the significance of an active ocean for mid-latitude predictability on time scales of 1-4 weeks. The atmospheric resolution is T_L159.

As part of its preparation of the Sixth Framework Programme (FP _6), the EU Commission invited expressions of interest in proposing 'Integrated Projects'. The Centre took the initiative to lead two such expressions of interest, one in global environmental monitoring (GEMS), and the other on seasonal to decadal forecasting (EURIPDES). The preparatory workshop in May 2002 for the GEMS proposal attracted more than 80 leading scientists from more than 25 institutes across Europe.

BMRC (K. Puri)

The current suite of global and limited area models at the Australian Bureau of Meteorology consists of:

- the global assimilation prediction (GASP) system, horizontal resolution T_L239 and 29 levels;
- the limited area prediction system (LAPS), horizontal resolution $0.375^o \times 0.375^o$ and 29 levels;
- the tropical limited area prediction system with the same resolution;
- the mesoscale limited area prediction system, horizontal resolution $0.125^o \times 0.125^o$ and 29 levels;
- the tropical cyclone limited area prediction system, horizontal resolution $0.15^o \times 0.15^o$ and 19 levels.

In addition a $0.05^o \times 0.05^o$ version of the model is run operationally twice a day for domains covering Melbourne and Sydney, with hourly output then being used to drive a CSIRO photochemical model for use by the Environment Protection Authorities.

An upgrade in operational LAPS to use 1D-Var assimilation of satellite radiances together with hourly radiation calculations (instead of 3-hourly currently) and soil moisture nudging was implemented in September 2002 following extensive testing. A detailed bulk explicit microphysics scheme has been implemented in LAPS and is currently undergoing testing prior to operational implementation.

An extended version of the global system (50 vertical levels with the top level at 0.1 hPa) has been developed which allows the full forward calculation of ATOVS radiance first-guess values in the 1D-VAR retrieval scheme. Extensive global assimilation experiments have been conducted and medium-range prediction performance in the stratosphere has been substantially improved. Scatterometer (QuickScat) data are now being assimilated on an experimental basis within GASP (T_L239/L33 vs operations T_L239/L29), and has shown a modest positive impact on medium-range prediction in the Southern Hemisphere. Quality control procedures have been supplemented with background checks of wind direction to remove incorrectly de-aliased data. The scatterometer data is expected to be included into the operational global system as part of the next major upgrade.

A significant amount of work has been carried out in implementing message passing in both the global and regional systems in preparation for the BoM tender for a supercomputer upgrade in 2004. The Meteorological Archival and Retrieval System (MARS), a software package developed at the ECMWF, was made available to the Bureau late in 1998. MARS has now been implemented in the Bureau and is currently used to archive selected global model and global ensemble system output, in addition to research experimental data.

POAMA (Predictive Ocean Atmosphere Model for Australia) is a seasonal to inter-annual climate prediction system based on coupled ocean and atmosphere general circulation models. It was developed in a joint project involving the Bureau of Meteorology Research Centre (BMRC) and CSIRO Marine Research (CMR), with some funding coming from the Climate Variability in Agriculture Program (CVAP) of Land and Water Australia. The POAMA model is a significant improvement over earlier versions of coupled models for seasonal forecasting at BMRC. It uses the latest ocean and atmosphere general circulation models. In addition real time oceanic and atmospheric initial states are used to initialise the coupled model. These are provided by an ocean data assimilation system that is run in real time as part of the POAMA system and by the Bureau of Meteorology operational weather analyses.

The atmospheric component of the coupled model used in POAMA is the Bureau of Meteorology unified atmospheric model (BAM). A modified convection closure is used because it allowed the model to have a good representation of the MJO. It has a horizontal spectral resolution of T47 and has 17 vertical levels. The ocean model component is the Australian Community Ocean Model version 2 (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. The maximum depth is 5,000 metres. The ocean and atmosphere models are coupled using the Ocean Atmosphere Sea Ice Soil (OASIS) coupling software (developed by CERFACS, France). This coupler gives high flexibility for changing model components in the future as models further improve. The ocean data assimilation scheme is based on the optimum interpolation (OI) technique. Only temperature observations are assimilated and only measurements in the top 500 are used.

The POAMA system has been run in real-time every day by the Bureau of Meteorology operations branch since 1st October 2002. A 9 month coupled model forecast is produced daily in real-time using the very latest ocean state. The system has displayed significant skill in predicting features of the 2002 El Niño event.

Météo-France (Dr. M. Déqué)

Météo-France is both an operational forecast and a research centre. As an operational forecast, Météo-France produces each month a 4-month forecast. The set-up is the same as the ELMASIFA (see below) project, but with ARPEGE.3 and 9 members. These forecasts are not delivered to the public, but displayed on intranet and delivered in the framework of conventions with customers (insurance, energy). The research activities are driven by international projects (mostly EU-funded). The main projects of the last 5 years are:

PROVOST: In this project led by ECMWF, four GCMs using observed SSTs attempted to reproduce forecasts during the ERA15 period. The results showed a good performance in the tropical belt and significant scores in winter northern mid-latitude, although it is difficult to get a stable estimate of the scores over Europe.

ELMASIFA: This project used ARPEGE.1 to evaluate the winter rainfall over North Africa. The experimental set-up was similar to PROVOST, except that instead of using observed SSTs, a statistical prediction scheme was used. The predictability, good with observed SST, disappears with predicted SST, except along the Atlantic coast.

POTENTIALS: This project is similar to PROVOST, except that a statistical correction is introduced into the model (ARPEGE.1 then ARPEGE.3) equations to correct the tendency error. When a constant correction is applied, the systematic error is reduced, as expected, but the scores are unchanged. Using a specific correction for each year, which is not applicable directly in real-time forecasting, brings some score improvement.

DEMETER: This project, led by ECMWF, is the sequel of PROVOST. Seven models are used, the lead time is 6 months, the GCMs are coupled to an ocean model, and the forecast period is ERA40. Preliminary results show that coupling provides better scores in the tropics than statistical SST prediction, whereas the opposite is obtained in the midlatitudes.

Japanese Meteorological Agency (Dr. K. Saito)

Numerical weather prediction at JMA started in 1959, and the current system, which was implemented in March 2001, is the 7th generation. The current main computer is the Hitachi SR8000E1, which attains 768 Gflops by 80 nodes. JMA has five main forecast models. The global spectral model (GSM) with T213L40 has been operated twice a day. The forecast times are 216 hours (12 UTC) and 90 hours (00 UTC). The regional spectral model (RSM) is a model for short-range forecast and covers East Asia with a horizontal resolution 20 km (L40). This model is run twice a day, and forecast time is 51 hours. The mesoscale model (MSM) is for disaster prevention and covers Japan and its surrounding areas with a horizontal resolution 10 km (L40). This model is run 4 times a day (6 hourly), and forecast time is 18 hours. The typhoon model (TYM) is for track and intensity forecast of tropical cyclones. Horizontal resolution is 24 km (L25) and is run 4 times a day (6 hourly) up to 84 hours when a tropical cyclone exists in the Northwestern Pacific. The Ensemble prediction model (EPS) is for one-week and long-range forecasts. For one-week forecast, a low-resolution version of GSM (T106L40) is used and 25 members are run daily (216 hours) from 12 UTC. For long range forecast, the T106 model with 26 members is run once a week up to 34 days, and the T63 model with 31 members is run once a month up to 4 months and twice a year up to 8 months.

JMA implemented a 3D-Var data assimilation system for global analysis in September 2001 resulting in improved performance of the GSM. The RMSE of the 500 hPa temperature field and the 250 hPa wind field of northern hemisphere decreased by about 10 %. Mean positional error of typhoon tracking of GSM scored a minimum in 2002. A newly developed 4D-Var data assimilation system (Meso 4D-Var) was implemented for MSM analysis in March 2002. In the operational 4D-Var assimilation system, an incremental approach is taken with an inner loop model with resolution of 20 km L40 for high-speed processing. Inner forward model has nonlinear full-physics model, while inner backward adjoint model uses reduced physics (grid-scale condensation, moist convective adjustment, simplified vertical diffusion and simplified long wave radiation). Consecutive 3-hour assimilation windows are adopted, and minimization processes are limited up to 20 iterations for efficiency. Using 40 nodes of the Hitachi-SR8000, 15 minutes elapsed time is required for one 3-hour assimilation window. Radar-AMeDAS (automated rain gauge network of JMA) precipitation data is assimilated. Threat scores of precipitation improved with the implementation of the 4D-Var. Use of wind data of the domestic ACARS from the Japan Air Line started in August 2002. Using ACARS data in the Meso 4D-Var, positional errors of the low-level wind shear lines can be reduced.

UK Met Office (Dr. A. Lorenc)

A new version of the Unified Model was made operational on 7th August 2002, in the global and mesoscale model. It uses the non-hydrostatic form of the governing equations, making the model suitable for use at very high resolutions (a research programme is underway for its use in convective-scale NWP). The equation set also includes the extra terms normally ignored when making the shallow atmosphere assumption, thus allowing a complete representation of the Coriolis force. A semi-implicit scheme is used to solve the governing equations. It is designed to conserve mass, mass-weighted potential temperature and moisture, and angular momentum. This ensures the integrity of the solutions when undertaking the long integrations required for climate-change experiments. The equations are integrated using a predictor-corrector method, which requires the solution of a 3-dimensional, Helmholtz-type equation using a generalised conjugate residual technique. Initial estimates of the prognostic variables are obtained by semi-Lagrangian advection using a two-level scheme. For potential temperature a non-interpolating scheme is used in the vertical. For moisture, quintic interpolation is used in the vertical only, since lower order interpolation in the vertical results in excessive drying in the tropical lower stratosphere. The new model uses physical parametrizations developed also for climate simulations, with boundary layer, radiation, cloud ice and convection in particular differing significantly from the previous global NWP model.

The main results from extensive pre-operational trials were reduced errors in the northern hemisphere, improved forecasts of aviation winds, and improved position and intensity of tropical cyclones. Forecasts scores in the southern hemisphere were worse; after these trials this was found to be due to an error in the time interpolation of the forecast to ATOVS soundings, now corrected. Tropical circulations were degraded; the Hadley circulation shows a clear spin-up with forecast range. The error structure suggests a thermally direct circulation error in the upper troposphere leading to an increase in the flow at 250hPa and a thermally indirect circulation error in the lower troposphere, which reduces the poleward flow in the boundary layer. This error structure in the vertical may be linked to incorrect distribution of diabatic heating in the vertical in the tropics. This error remains the major obstacle to use of the new model version in climate simulations, and is the subject of intensive study, using (among others) aqua-planet simulations.

The non-hydrostatic model is the core to basic research projects into various problems associated with modelling and data assimilation with partially resolved convection. These should lead to the ability to implement convective scale forecasting (at about 2km resolution) after the next computer upgrade in 2008.

The Met Office have successfully developed and tested a global 4D-Var, and are now proceeding with parallel projects to develop that system for operational implementation in 2004 and to research methods for including assimilation of cloud and precipitation data in global and, more importantly, higher resolution limited area 4D-Var.

Deutscher Wetterdienst (Dr.D. Majewski)

The strategic goal of NWP at DWD can be summarised as follows: “DWD needs a reliable NWP system which forms the basis of largely automated short range weather forecast system”. Thus NWP research and development at DWD concentrates on local short range *weather* forecasts with main emphasis on the hydrological cycle.

With the installation of the new computer system at the end of 2001, an IBM RS/6000 SP with 80 nodes (16 Power III processors), the operational NWP suite of DWD can be upgraded in 2003. The mesh size of the global model GME will be reduced from 60 to 40 km, and the number of layers increased from 31 to 40. For global data assimilation, the development will concentrate on a 1D-Var scheme for the usage of satellite radiances (ATOVS data), scheduled for 2003, and the implementation of a 3D-Var scheme (physical space assimilation system) during 2004. For the nonhydrostatic local model LM with a mesh size of 7 km, the model domain will be increased from 325x325 to 750x638 grid points to cover whole of Europe, and the number of layers will be increased from 35 to 40.

To aid short range forecasts of severe weather, a model-based system for nowcasting and very short range forecasting will be developed at DWD by the end of 2005. A very high resolution version of LM with a mesh size of 2.8 km (to allow for explicit prediction of deep convection) will provide 18-h predictions every three hours (based on 00, 03, 06, ..., 15, 18, 21 UTC analyses). At this scale, the proper initialisation of moisture, clouds and precipitation based on GPS, satellite and radar data will be essential.

Russian Hydrometeocentre (HMC) and the Voeikov Main Geophysical Observatory (MGO) (Dr. V. Kattsov)

An intercomparison was made of experimental one-month weather forecasts (1 April 2001 – 1 April 2002) between the HMC and the MGO. The forecasts were intercompared for the averages over the first ten days; over the first twenty days, and over the entire month. The HMC forecasts were based on a five-member ensemble obtained from a T42L15 model for the first ten days and then regression for the rest of the month. Initial states for the ensemble were obtained using lagging techniques with a 12-hour interval. The MGO forecasts comprised an eleven-member ensemble obtained from a T30L14 model for the three ten day periods. Initial states for the MGO ensemble were obtained using breeding techniques. Concurrently, starting from December 2001, MGO one-month forecasts were also produced using a higher resolution version of the MGO model – T42L14, so that by the end of the intercomparison, the MGO had provided two full sets of monthly forecasts obtained from the two versions of the model. The intercomparison was carried out for the extratropical Northern Hemisphere, for sea-level pressure, temperature at the 850 hPa geopotential surface height, and height of the geopotential surface 500 hPa. Additionally, 2m air temperature forecasts were intercompared for 70 meteorological stations in Russia and neighbouring countries.

A new international effort in seasonal forecasting in the framework of Asia-Pacific Economic Cooperation (APEC) has been started. The objective of APCN (APEC Climate Network) is to develop a multi-model ensemble system (MMES) for seasonal prediction. The MMES will be operated using boundary conditions provided by multi-model ensemble SST predictions. The forecast system includes procedures for model output collection, bias correction, statistical downscaling, super-ensemble for blending the predictions of different models, and verification. The data collection will be linked to the international Seasonal Prediction Model Intercomparison Project (SMIP2/HFP) of the CLIVAR Working Group on Seasonal to Interannual Prediction (WGSIP) under the auspices of WCRP. In addition to basic research related to climate predictability, this project will also assess the economic value of MMES and develop methods for applying MMES in the industrial sectors. By October 2002, institutions from Canada, China, Chinese Taipei, Korea, Russia, and USA had participated in this effort.

Canadian Meteorological Centre (Dr.J. Côté)

The surface model ISBA (Interactions Soil-Biosphere Atmosphere) was implemented in September 2001 in the regional configuration. ISBA makes use of the tiling approach with each tile split into 4 surface types: land, glaciers, water, sea-ice, and the fluxes are aggregated. A sequential analysis of surface variables was implemented at the same time. These changes have resulted in the model being dryer near the surface leading to improved precipitation.

Another important implementation took place in December 2001. The sub-grid scale orographic blocking now includes the Lott and Miller parameterization which greatly improves the scores especially in winter. Another significant implementation at the same time was an improvement of the data assimilation scheme which allowed the 3D-Var scheme to include more data sources. The new system assimilates the temperature at 27 significant levels rather than the geopotential at 16 mandatory levels that were used previously. Data assimilated now includes: temperature data from AMDAR and ACARS, satwinds from GOES 8 & 9, additional TOVS channels below 500 hPa, additional data from AMSU-B, QuickScat, wind profiler, GOES radiances.

All components were tested individually, then together over many weeks of winter and summer 2001. The impact due to analysis changes was more important in the short range, and in summer situations. The impact due to model changes was more important in the medium range, and in winter situations. Parallel runs were made from mid-October 2001 to mid-December 2001 and the subjective evaluation by the Canadian Meteorological Centre operational meteorologists was very positive. The results were similar for the regional system.

A mesoscale GEM at 2.5 km (0.0225°) horizontal resolution and 43 vertical hybrid levels with a top at 10 hPa is being developed. The physical package includes Kong and Yau explicit condensation, no deep convection, shallow convection, and ISBA surface scheme. The model is integrated for 12h and is initialised with operational regional analysis at 24 km. A region in Southern British Columbia was chosen initially. The preliminary results are physically and dynamically reasonable: the model behaves like the operational GEM and realistic mesoscale phenomena are generated.

4D-Var is being developed around two units: the 3D-Var module and the distributed-memory GEM-DM model. Coupling of those two units is in the spirit of the PALM approach where 3D-Var and GEM-DM are executed simultaneously and exchange objects.

NCEP (Dr. S. Lord)

NCEP is in the process of converting its operations to a new IBM computer, which will have 2.5 times the processing capability of its current machine. Operations are expected to begin in the spring of 2003. A new Global Forecasts Systems (GFS) T254, 64 level system was implemented in late October 2002. The new GFS extends vertically to 0.2 hPa, and has some new scientific improvements including new background error specification, use of microwave sounding data over some land areas, and a new sampling procedure for sounding radiances. In the past year, a lot of work has been done on assessing the global model climate. Anomaly correlation scores for the Atlantic, Pacific North American, and Antarctic Oscillations have been calculated for daily 15-day operational forecasts. Precipitation and upper tropospheric divergence anomalies have been assessed using year long runs with observed Sea Surface Temperature and various model resolutions. The height of the model top and vertical resolution are important factors controlling tropical precipitation patterns. Improvements to marine forecasting over the past year have included introduction of a Regional Ocean Forecast system, run daily over the Western North Atlantic, and an upgraded Wavewatch model with sub-grid island blocking effects.

China Meteorological Administration (Dr. Chen Dehui)

A T213L31 spectral model (originated from ECMWF) has been put into real-time parallel running in comparison with operational T106L19 since June 2001. The statistical and synoptic verification revealed that the T213L31 model gives better performance than T106L19.

The Regional HLAFS_L20 system has been tested by increasing the resolution from 50 km to 25 km and introducing an explicit scheme for parameterization of grid-point scale precipitation. This resulted in improvement on precipitation forecasts. There has been no significant improvement on accuracy of tropical cyclone track prediction by the operational MTTP (Model for Typhoon Track Prediction).

An air quality forecast system based on a non-static advection and dispersion box-model has been developed and is experimentally running on the Chinese-made parallel computer (SW-I) since June 5, 2001, nested with MM5. The system is run once every day to predict the air pollution index (API) of the pollutants (SO2, NO2, PM10 and CO) and PPI for 47 cities across China. A numerical model for sand & dust-storm prediction is under development at NMC/CMA.

No changes have been made for the Ensemble Prediction System of Medium Range Forecast (SV, 32 members, T106L19).

Plans for future include:

- T213L31 assimilation and forecast system with OI analysis scheme will be run fully operationally by the end of 2002.
- The new version of high resolution (0.25°) HLAFS will be put into quasi-operational running in late 2002, and will be run fully operationally in 2003.
- The development of T213L31 system with 3D-Var scheme will be aimed at conducting real-time experiments during next year.
- The migration of MTTP system to the parallel computer (Chinese-made SW-I) including the parallelization of the typhoon track prediction model will be finished by the end of 2002, and experiments performed with a version with resolution of 25 km.
- The new dispersion forecasting system at RSMC Beijing for Environmental Emergency Response (EER) will be established based on the HYSPLIT4 (HYbrid Single-Particle Lagrangian Integrated Trajectory) model from ARL/NOAA in the early of 2002. Some experiments will continue to be conducted to test its performance and the possibilities of extending the application to other fields during next year.

6. OTHER WGNE ACTIVITIES AND FUTURE EVENTS

Publications

One publication had been produced in the WGNE "blue-cover" numerical experimentation since the seventeenth session of the group, namely the annual summary of research activities in atmospheric and oceanic modelling (No. 32, produced in April 2002), again printed and distributed directly by RPN, Montreal.

This year a major step was taken in handling the April 2002 report electronically. Contributions were submitted as an attachment to an e-mail message, or through the web site www.cmc.ec.gc.ca/rpn/wgne. A few contributions still arrived by mail. An electronic version of the report was produced and is available on the web site. About 200 hard copies have also been produced and distributed to those who preferred them.

WGNE thanked Dr. H. Ritchie who retired this year as the editor of WGNE 'Blue Book '. WGNE noted with immense appreciation the excellent editorial service rendered by Dr. Ritchie for four years. WGNE welcomed the new editor Dr. J. Côté who also thanked Dr. Ritchie and noted that this will be a year of consolidation for the electronic publication with the aim of making the process more robust. Forms will be used on the web to submit articles including abstracts.

Next session of WGNE and GMPP and other events

At the kind invitation of the Centro de Previsão de Tempo e Estudos Climáticos (CPTEC), Brazil, the next session of the WGNE, the nineteenth, would be held in Salvador, Brazil, 10-14 November 2003.

	GLOBAL "NOW"	GLOBAL 3-5 YEARS	REGIONAL "NOW"	REGIONAL 3-5 YEARS	COMPUTING "NOW"	COMPUTING 3-5 YEARS
BMRC	T_L239 L29	T_L479 L60	0.125^o L29	0.05^o L60	40 GFLPS NEC SX-5	450 GFLPS NEC SX6
CMA	T213 L31	GRAPES 40KML35	0.25^o L20	GRAPES 15KML40	85 GFLPS IBM	1000 GFLPS ?
CMC	0.9^o L28	0.3^o L60	15KM L42	WINDOWS 2KM L60	200 GFLPS NEC SX-5/SX-6	1500 GFLPS IBM
DWD	60KM L31	30KM L35	7KM L35	2KM L50	240 GFLPS IBM	450 GFLPS IBM
ECMWF	T_L511 L60	T_L799 L90	N/A	N/A	600 GFLPS IBM	2400 GFLPS IBM
JMA	T213 L40	T_L959 L60	10KM L40	5KM L50	80 GFLPS HITACHI SR8000	2400 GFLPS ?
METEO-FRANCE	T199 L31	T403 L41	9.5KM L31	2.5KM ?	400 GFLPS FUJITSU	?
UKMO	60KM L38	40KM L50	12KM L38	2KM L50 10KM L48	150 GFLPS T3E	1900 GFLPS NEC SX6
NCEP	T254 L64	45KM L70	12KM L50	8KM L70	2 at 75GFLPS IBM	2 at 900 GFLPS ?
FNMOC/ NRL	T239 L30	T480 L60	6KM L40	3KM L60	100 GFLPS 03K	650 GFLPS ?

GFLPS = Indicator in Gigaflops of sustained computing capacity
"NOW" = Actually operational or within a few months

METRICS FOR OPERATIONAL NWP CENTERS AS REPORTED TO WGNE - NOVEMBER 2002

Updated 20 June 2003

7. CLOSURE OF SESSION

On behalf of all participants, Dr. K. Puri, Chair of WGNE, and Dr. D. Randall, Chair of GMPP, expressed deep appreciation to Météo-France for hosting the session of WGNE and GMPP, and the excellent facilities and hospitality offered. The opportunity of interacting with many scientists and experts at Météo-France and hearing first hand of the research and work going ahead had been very valuable. Sincere gratitude was voiced to Dr. P. Bougeault and supporting staff for the excellent arrangements, unstinting assistance, and refreshments that had been provided.

The joint eighteenth session of WGNE/sixth session of GMPP was closed at 12.30 hours on 22 November 2002.

APPENDIX A

LIST OF PARTICIPANTS

1. Members of the CAS/JSC Working Group on Numerical Experimentation

Dr. K. Puri (Chairman)	Bureau of Meteorology Research Centre GPO Box 1289K Melbourne, Victoria 3001 Australia Tel: +61-39-669-4433 Fax: +61-39-669-4660 e-mail: K.Puri@bom.gov.au
Dr. P. Bougeault	Centre National de Recherches Météorologiques Météo-France CNRM/GMME/D 42, Avenue Coriolis F-31057 Toulouse Cedex France Tel: +33-5-61-07-93-58 Fax: +33-5-61-07-96-26 e-mail: philippe.bougeault@meteo.fr
Dr. Chen Dehui	National Meteorological Centre China State Meteorological Administration 46 Baishiqiaolu Road Western Suburb Beijing 100081 China Tel: +86-10-6840-8074 Fax: +86-10-6217-2016 e-mail: chendh@rays.cma.gov.cn
Dr. V. Kattsov	Voeikov Main Geophysical Observatory 7 Karbyshev Street 194021 St. Petersburg Russian Federation Tel: +7-812-247-8668 Fax: +7-812-247-8661 e-mail: kattsov@main.mgo.rssi.ru
Dr. A.C. Lorenc	Forecast Research Division Met Office London Road Bracknell, Berkshire RG12 2SZ United Kingdom Tel: +44-1344-856-227 Fax: +44-1344-854-026 e-mail: Andrew.Lorenc@metoffice.com
Dr. M. Miller	European Centre for Medium-Range Weather Forecasts Shinfield Park Reading, Berkshire RG2 9AX United Kingdom Tel: +44-1189-499-070 Fax: +44-1189-869-450 e-mail: mmiller@ecmwf.int

Dr. J. Côté

Recherche en Prévision Numérique
Meteorological Service of Canada
West Isle Office Tower 500
2121 Trans-Canada Highway
Dorval, Québec H9P 1J3
Canada
Tel: +1-514-421-4742
Fax: +1-514-421-2106
e-mail: Jean.Cote@ec.gc.ca

Dr. K. Saito

Numerical Prediction Division
Forecast Department
Japan Meteorological Agency
1-3-4 Otemachi
Chiyoda-ku, Tokyo 100-8122
Japan
Tel: +81-3-3211-8408
Fax: +81-3-3211-8407
e-mail: ksaito@npd.kishou.go.jp

Dr.D. Majewski

Deutscher Wetterdienst
Frankfurter Strasse 135
D-63067 Offenbach am Main
Germany
Tel: +49-69-8062-2728
Fax: +49-69-8062-3721
e-mail: detlev.majewski@dwd.de

Dr. D. Williamson

National Center for Atmospheric Research
Climate and Global Dynamics Division
P.O. Box 3000
Boulder, CO 80307-3000
USA
Tel: +1-303-497-1372
Fax: +1-303-497-1324
e-mail: wmson@ncar.ucar.edu

Dr. S. Lord

NCEP/Environmental Modelling Center
Rm 207, World Weather Building
5200 Auth Road
Camp Springs, MD 20746
USA
Tel: +1-301-763-8000 ext. 7202
Fax: +1-301-763-8545
e-mail: Stephen.Lord@noaa.gov

2. Members of the GEWEX Modelling and Prediction Panel

Dr. D.A. Randall (Chairman)

Department of Atmospheric Science
Colorado State University
Fort Collins, CO 80523-1371
USA
Tel: +1-970-491-8474
Fax: +1-970-491-8428
e-mail: randall@redfish.atmos.colostate.edu

Professor B. Holtslag
Wageningen University
Meteorology and Air Quality Section
Duivendaal 2
NL-6701 AP Wageningen
The Netherlands
Tel: +31-31-748-5519
Fax: +31-31-748-2811
e-mail: Bert.Holtslag@user.metair.wag-ur.nl

Dr. S. Krueger
Department of Meteorology
University of Utah
135 South 1460 East, Room 819
Salt Lake City, UT 84112-0110
USA
Tel: +1-801-581-3403
Fax: +1-801-585-3681
e-mail: skrueger@met.utah.edu

Dr. J. Polcher
Laboratoire de Météorologie Dynamique du CNRS
Tour 25, 5ème étage
BP 99, 4 Place Jussieu
75252 Paris Cedex 05
France
Tel: +33-1-4427-4763
Fax: +33-1-4427-6272
e-mail: Jan.Polcher@lmd.jussieu.fr

3. Invited experts and observers

Drs A. Boone, F. Bouttier, J.-L. Brenguier, M. Déqué, H. Douville, J.-F. Geleyn, A. Joly, E. Martin, J. Nicolau, J. Noilhan, J. Pailleux, J.-L. Redelsperger, S. Valcke and S. Westrelin attended from Météo-France.

Dr. P. Gleckler
PCMDI
Program for Climate Model diagnosis and Intercomparison
Lawrence Livermore National Laboratory, L-103
P.O. Box 808
Livermore, CA 94550
USA
Tel: +1-925-422-7631
Fax: +1-925-422-7675
e-mail: pgleckler@llnl.gov

Professor R. Laprise
UQAM
Department of Earth and Atmospheric Sciences
CP 8888-Succ. Centre-ville
Montréal H3C 3P8
Canada
Tel: +1-514-987-3000, ext. 3302
Fax: +1-514-987-4277
e-mail: laprise.rene@uqam.ca

Dr. J. Roads	University of California, San Diego 9500 Gilman Drive, MC 0224 La Jolla, CA 92093-0224 USA Tel: +1-858-534-7776 Fax: +1-858-534-2099 e-mail: jroads@ucsd.edu
Professor A. Thorpe	NERC Centres for Atmospheric Science Department of Meteorology University of Reading Earley Gate PO Box 243 Reading, Berkshire RG6 6BB United Kingdom Tel: +44-118-931-6452 Fax: +44-118-931-6462 e-mail: a.j.thorpe@reading.ac.uk
Dr. P. Try	International GEWEX Project Office 1010 Wayne Avenue, Suite 450 Silver Spring, MD 20910 USA Tel: +1-301-565-8345 Fax: +1-301-565-8279 e-mail: gewex@gewex.org

4. Secretariat Staff

Dr. V. Satyan	Joint Planning Staff for the World Climate Research Programme World Meteorological Organization Case Postale No. 2300 CH-1211 Geneva 2 Switzerland Tel: +41-22-730-8418 Fax: +41-22-730-8036 e-mail: satyan_v@gateway.wmo.ch
Dr. Zhao Chong Lei	Atmospheric Research and Environment Programme Department World Meteorological Organization Case Postale No. 2300 CH-1211 Geneva 2 Switzerland Tel: +41-22-730-8211 Fax: +41-22-730-8049 e-mail: Lei_Z@gateway.wmo.ch

APPENDIX B

WMO STATEMENT ON THE SCIENTIFIC BASIS FOR, AND LIMITATIONS OF, WEATHER AND CLIMATE FORECASTING

1. Introduction

1.1 Every day around the world, the NMSs and the private sector meteorological service providers of the Member States and Territories of WMO provide hundreds of thousands of forecasts and warnings of weather and climate conditions and events. These forecasts and warnings provide information for the benefit of the community at large and for a wide range of specialized user sectors, on a broad spectrum of atmospheric phenomena ranging from those with timescales of seconds to minutes and space scales of metres to kilometres, such as severe storms, through to those, such as *El Niño*-related drought, with multi-year and global impact. The forecast information provided is used to inform and improve decision making in virtually every social and economic sector and the globally aggregated economic benefits of meteorological services are reckoned to be of the order of hundreds of billions of United States dollars.

1.2 The capacity to provide these socially- and economically-beneficial services to the citizens of the 185 Members of WMO results from the operation of the unique international system of cooperation of the WMO World Weather Watch Programme which is based on:

(a) The collection and international exchange of the global observational data that are essential to describe the current (initial) state of the atmosphere (and the underlying land and ocean) at any point in time;

(b) The fact that the physical and dynamical processes governing the behaviour of the atmosphere and ocean can be represented in numerical models which are capable of providing forecasts of daily weather conditions with significant skill out to several days from the 'initial' state as well as useful indications, in certain circumstances, of general trends of climate for months and seasons ahead;

(c) The existence of a coordinated international meteorological system of global, regional and national data-processing and modelling centres producing real-time products from which skilled professional forecasters are able to prepare forecasts and warnings in forms that are relevant and useful to the user community;

(d) The ability to monitor extreme events in real-time and to issue warnings by combining classical meteorological observations, model output and information from remote-sensing systems such as satellites and radar.

1.3 The scientific understanding and technological capabilities underlying this globally cooperative system of weather and climate forecasting have made enormous progress over the past 25 years as a result, in particular, of such cooperative international research programmes as the WMO/ICSU Global Atmospheric Research Programme, the WMO World Weather Research Programme and the WMO/ICSU/IOC World Climate Research Programme. The skill levels and utility of the resulting forecasts and warnings have steadily increased. Indeed three-day forecasts of surface atmospheric pressure are now as accurate as one-day forecasts 20 years ago. But the observational database necessary to describe the 'initial' state of the atmosphere will always be limited by considerations of scale and measurement accuracy, the processes governing the behaviour of the atmosphere are non-linear and the phenomenon known as chaos imposes fundamental limits on predictability. While new techniques are emerging which help potential users of weather and climate forecasts to understand better, and make allowance for, the inherent uncertainties in the forecasts, the WMO Executive Council believes it is important that all those who make use of such forecasts in decision making should be made better aware of both their scientific foundation and their scientific and practical limitations. It therefore requested that CAS prepare a statement on the current status of weather and climate forecasting.

1.4 This statement has been prepared by CAS with input from other WMO and external scientific organizations and programmes including the World Climate Research Programme. It was approved by the thirteenth session of CAS in Oslo in February 2002 and endorsed by the Executive Council at its fifty-fourth session in June 2002. It is provided for the information of all those with an interest in the scientific foundations and limitations of weather and climate forecasting on timescales from minutes and hours through to decades and centuries.

2. The science of weather forecasting

Dynamical and physical processes within the atmosphere, and interactions with the surroundings (e.g. land, ocean, and ice surfaces), determine the evolution of the atmosphere and, hence, the weather. Scientifically-based weather forecasts are possible if the processes are well enough understood and if the current state of the atmosphere is well known enough, for predictions to be made of future states. Weather forecasts are prepared using a largely systematic approach, involving observation and data assimilation, process understanding, prediction and dissemination. Each of these components has, and will continue, to benefit from advances in science and technology.

2.1 Observations and data assimilation

2.1.1 Over the past few decades, substantial advances in science have resulted in improved and more efficient methods for making and collecting timely observations, from a wide variety of sources including radar and satellites. Using these observations in scientifically-based methods has caused the quality of weather forecasts to increase dramatically, so that people around the world have come to rely on weather forecasts as a valued input to many decision-making processes.

2.1.2 Computer-generated predictions are initialized from a description of the atmospheric state built from past and current observations in a process called data assimilation, which uses the NWP model (see paragraph 2.3.2) to summarize and carry forward in time information from past observations. Data assimilation is very effective at using the incomplete coverage of observations from various sources to build a coherent estimate of the atmospheric state. But, like the forecast, it relies on the NWP model and cannot easily use observations of scales and processes not represented by the model.

2.1.3 The international scientific community is emphasizing the still very poorly observed areas as being a limiting factor in the quality of some forecasts. As a con-sequence, there is a continued need for improved observation systems and methods to assimilate these into NWP models.

2.2 Understanding of the atmosphere: inherent limitations to predictability

2.2.1 The scientific understanding of physical processes has made considerable progress through a variety of research activities, including field experiments, theoretical work and numerical simulation. However, atmospheric processes are inherently non-linear and not all physical processes can be understood or represented in NWP models. For instance, the wide variety of possible cloud water and ice particles must be highly simplified, as are small cumulus clouds that can lead to rain showers. Continued research effort using expected improvements in computer technology and physical measurements will enable these approximations to be improved. Even then, it will still not be possible to represent all atmospheric motions and processes.

2.2.2 There is a wide spectrum of patterns of atmospheric motion, from the planetary scale down to local turbulence. Some are unstable and are arranged so that flow is amplified using, for example, energy from heating and condensation of moisture. This property of the atmosphere means that small uncertainties about the state of the atmosphere will also grow, so that eventually the unstable patterns cannot be precisely forecast. How quickly this happens depends on the type and size of the motion. For convective motions such as thunderstorms, the limit is of the order of hours, while for large scales of motion it is of the order of two weeks.

2.3 Weather prediction

2.3.1 *Nowcastin*g: Forecasts extending from 0 out to 6 to 12 hours are based upon a more observations-intensive approach and are referred to as nowcasts. Traditionally, nowcasting has focused on the analysis and extrapolation of observed meteorological fields, with a special emphasis on mesoscale fields of clouds and precipitation derived from satellite and radar. Nowcast products are especially valuable in the case of small-scale hazardous weather phenomena associated with severe convection and intense cyclones. In the case of tropical cyclones, now-casting is an important detection and subsequent short-term prediction approach that provides forecast value beyond 24 hours in some cases. However, the time rate of change of phenomena such as severe convection is such that the simple extrapolation of significant features leads to a product that deteriorates rapidly with time - even on timescales of the order of one hour. Thus, methods are being developed that combine extrapolation techniques with NWP, both through a blending of the two products and through the improved assimilation of detailed mesoscale observations. These are inherently difficult tasks and, although accuracy and specificity will improve over coming years, these products will always involve uncertainty regarding the specific location, timing and severity of weather events such as thunder and hail storms, tornadoes and downbursts.

2.3.2 *Numerical weather predictio*n: Forecasts for lead times in excess of several hours are essentially based almost entirely on NWP. In fact, much of the improvement in the skill of weather forecasts over the past 20 years can be attributed to NWP computer models, which are constructed using the equations governing the dynamical and physical evolution of the atmosphere. NWP models represent the atmosphere on a three-dimensional grid, while typical operational systems in 2001 use a horizontal spacing of 50–100 km for large-scale forecasting and five to 40 km for limited area forecasting at the mesoscale. This will improve as more powerful computers become available. Only weather systems with a size several times the grid spacing can be accurately predicted, so phenomena on smaller scales must be represented in an approximate way using statistical and other techniques. These limitations in NWP models particularly affect detailed forecasts of local weather elements, such as cloud and fog and extremes such as intense precipitation and peak gusts. They also contribute to the uncertainties that can grow chaotically, ultimately limiting predictability.

2.3.3 *Ensemble predictio*n: Uncertainty always exists - even in our knowledge of the current state of the atmosphere. It grows chaotically in time, with much of the new information introduced at the beginning no longer adding value, until only climatological information remains. The rate of growth of this uncertainty is difficult to estimate since it depends upon the three-dimensional structure of the atmospheric flow. The solution is to execute a group of forecasts – an ensemble - from a range of modestly different initial conditions and/or a collection of NWP models with different, but equally plausible, approximations. If the ensemble is well designed, its forecasts will span the range of likely outcomes, providing a range of patterns where uncertainties may grow. From this set of forecasts, information on probabilities can be derived automatically, tailored to users' needs. Forecast ensembles are subject to the limitations of NWP discussed earlier. Additionally, since the group of forecasts is being computed simultaneously, less computer power is available for each forecast. This requires grid spacings to be increased, making it more difficult to represent some severe weather events of smaller horizontal scale. Together with the limited number of forecasts in an ensemble, this makes it harder to estimate probabilities of very extreme and rare events directly from the ensemble. Moreover it is not possible to modify the NWP models used to sample properly modelling errors, so sometimes all models will make similar errors.

2.3.4 *Operational meteorologis*t: There remains a critical role for the human forecaster in interpreting the output and in reconciling sometimes seemingly conflicting information from different sources. This role is especially important in situations of locally severe weather. Although vigorous efforts are being made to provide forecasters with good quality systems such as interactive workstations for displaying and manipulating the basic information, they still have to cope with vast amounts of information and make judgements within severe time constraints. Furthermore, forecasters are challenged to keep up to date with the latest scientific advances.

3. Prediction at seasonal to interannual timescales

3.1 Beyond two weeks, weekly average predictions of detailed weather have very low skill, but forecasts of one-month averages, using NWP with predicted sea-surface temperature anomalies, still have significant skill for some regions and seasons to a range of a few months.

3.2 At the seasonal timescale, detailed forecasts of weather events or sequences of weather patterns are not possible. As mentioned above, the chaotic nature of the atmosphere sets a fundamental limit of the order of two weeks for such deterministic predictions, associated with the rapid growth of initial condition errors arising from imperfect and incomplete observations. Nonetheless, in a limited sense, some predictability of temperature and precipitation anomalies has been shown to exist at longer lead times out to a few seasons. This comes about because of interactions between the atmosphere, the oceans, and the land surface, which become important at seasonal timescales.

3.3 The intrinsic timescales of variability for both the land surface and the oceans are long compared to that of the atmosphere, due in part to relatively large thermal inertia. Ocean waves and currents are slow in comparison to their atmospheric counterparts, due to the large differences in density structure. To the extent that the atmosphere is connected to the ocean and land surface conditions, then, a degree of predictability may be imparted to the atmosphere at seasonal timescales. Such coupling is known to exist particularly in the tropics, where patterns of atmospheric convection ultimately important to global scale weather patterns are quite closely tied to variations in ocean surface temperature. The most important example of this coupling is found in the ENSO phenomenon, which produces large swings in global climate at intervals ranging from two to seven years.

3.4 The nature of the predictability at seasonal timescales must be understood in probabilistic terms. It is not the exact sequence of weather that has predictability at long lead times (a season or more), but rather some aspects of the statistics of the weather – for example, the mean or variance of temperature/precipitation over a season - that has potential predictability. Though the weather on any given day is entirely uncertain at long lead times, the persistent influence of the slowly evolving surface conditions may change the odds for a particular type of weather occurring on that day. In rough analogy to the process of throwing dice, the subtle but systematic influence of the boundary forcing can be likened to throwing dice that are “loaded”. On any given throw, we cannot foretell the outcome, yet after many throws the biased dice will favour a particular outcome over others. This is the sort of limited predictability that characterizes seasonal prediction.

3.5 Currently, seasonal predictions are made using both statistical schemes and dynamical models. The statistical approach seeks to find recurring patterns in climate associated with a predictor field such as sea-surface temperature. Such models have demonstrated skill in forecasting *El Niño* and some of its global climate impacts. The basic tools for dynamical prediction are coupled models - models that include both the atmosphere and the other media of importance, particularly the oceans. Such models are initialized using available observations and integrated forward in time to produce a seasonal prediction. The issue of uncertainty is handled using an ensemble approach, where the climate model is run many times with slightly different initial conditions (within the range of observation errors or sampling errors). From this, a distribution of results is obtained, whereupon statistics of the climate can be estimated. Recently, encouraging results have been obtained from ensemble outputs of more than one model being combined.

3.6 There are several limitations attending current predictions. Most coupled models (and to a lesser extent uncoupled models) exhibit some serious systematic errors that inevitably reduce forecast skill. Data availability is a limitation for both statistical models and for dynamical models. In the latter case, very limited information is available for much of the global oceans and for the land surface conditions. Also, current initialization methods do not account properly for systematic model errors, further limiting forecast performance. A final set of limitations arises for practical reasons. Due to resource requirements, most seasonal predictions cannot be done at resolutions comparable to weather prediction. Furthermore, rather small ensemble sizes (of the order of 10) are used for some models, certainly less than is optimal for generating robust probabilistic forecasts. Current research is addressing the potential for regional “downscaling” of climate forecasts by various means and the possibilities for more detailed probabilistic climate information from expanded ensembles of one or more models.

3.7 Possible use of seasonal forecasts is currently being explored in various contexts. In each case, effective use will require careful attention to the issue of uncertainty inherent in seasonal forecasts. Future advancements can be expected to improve the estimates of uncertainty associated with forecasts, thus allowing better use of forecast products.

4. Projection of future climate

4.1 As explained above, based on the current observed state of the atmosphere, weather prediction can provide detailed location and time-specific weather information on timescales of the order of two weeks. Some predictability of temperature and precipitation anomalies has been shown to exist at longer lead times out to a few seasons. This comes about because of inter-actions between the atmosphere, the oceans, and the land surface, which become important at seasonal timescales. At longer timescales, the current observed state of the atmosphere and even those large-scale anomalies which provide predictive skill at seasonal to interannual timescales are no longer able to do so due to the fundamental chaotic nature of the Earth-atmosphere system. However, long-term changes in the Earth-atmosphere system at climate timescales (decades to centuries) are dependent on factors which change the balance of incoming and outgoing energy in the Earth-atmosphere system. These factors can be natural (e.g. changes in solar output or volcanoes) or human induced (e.g. increased greenhouse gases). Because simulations of possible future climate states are dependent on prescribed scenarios of these factors they are more accurately referred to as “projections” not “predictions” or “forecasts”.

4.2 In order to perform climate projections, physically-based climate models are required in order to represent the delicate feedbacks which are crucial on climate timescales. Physical processes and feedbacks that are not important at NWP or even at the timescales of seasonal prediction become crucial when attempting to simulate climate over long periods, e.g. cloud-radiation interaction and feedback, water vapour feedback (and correctly modelling long-term trends in water vapour), ocean dynamics and processes (in particular an accurate representation of the thermohaline circulation). The treatments of these key features are adequate to reproduce many aspects of climate realistically though there remain many uncertainties associated with clouds and aerosols and their radiative effects, and many ocean processes. Nevertheless, there is reasonable confidence that state-of-the-art climate models do provide useful projections of future climate change. This confidence is based on the demonstrated performance of models on a range of space timescales.

4.3 Notably, the understanding of key climate processes and their representation in models (such as the inclusion of sea-ice dynamics and more realistic ocean heat transport) has improved in the past few years. Many models now give satisfactory simulations of climate without the need for non-physical adjustments of heat and water fluxes at the ocean-atmosphere interface used in earlier models. Moreover, simulations that include estimates of natural and anthropogenic forcing are well able to reproduce observed large-scale changes in surface temperature over the twentieth century. This large-scale consistency between models and observations lends confidence in the estimates of warming rates projected over the next century. The simulations of observed natural variability (e.g. ENSO, monsoon circulations, the North Atlantic Oscillation) have also improved.

4.4 On the other hand, systematic errors are still all too apparent, e.g. in simulated temperature distributions in different regions of the world or in different parts of the atmosphere, in precipitation fields, clouds (in particular marine stratus). One of the factors that limits confidence in climate projections is the uncertainties in external forcing (e.g. in predicting future atmospheric concentrations of carbon dioxide and other greenhouse gases, and aerosol loadings).

4.5 As with NWP and seasonal forecasts, ensembles of climate projections are also extremely important. Ensembles enable the magnitude and effects of natural climate variability to be gauged and affect its impact on future projections, and thereby permit any significant climate change signal to be picked out more clearly statistically (the magnitude of natural climate variability will be comparable with that of climate change for the next few decades).

5. Dissemination to end-users

5.1 The weather forecasts have to be communicated to a vast array of users such as emergency managers, air traffic controllers, flood forecasters, public event managers, etc. in a timely and user-applicable form. This in itself poses another major challenge that is increasingly benefiting from advances in information technology. Predictions at seasonal to interannual timescales and climate projections are also being used by an increasingly wide range of users.

5.2 The value of forecasts to decision makers is greatly enhanced if the inherent uncertainty can be quantified. This is particularly true of severe weather, which can cause such damage to property and loss of life that precautions may be well advised even if the event is unlikely, but possible. Probabilities are a natural way of expressing uncertainty. A range of possible outcomes can be described with associated probabilities and users can then make informed decisions allowing for their particular costs and risks.

5.3 Forecasts expressed as probabilities, or ensembles, contain much more information than deterministic forecasts, and it is difficult to convey it all to users. Broadcast forecasts can only give a broad picture of the most likely outcome, with perhaps some idea of important risks. Each user's decision may be based on the probabilities of a few specific occurrences. What these are, and the probability thresholds for acting on the forecasts, will differ. So for important user decisions it is necessary to apply their particular criteria to the detailed forecast information.

6. Conclusions

6.1 The skill in weather forecasting has advanced substantially since the middle of the twentieth century, largely supported by the advancement of computing, observation and telecommunications systems, along with the development of NWP models and the associated data-assimilation techniques. This has been greatly facilitated because of the vast experience of both forecasters and decision makers in producing and in using forecast products. Nevertheless, each component within the science and technology of weather forecasting and climate projection has its own uncertainties. Some of these are associated with a lack of a complete understanding of, or an inherent limitation of, the predictability of highly complex processes. Others are linked still to the need for further advances in observing or computing technology, or to an inadequate transfer between research and operations. Finally, one cannot underestimate the importance of properly communicated weather forecasts to well-educated users.

6.2 Without a doubt, significant benefits will result from continued attention to scientific research and the transfer of knowledge gained from this work into the practice of forecasting. Furthermore, a recognition of the limitations of weather forecasts and climate projections, and when possible, an estimate of the degree of uncertainty, will result in the improved use of forecasts and other weather information by decision makers. Ultimately the objective is for the scientific and user communities to work better together, realizing even greater benefits.

APPENDIX C

The WGNE survey of verification methods for numerical prediction of weather elements and severe weather events

Philippe Bougeault
Météo-France, Toulouse

January 2003

1. Introduction

In its 13th session, the WMO/Commission for Atmospheric Science tasked WGNE to prepare a position paper on high-resolution model verifications, oriented towards weather elements and severe weather events (item 5.3.10 of the abridged final report, document WMO-N°941). This recognizes the specific difficulty of traditional verification methods in providing a useful measure of model performance at high resolution and for intense events. First, the verification of mesoscale events is limited by the insufficient density and quality of the observing networks. Second, the related weather elements may be on the edge of predictability, or entirely stochastic from the perspective of current NWP models. As such, the traditional verification methods based on instantaneous comparison of analyzed and predicted fields may not yield useful information, and new methods are needed. Third, there is a great expectation that mesoscale models will deliver products of direct relevance to end- users, and consequently much work is done on the development of user-oriented verifications, but the needs are not the same for user-oriented and developers-oriented verifications.

The verification of numerical models against observations has several purposes. For instance: (i) provide a measure of the progress of the forecast skill over the years; (ii) compare the merits of two versions of a forecasting system in order to decide which is the best for operations; (iii) understand where the problems are and what aspects of the system need refinements; (iv) compare the relative value of two different systems for a specific category of users. No single verification system can be optimal for all of these tasks and there is a need to issue guidance on what methods are good for what purpose. The purpose of the present paper is to report on a survey of methods currently in use or under development in many operational NWP centers, and to provide guidance on desirable features for verification methods, based on shared experience.

The organization of the paper is as follows: Section 2 is a list of the available sources and recent discussions. Section 3 summarizes the logical process of verification and discusses some “recommended” methods, depending on a range of issues. Section 4 focuses on the topic of severe weather. Finally Section 5 summarizes the replies of various centers to the survey.

2. A short review of available sources

The subject of verification is a very active area. The most common methods are presented by Stanski et al. (1989). A quick overview of recent developments can be obtained from the Internet site on Verification Methods maintained by E. Ebert at BMRC, see http://www.bom.gov.au/bmrc/wefor/staff/eee/verif/verif_web_page.html. A detailed glossary is available at http://www.sel.noaa.gov/forecast_verification/verif_glossary.html#catfcst. An early discussion of verification techniques for high resolution models and related problems can be found in Anthes (1983). Some general concepts are discussed by Murphy (1991, 1993) and Murphy and Winkler (1987). A classic book on statistical methods is Wilks (1995). The subject was discussed in 1998 at a NCAR Workshop on Mesoscale Model Verification (Davis and Carr, 2000). A very recent discussion is given by Mass et al. (2002) in the context of the evaluation of a mesoscale model over the Pacific Northwest. Under the auspices of WGNE, a systematic inter-comparison of model precipitation forecasts against high-resolution rain gauges (and sometimes radars) is now conducted in several centers (Ebert et al., 2002). These papers also contain some discussions of the best approach to verification at the mesoscale.

Verification methods at high-resolution are currently a subject of debate, with many on-going meetings. Here are a few recent examples: The European COST717 action on the use of radars in NWP has published a review of current methods and tools for verification of numerical forecasts of precipitation (written by C. Wilson, MO). This is available on http://www.smhi.se/cost717 . The European Short Range Network on Numerical Weather Prediction held a workshop in De Bilt in April 2001. Their discussion on verification methods can be found on http://srnwp.cscs.ch/leadcenters/ReportVerification.html The World Weather

Research Program (WWRP) organized a workshop on QPF verification methods (Prag, May 2001). The report may be found on http://www.chmi.cz/meteo/ov/wmo . Another workshop devoted to the definition of more meaningful methods took place at NCAR in August 2002, see http://www.rap.ucar.edu/research/verification/ver_wkshp1.html . The WWRP Conference on Quantitative Precipitation Forecasting (Reading, September 2002) also had a session on verification methods. A summary of the session can be found on http://www.royal-met-soc.org.uk/qpfproc.html . The WWRP also recently initiated a Working Group on Verification Methods.

A general consensus of these discussions seems to be: (i) new methods are needed to deal with the verification of mesoscale models; (ii) the international exchange of observations need to be enhanced; (iii) the intercomparison of model scores can be useful, but only if it is done with great care.

3. Methodology of Verification

The logical process of verification against observations can be divided in five steps: (i) the choice of a set of observations for verification; (ii) the technique to compare a single model forecast to a single observation; (iii) the aggregation of model/observation pairs in ensembles of a convenient size; (iv) the use of statistics to condense the information contained in the joint distribution of model/observation pairs (v) the use of additional information to help interpret the scores, in particular their statistical significance.

3.1 Observations available for verification of weather elements

The most commonly used observations for verification of weather elements are surface precipitation from rain gauges. The accumulation period is quite variable, from a few minutes to 24 hours. The use of the shorter accumulation periods should be encouraged for high-resolution models, with a view of matching the accumulation period to the model resolved time scales. Surface air temperatures, humidity and winds are also widely used. Cloud cover reports from surface stations are sometimes cited. The use of more advanced observation systems, such as meteorological radars and satellite cloud cover, is incipient and should be encouraged, although they are posing an obvious problem of accuracy, especially in mountainous areas. The use of a standard Z-R relationships for radar data is insufficient for heavy rainfall because of attenuation. The observation uncertainty should always be kept in mind when building a verification system. A few centers are developing verifications of other weather elements: Hail is reported in Synop observations, and specific detection networks exist in some parts of the world. Visibility is a subject of much interest, and reliable measurements are now available. Wind gusts are also commonly measured and predicted, and so deserve a specific verification. Ground skin temperature can be measured by satellite and is predicted by models, it should therefore also be verified.

3.2 Controlling the quality of the observations

This is a key step in the whole process. Most modern NWP systems have adopted a double quality-check procedure. In a first step, observations are checked for gross errors (unit problems, unphysical values, internal lack of consistency). Then they are compared to the model (see next subsection) and in case of a large differences between a model-derived value and the actual observation, other observations close-by are checked to ascertain whether the suspicious value is isolated, in which case it is discarded. This involves a considerable degree of empiricism, and could be at the origin of large differences in the results of various verification packages. There is a need for international exchange and comparison of the procedures involved in the quality control of observations, with due regards to differences inherent to the diversity of observing networks. The quality control methods might also be different for various verification purposes. For instance, an observation unrepresentative of the scale resolved by a model could be discarded as part of the quality control procedure when the verification is oriented towards model assessment, while it should be retained when the verification is user-oriented. This problem is even more important for high-resolution models.

3.3 Comparing the model with the observations

The way in which forecasts and observations are matched becomes more important for mesoscale verification because of the sampling limitations of both observations and forecasts for small-scale structures and processes. The best strategy obviously depends on the density and quality of the observing network, the resolution of the model, the type of observation considered, etc… This is highly variable around the world, so it is no surprise that meteorologists facing different situations in different countries have developed a large variety of methods, and sometime even vocabulary. Point observations contain information on all space and time scales, but usually drastically under-sample finer space and time scales. It is often considered preferable to treat the observations as estimates of area or time averages rather than to carry out an analysis of under-sampled fields. Such analyses artificially treat the point observations as if they contain information on only those scales which can be represented by the grid on which the analysis is done. Analysis of observations has the effect of eliminating from the verification the component of the error due to the inability of the model to represent scales smaller than its grid allows.

When the resolution of the observing network is larger than the model, the observations should clearly be up-scaled to the model resolution. A simple and efficient technique has recently been described by Cherubini et al. (2001) in the context of ECMWF model 24h accumulated rainfall verification: the climate observing network for 24h rainfall is significantly denser over most of Europe than the ECMWF model grid. It is therefore adequate to compute the arithmetic average of all the climate stations falling inside each model grid box. This more representative “super-observation” is then compared to the model grid value. This dramatically improves the model performance (especially the FBI and ETS scores at threshold 0.1 mm), and shows that the previous comparisons to the closest SYNOP rain observation were misleading.

A more common case is, however, when the model resolution is higher than the observing network. This will be true for most meso-scale models and weather parameters. One simple technique is then to interpolate the model prediction to the location of the observation, but this has the effect of smoothing the model result, and could result in a biased interpretation of its capacity to deal with extreme events. A common technique is to use the value at the nearest grid point to the observation location, ignoring the corresponding error on location.

Observations may not always be representative for the average model grid box (in fact they rarely are). Various representativity problems are due to the ground altitude (for temperature), to exposure effects (for wind and rain), to land cover heterogeneity (for temperature and humidity). Most centers use a standard vertical gradient of temperature to correct for altitude differences. Some schemes have been developed to correct wind forecasts for exposure effects and rain observations for altitude effects. With the rapid development of surface schemes using ‘tiles’, it may become possible to compare an observed temperature with one of several temperatures within the grid box (the one corresponding to the model land cover type matching the observation best). This may generate a need to have additional meta-data attached to the surface observations, indicating what is the immediate environment of the observation station (e.g. crops, lake, forest, urban, etc…).

The computation of the ‘model equivalent’ to the observations for verification purposes shares many aspects with the computation of the ‘observation operators’ in the variational data assimilation techniques. The development of common software for these two aspects of the NWP suite is encouraged, for instance for the radar and satellite observations. Furthermore, the differences between observations and model, in observation space, is already computed to evaluate the cost function which is minimized in variational procedures. These computations may not need to be done again for model verification (Davis and Carr, 2000). However, differences in the set of considered observations or in the detail of these computations may become necessary for user-oriented and model-oriented verifications.

3.4 How to aggregate/stratify the results?

There is a need to find a trade-off between various constraints: ensembles of forecast/observation pairs should be large enough to carry a good statistical significance, but small enough to distinguish between various areas or time periods prone to different types of errors (eg various climate, or altitudes). Stratifying results by time of the day will allow one to spot errors on the diurnal cycle of temperature and other variables, presumably linked to deficiencies in the surface energy budget parameterization, or the soil humidity. Stratifying by lead time tells about how fast the model is deriving from the truth. Stratifying by the values of the observed parameter shows how the model performance degrades towards extreme values. Stratifying by geographical area, or altitude above sea level, helps to point out the relations between model errors and the terrain. Finally, the available manpower to inspect the results will usually set a practical upper limitation to the number of scores. It is impossible to know in advance what combination of parameters will be needed to solve rapidly any new problem, so it is advisable to store individual values in a relational database for the purpose of quickly forming new combinations. This approach is now used in several centers.

The full examination of the joint distribution of the forecasts/observations pairs is a powerful way to acquire a detailed understanding of the characteristics of a forecast system (Murphy and Winkler, 1987). The bi-variate distribution p(f,o) can be factorized in marginal distributions for observations p(o), forecasts p(f), and the conditional distribution of observations given the forecasts p(o|f) or forecasts given the observation p(f|o). An approach used at the Met Office is to look at the distribution of observations for given forecast events. This can be interpreted as the probability distribution of observations given a specific forecast. For a perfectly accurate NWP model, we would expect to observe a parameter in a given interval on every occasion when the forecast is it in that interval. A recent example of a distribution-oriented analysis of forecasts/observations pairs is provided by de Elia and Laprise (2002) (though they used only virtual observations supplied by a reference model run). They point to the fact that even for a globally unbiased forecast, the conditional bias (the bias of the forecast for a given value of the observed parameter) is in all cases towards the mean of the marginal forecast distribution. This should not be interpreted as an indication that the model is under-predicting. In fact, the conditional bias of the observations for a given forecast value is also towards the mean of the marginal observation distribution. This behavior is known as Galton's law in statistics.

3.5 Scoring deterministic forecasts

In practice, the bi-variate distribution often carries too much information and must be condensed by use of statistics. A large variety of statistical scores has been described in the literature, each of them having advantages and shortcomings. No single score can convey the full information, but it is often believed that a combination of a small number of well-chosen scores can provide a reasonable assessment of most model error distributions.

The definitions and main properties of the most common scores are explained e.g. on http://www.bom.gov.au/bmrc/wefor/staff/eee/verif/verif_web_page.html . Here is a short summary. Continuous statistics allow to measure how the values of forecasts variables differ in average from the values of observed variables. The mean error, or bias, is a useful basic information, but it does not measure the magnitude of the errors. The mean absolute error, the root mean square error, or the mean squared error all measure the average magnitude of the errors, with different weights of the largest errors. The anomaly correlation measures the correspondence or phase difference between the forecasts and observations without being sensitive to their absolute value. Categorical statistics are more appropriate to evaluate yes/no forecasts. They are often used to evaluate the capacity of models to predict that weather parameters will exceed a given threshold. A contingency table is constructed to count the correct predictions of observed events (hits), their non-prediction (misses), the prediction of a non-observed event (false alarms) and the correct prediction of non-observed events (correct negatives). These quantities are combined in various categorical statistics. The Accuracy (ACC= hits + correct negative divided by total) measure the fraction of all forecasts that were correct. It can be misleading because it is heavily influenced by the most common category, usually the "no event" in the case of weather. The Frequency Bias Index (FBI) measures the ratio of the frequency of forecast events (hits + false alarms) to the frequency of observed events (hits + misses). It indicates whether the forecast system has a tendency to underforecast (FBI<1) or overforecast (FBI>1) events. It does not measure how well the forecast corresponds to the observations, only relative frequencies. The Probability of Detection (POD= hits/ hits + misses) measures the fraction of observed events that were correctly forecast. It is sensitive to hits, good for rare events, but ignore false alarms. It can be artificially increased by issuing more "yes" forecasts to improve the number of hits. The False Alarm Ratio (FAR= false alarms/hits + false alarms) measures the fraction of "yes" forecasts in which the event did not occur. It ignores misses and can be artificially improved by issuing more "no" forecasts to reduce the number of false alarms. The Threat Score (or Critical Success Index) (TS= hits/ hits + misses + false alarms) measures the fraction of observed and/or forecast events that were correctly forecast. It is sensitive to hits, but penalizes both misses and false alarms. However, it does not distinguish the source of forecast error, and is sensitive to the frequency of events, since some hits can occur due to random chance. Thus in general, the Threat Score will be higher for a sequence of unusually numerous events, and this should not be interpreted as an indication that the forecasting system is becoming better. In order to correct for this effect, the Equitable Threat Score (ETS) measures the fraction of observed and/or forecast events that were correctly predicted, adjusted for hits associated with random chance in the forecast (ETS=(hits – hits(random))/ hits + misses + false alarms – hits(random), where hits(random) = (hits + misses)x(hits + false alarms)/total). This score is often used in the verification of rainfall forecasts because its "equitability" allows scores to be compared more fairly across different precipitation regimes. Along the same ideas, the Heidke Skill Score (HSS) measures the fraction of correct forecasts after eliminating those forecasts due purely to random chance. It measures the improvement over random chance. However, random chance is usually not the best forecast to compare to, and the HSS is sometimes computed with respect to climate or to persistence. More recently, the merits of the Odds Ratio (OR=hits * correct negative / misses * false alarms) have been argued (Stephenson, 2000; Goeber and Milton, 2002). The OR measures the ratio of the probability of making a hit to the probability of making a false alarm. It is appropriate for rare events, does not depend on marginal totals, and is therefore "equitable". It can easily be used to test whether the forecast skill is significant.

Multi-category forecasts can also be verified by building multi-category contingency tables. Scores can then be defined to quantify the degree of fit between the distributions of forecasts and verifying observations. The Accuracy and Heidke Skill Score are two examples of scores that can be easily generalized to account for multi-category forecasts.

Specific user-oriented scores are easily developed based on the above principles. Very often the main interests of users can be summarized in the two following questions: what is the probability that an even will occur when it is forecast? What is the probability that an event has been forecast when it occurs?

3.6 The double penalty problem

It is common observation that the objective scores for weather parameters can be worse for high resolution models than for low resolution models. Indeed, increased resolution generally produces better defined mesoscale structures, greater amplitude features and larger gradients. Thus, inevitable space and timing errors for weather-related parameters will lead to a larger RMSE than the smoother forecasts of a low resolution model. This is generally known as the 'double penalty' problem (see e.g. Anthès, 1983, or Mass et al., 2002). At the same time, there is a consensus that high-resolution numerical predictions are very useful to forecasters, even with small space and timing errors, because they point to the possibility of some important weather patterns happening in a given area, and because they convey some explanation of why and how this may happen (a conceptual model). A classical example is the forecast of isolated thunderstorms, where models are not expected to provide a very accurate location, but can be very informative regarding timing and severity. The need for verification techniques that allow for some tolerance to reasonably small space and time errors is universally recognized and central to much of the recent literature on the subject. One approach is to average the output of the high-resolution model to a lower resolution before applying the deterministic scores (this is sometimes called "hedging"). This may reveal the superiority of high-resolution models over low-resolution models, while direct comparison of model outputs interpolated to the station point would in general give a more favorable result for the low resolution model (Damrath, personal communication). However, smoothing model outputs will in general deteriorate their intrinsic behavior, such as forecast variance, spectrum of energy, and the frequency of intense events. This detrimental effect can be measured by other indicators, such as the Frequency Bias Score (see definition below). In general, it is recommended to consider several indicators to assess the quality of a model (e.g. the forecast variance should be close to the observed variance, the forecast bias should be very small, and the root mean square error should be reasonably small). An early paper on the usefulness of Control Statistics to avoid "playing the scores" is Glahn (1976).

Other approaches to circumvent the double penalty are reviewed by Davis and Carr (2000). Brooks et al. (1998) compute the probability density distribution associated with local severe weather reports on a single day, and evaluate the maximum skill of a forecast based on simple spatial averaging. This turns out to be fairly low (a CSI of 0.24 in his example). Thus, a hypothetical numerical forecast having a CSI of 0.09, despite being rather low in absolute value, represents 38% of the upper bound, and must be considered as relatively successful forecast. A most simple method used by de Elia and Laprise (2002) consists in allowing for a tolerance of one grid point to find the best match between the forecast and the observation. A more elaborated version of the procedure is to consider that all grid points within a given distance of a point of interest are equally likely forecasts of an event at this point. Thus, a probability of some threshold being exceeded at this point can be computed as the ratio of the number of neighboring grid points where it happens over the total number of grid points considered. The size of the area for these counts is subject to optimization. This probabilistic forecast must then be evaluated through appropriate scoring. An example of this approach is discussed by Atger (2001).

3.7 Scoring probabilistic forecasts

A good probability forecast system has three attributes: (i) Reliability is the agreement between the forecast probability and the mean observed frequency; (ii) Sharpness is the capacity of the system to forecast probabilities close to 0 or 1; (iii) Resolution is the ability of the system to resolve the set of sample events into subsets with characteristically different frequencies. Sharpness and resolution are somewhat redundant, and become identical when reliability is perfect. The most common measure of the quality of probabilistic forecasts is the Brier Score (Brier, 1950). It measures the mean squared probability error, and ranges from 0 to 1, with perfect score 0. Murphy (1973) showed that the Brier Score can be partitioned into three terms accounting respectively for reliability, resolution, and uncertainty. The Brier Score is sensitive to the frequency of the event: the more rare the event, the easier it is to get a good BS without having any real skill. The Ranked Probability Score (RPS) measures the sum of squared differences in cumulative probability space for a multi-category probabilistic forecast. It penalizes forecasts more severely when their probabilities are further from actual outcome. As the BS, it ranges from 0 to 1, with perfect score 0.

Reliability is specifically measured by reliability diagrams, where the observed frequency is plotted against forecast probability, divided into a certain number of groups (Wilks, 1995). Perfect reliability is achieved when the results are aligned along the diagonal of the diagram. A shortcoming of reliability diagrams is that one needs a large number of forecasts to generate a meaningful diagram. An alternative approach known as the multi-category reliability diagram (Hamill, 1997) allows to accumulate statistics from a reduced number of forecasts. In the case of ensemble forecasts, an additional useful evaluation is given by the Rank Histograms (also called Talagrand diagrams). The rank is the position of the verifying observation relative to the ranked ensemble forecast values. For a reliable ensemble, the Rank Histogram should be approximately uniform, meaning that an observation is equally likely to occur near any ensemble member.

Probabilistic forecasts can be tailored to the use of any specific user category, by adjusting the probability threshold required to make a yes/no decision. Of course, this will induce a simultaneous change of the POD and of the FAR. An increase of POD will be achieved at the cost of an increase in FAR. Diagrams showing how the POD and FAR rate change with the decision criteria are called Relative Operating Characteristics curves (ROC). They describe how a forecast system can meet simultaneously the needs of various users categories, and therefore contain a lot of information. In contrast, a deterministic forecast system will be represented by a single point on such a diagram. It is expected that the curve describing the probabilistic system results pass above the point describing the deterministic system, showing the superiority of the probabilistic approach. The area under the ROC curve is frequently used as a global indicator of the quality of a probabilistic forecast system. However it tells nothing about reliability. Another increasingly used measure of the quality of probabilistic forecasts is the Potential Economic Value, which conveys about the same information as the ROC curve, translated in potential gain for any category of users, stratified by their Cost/Loss parameter (e.g. Richardson, 2000).

3.8 Additional information necessary to interpret the scores

An essential information is the uncertainty associated with the above statistics. A related question is the statistical significance of the comparison between two forecasting systems on a given series of weather events. This is especially important for severe weather, since the number of events is often small. Hamill (1999) discusses a number of limitations of common hypothesis tests in weather forecast verification, such as spatial correlations and non-normality of errors. He proposes new methods such as re-sampling techniques, that allow to evaluate the uncertainty associated to statistical scores such as the widely used ETS. Similar techniques are also applicable to the probabilistic scores. Atger (2001) has applied this method in the context of QPF. The sample of events was randomly halved into two sub-samples, and the score differences between the two sub-samples were evaluated. The process was repeated a large number of times and resulted in an evaluation of the uncertainty in the scores.

Another recommended point of comparison is with straightforward forecasting techniques, such as climate, persistence, or chance. This is embodied in a number of the above-mentioned scores. Finally, it is considered that computing scores on the verifying analysis (or on the model initial state) is a good point of comparison.

3.9 Research in verification methods

The development of new verification methods is an active area of research. Most of the methods discussed above will tell little about exactly what the error is, or why there is an error. Therefore recent efforts are directed towards the development of methods that could help the modelers to improve models. The need to identify the spatial scales involved in a given error was already mentioned by Anthès (1983). Scale separation techniques are being developed, base e.g. on wavelets (Briggs and Levine, 1997). The objective is to identify at what scales the greatest error is occurring, and whether the model resolves all of the scales that can be measured in the observations. Zepeda-Arce et al. (2000) propose a method consisting in upscaling from fine to coarse resolution by simple averaging, and computing verification scores as a function of both threshold and resolution. If the scores improve very quickly towards coarser resolution, there is an indication that the forecast is good. Fuzzy verification techniques under developments at BMRC and DWD try to deal with uncertainty in both forecasts and observations. Finally, the examination of model energy spectra and their evolution over time has often been recommended to verify the realism of simulations.

The use of object-oriented techniques is also developing rapidly. This is making sense when it is possible to associate unambiguously an observed weather object with its forecasted counterpart. A most classical application of this is the verification of the skill in forecasting the track of tropical cyclones. The score is based on the distance between the observed and forecasted tracks of the cyclone center, assuming that the association between the observed and forecasted cyclone is a simple issue. Hoffman et al. (1995) have proposed a generalization of this approach to other types of events, and Ebert and McBride (2000) have implemented a similar system, the Contiguous Rain Area method (CRA), now routinely used at BMRC, Australia. They show that the total RMSE of a precipitation forecast can be decomposed into three components, describing respectively a displacement error, a volume error and a pattern error. A systematic evaluation of these three components of error on a long period helps in understanding what the problems of the model are. It also allows to define the 'hits' and 'false alarms' cases with a certain tolerance, consistent with the forecasters opinion of a useful forecast. It should be noted that full application of this technique is only possible when the forecast and observed rain systems are completely enclosed within the verification domain. Moreover, application to local storms would probably be hampered by the difficulty of associating unambiguously an observed and a forecast event without a human intervention, except for the strongest cases.

4. The severe weather problem

Severe weather poses a special problem because it is unfrequent, poorly documented by observations, and at the limit of predictability. Quantitative verifications are therefore more difficult and their statistical significance is always poor. At the same time, it is recognized that a poor numerical forecast in absolute terms can be of great value if it is well interpreted by an experienced forecaster. This may be seen as an extreme example of the "double penalty" problem. In addition of a tolerance on space and time, a tolerance on the value of weather-related parameters must often be accepted in the case of extreme values. For instance, in a region where a daily accumulated precipitation larger than 200 mm is a rare event, a 200 mm forecast represents a bad forecast if the observed value is more frequent (say, 50 mm), but a useful forecast if the observed value is 350 mm. So, the same absolute error can have various significance depending on how the forecast is placed with respect to climate. The issue is made more complex by the scale difference between model and observations. In many cases indeed, we should not expect the current models to reproduce the maximum values of weather parameters observed in extreme events because their resolution is too low. We should however design methods to diagnose severe weather based on the existing models, and thoroughly verify the validity of these diagnostics.

The linear error in probability space method (LEPS, Ward and Folland, 1991) is an early attempt to deal with this problem. If f is the forecast, o the observation, and F(o) the cumulative probability density function of o, (i.e. the probability that the observation is smaller then o), the LEPS measure of the error is the difference F(f)-F(o). Therefore, large differences between f and o are less penalized if they occur near extreme values of the distribution of o. The minimum error is 0 and the maximum error is 1.

The Extreme Forecast Index, developed recently at ECMWF (Lalaurette, 2002) provides a generalization to probabilistic forecasts. The extreme forecast index (EFI) is a measure of the difference between a probabilistic forecast and a model climate distribution. In order to avoid a dependence on the climate of the region under study, it is desirable that such an index do not scale like the forecast parameter, but varies from –1 (an extreme negative value) to +1 (an extreme positive value). To achieve this goal, the EFI is formulated in the probability space: for a given location on Earth and a given meteorological parameter, one associates to each proportion p of the ranked model climate records a parameter threshold $q_c(p)$, known as the percentile of the distribution: $q_c(0)$ is the absolute minimum, $q_c(0.5)$ the median, $q_c(1)$ the absolute maximum. We then define $F_f(p)$ as the probability with which a probabilistic forecast predicts that the observation will be below $q_c(p)$, and write $EFI= \int (p-F_f(p) \, dp$. The index cumulates the differences between the climate and forecast distributions. $F_f(p)=p$ only in the case where forecast probability distribution is exactly the same as the climate, and in this case EFI=0. This will be also true for a deterministic forecast calling for the median value of the climate record. Furthermore, EFI=+1(-1) only if all possible values in the forecast are above (below) the highest (lowest) value of the climate record. In practice, an exponent (3) is used in order to have the EFI varying more rapidly near the extreme values. One limitation of the EFI is the need to have a good representation of the model climate. In practice, this can only be obtained by running a constant version of the model (or nearly constant) during several years. There is some hope that the time period needed to accumulate enough statistics can be considerably reduced by using ensemble predictions, providing many realizations of the forecast every day. In order to verify the EFI forecast, the model analysis or short-range forecast can be used. Contingency tables can be constructed to count the number of occasions when the EFI prediction performed well or bad in exceeding a given value. Thus, categorical scores can be produced for the EFI prediction. Also, to account for under- or over-prediction of extreme events by a model, one may decide to issue a warning when the EFI forecast exceeds a value lower or higher than the target, and construct ROC curves. This type of verification is believed to be extremely useful to increase and assess the capacity of a NWP model to predict extreme events, with due regards to its systematic biases. The ECMWF EFI system is being developed in the frame of a medium resolution ensemble prediction system, but it is believed that a similar approach could be adopted for deterministic or probabilistic forecasts from a high resolution model, provided a convenient knowledge of the model climate is at hand.

5. Main verifications of weather elements currently performed at operational centers

A survey of methods currently in use or in development has been performed, focusing on the verification of weather elements. The following is a summary of replies by operational NWP centers, indicating only the major efforts. There is no intention to provide an exhaustive list of verifications performed by these centers.

Australia: the rainfall forecast is verified against an analysis of 24-hour rain gauge data over continental Australia. The resolution of the analysis is 0.25 degrees, and the analysis is remapped to the model resolution. The basic verification relies on bias, RMSE, and contingency tables from which various categorical scores are computed. The statistics are written to files and saved for various aggregation and display schemes. An object-oriented verification (Contiguous Rain Area method, Ebert and McBride, 2000) is also performed on up to four individual rain systems per day. The location, volume, and pattern errors are computed, as well as errors in rain area and intensity. This is considered very useful for extreme events. Some work is in progress with radar data.

Canada: Bias and RMSE of wind, temperature, dew point, and surface pressure to surface and upper air stations are routinely monitored. For precipitation, bias and threat scores for various thresholds are computed to the synop stations, and more recently to a higher resolution SHEF (standard hydrometeorological exchange format) network (Belair et al., 2000). Work on the North American radar data has started and will be used to assess the relative importance of the various physical processes in the model.

China: 400 stations were carefully chosen over China's territory for precipitation forecasts verification. Both NWP models and subjective forecasts are interpolated to the location of these stations, and the verification is done routinely. It is based on threat scores and bias for various thresholds (0.1, 10, 25, 50 and 100 mm/24 hours).

France: About 1200 synoptic and automated surface weather stations are used. The parameters subject to systematic verification are the precipitation, the cloud cover, the temperature and humidity at 2m, the wind speed and direction and the intensity of the wind gusts. The nearest model grid point is used to compare with the point observations. The biases and RMSE are computed. In addition contingency tables are computed for precipitation (4 classes) and cloud cover (3 classes). All observations and forecasts at each point station are retained in a single database in order to conduct analyses of the model performance by sorting stations according to various criteria. Further contingency tables are being developed for wind speed and wind gusts. Work is in progress concerning the use of radar data to verify the precipitation forecast, and object-oriented methods.

Germany: At DWD, verification of precipitation is done using a high density network of observations (around 3600 sites with daily totals). The following verification strategies are used: (i) user-oriented verification: comparison of observations with forecasted values at the nearest grid point of the model or with an interpolated value from the surrounding grid points; (ii) modeler-oriented verification: computation of super-observations in different grids (1°x1° grid for WGNE; in the grid of the global model; in the grid of the regional model). Verification using Synop data is also done for the operational models.

Japan: JMA operates a high-resolution surface observation network named the Automated Meteorological Data Acquisition System (AMeDAS), which consists of 1300 raingauges, 200 snowgauges and 800 thermometers, aerovanes and heliographs all over Japan. Its estimated grid spacing is about 17km for raingauges and 20 km for other facilities. The AMeDAS data are used to verify forecast performance on both precipitation and surface temperature. The observational data are converted into a set of uniform grid data in 80 km mesh and the forecasts are compared with the gridded observations. This method is adopted to avoid discontinuity caused by changes in model resolution and to reduce sampling error of observation. JMA also operates 20 radar sites and produces a precipitation analysis over Japan by compositing radar reflectivities and AMeDAS raingauge data. This analysis is used to evaluate the forecast skill of the mesoscale model at three different resolutions: 10, 40 and 80km, and for time periods of 1, 3, and 6 hours. The regional spectral model is verified with the same data at 20, 40, and 80km resolution. Standard categorical scores are computed, such as threat score, bias score and equitable threat score.

UK: Operationally, the UK mesoscale forecasts are assessed by a summary index based on five parameters: 1.5m temperature, 10m wind, 6h accumulated precipitation, total cloud cover and visibility. Skill scores from T+6 to T+24 are used, with 42 stations used as truth. For temperature and wind the skill scores are based on mean square errors compared to persistence. Equitable threat scores are used for precipitation, cloud cover and visibility with thresholds of 0.2, 1.0, 4.0 mm/6h, 2.5, 4.5, 6.5 oktas for clouds, and 5km, 1km, and 200m for visibility. UK is also making a considerable effort to use radar composites to verify precipitation. Within the

NIMROD system, ground clutter, corrupt images, and anomalous propagation effects are removed, the vertical profile of reflectivity is taken into account, and calibration against gauges is adjusted once per week.

United States: at the U.S. National Centers for Environmental Prediction, model forecasts of surface and upper-air fields are verified against a myriad of observational data, including height, temperature, wind and moisture observations from radiosondes, dropsondes, land and marine observation stations; temperature and wind from aircraft at flight level and during ascent/descent; and upper air winds from pibals, profilers, satellite derivations and doppler radar VAD product. Model fields are interpolated to the location of the observation for the comparison. The extensive verification database allows evaluation of model performance from a variety of angles. Daily (12Z-12Z) precipitation verification is performed using a 0.125 degree precipitation analyses over the contiguous United States based on 7,000-8,000 daily gauge reports which are quality-controlled with radar and climatological data. The verification is done on 80-km and 40-km grids for NCEP operational models and various international models, and on a 12-km grid for NCEP's mesoscale non-hydrostatic nested model runs. Precipitation fields (forecast and observed) are mapped to the verification grids. From the precipitation forecast/observed/hit statistics collected for the verification domains (Continental US and 13 subregions), 26 different scores can be calculated, among which are equitable threat, bias, probability of detection, false alarm rate and odds ratio. Limited 3-hourly verification is also performed using NCEP's hourly 4-km multi-sensor precipitation analysis based on radar and automated hourly gauge reports. Monthly/month-to-date precipitation verification graphics are available at:
http://wwwt.emc.ncep.noaa.gov/mmb/ylin/pcpverif/scores/ .

Except for precipitation, where raingage observations are used in a procedure similar to that described above, global (medium range, 15 days) and regional (short-range, 48 h) ensemble forecasts generated operationally at NCEP are currently evaluated against gridded NWP analysis fields. Beyond the traditional scores (root mean square error and anomaly correlation for the individual members and the ensemble mean field) analysis rank histograms and histograms assessing the time consistency of consecutive ensemble forecasts are also computed (Toth et al, 2003). Probabilistic forecasts derived from the ensemble, including spread and reliability measures, are evaluated using a variety of standard probabilistic verification scores including the Brier Skill Score, Ranked Probability Skill Score, Relative Operating Characteristics, and Economic Value of forecasts. The latter two measures are also computed for single higher and equivalent resolution "control" forecasts originating from unperturbed initial fields, allowing for a comparative analysis of the value of a single higher resolution forecast and a lower resolution ensemble of forecasts.

Russia: The grid-point values of the non-hydrostatic meso-scale model are compared with nearby stations directly. The verified parameters are: surface pressure (bias, mean absolute and rms errors); surface temperature (bias, mean absolute and rms errors, relative error); wind (mean absolute vector error, mean absolute speed and direction errors, speed bias, scalar and vector RMSE); precipitation (an ensemble of scores based on contingency tables).

6. Conclusions

While it is impossible to cover the whole field of verification techniques in this survey, several conclusions emerge from the review of current works and debates:

1. As high resolution models are expected to provide results in direct relevance to user needs, there is a growing pressure to develop ‘user-oriented’ verifications. These can depart significantly from the more traditional model-oriented verifications, for instance in the choice of actually used observations or scrutinized model scales. Since it is difficult to accommodate several different needs in the same software, it is recommended to separate clearly the user-oriented and model-oriented parts of the verification packages.
2. The resolution of the observing networks is now often inferior to the resolution of NWP models. This calls for an improvement of observing networks, and design of more adequate verification techniques, especially for weather elements. Enhanced international exchange of high resolution data should also be encouraged.
3. The difference in horizontal scale between the forecast and the observations is too often neglected. However, no really adequate technique appears to exist to deal with this problem, except in some very special cases where upscaling of observations is possible.
4. The detailed prediction of some weather elements often appears at the limit of current NWP models capacity. This is due to specific predictability problems, and to remaining weaknesses in the model formulation, observations, and data assimilation techniques.
5. The double penalty problem remains a central issue with which many verification scientists are struggling. Several approaches to this problem are pursued. (i) Use of convenient battery of scores

(e.g. ETS and FBI); (ii) up-scaling the verification; (iii) formulate the forecast in a probabilistic way, either by use of ensembles, or by use of a collection of neighbor grid points and time steps.

6. Because of the intrinsic predictability limitation, the verification problem for weather elements at high resolution is better posed in a probabilistic way. There is a need to develop probabilistic formulations of the forecast and adequate verifications.
7. Severe weather verification poses a specific problem, and currently requires a verification in the probability space (such as LEPS or EFI). The relative frequency of severe events should be matched between model and observation rather than their quantitative representation. This requires a good knowledge of the model climate.
8. The verification problem shares many aspects with the data assimilation problem, for instance the computation of observation operators and of differences between forecasts and observations. This should be recognized and exploited in the development and maintenance of software.
9. A set of standard verifications should be defined for weather elements from high resolution NWP. This may be the subject of future work of the WGNE.
10. Verification scores should always be accompanied by information on the uncertainty and/or statistical significance. Extreme cases are very limited in number, and verification without proper account of uncertainty may easily result in wrong conclusions. Comparison with more simple forecasting methods, such as climate, persistence, and chance should also be provided as a reference. Also, the model analysis (initial field) should be scored with the same technique as the forecast in order to provide a reference. Some scores are more easily amenable to uncertainty computation and more "equitable" (in the sense that they are less sensitive to the sample composition). Recently, the Odds Ratio has been claimed to possess those qualities.

Advanced verification techniques are under development in various centers to provide model developers with more appropriate information on the origin of errors and the realism of models. Verification can have a number of different objectives and no single technique can address all objectives at once. Verification will remain a complex and important subject. It is believed that the search for better verification methods is one powerful way to reach better forecasts.

Acknowledgements: This paper has benefited of much input from various people. I am particularly indebted to Elizabeth Ebert (Australia), René Laprise, Jocelyn Mailhot and Laurie Wilson (Canada), Chen De Hui (China), François Lalaurette (ECMWF), Samuel Westrelin and Frederic Atger (France), Ulrich Damrath (Germany), Goda Harutoshi (Japan), Gerard Kats (Netherlands), Dmitry Kiktev (Russia), Martin Goeber, Andrew Lorenc and Clive Wilson (UK), Stephen Lord and David Williamson (USA).

References

Anthes, R.A., 1983: Regional models of the atmosphere in middle latitudes. Mon. Wea. Rev., 111, 1306-1330.

Atger, F., 2001: Verification of intense precipitation forecasts from single models and ensemble prediction systems. Nonlinear Processes in Geophysics, 8, 401-417.

Belair, S., A. Methot, J. Mailhot, B. Bilodeau, A.Patoine, P. Pellerin and J. Coté, 2000: Operational implementation of the Fritsch-Chappel scheme in the 24-km Canadian regional model. Wea. Forecasting, 15, 257-274.

Brier, G.W., 1950: Verification of forecasts expressed in terms of probability. Mon. Wea. Rev., 78, 1-3.

Briggs, W.M., and R.A. Levine, 1997: Wavelets and field forecast verification. Mon. Wea. Rev., 125, 1329-1341.

Brooks, H.E., M. Kay and J.A. Hart, 1998: Objective limits to forecasting skill of rare events. Preprints, 19th Conf. On Severe Local Storms, Minneapolis, MN, Amer. Meteor. Soc, 552-555.

Cherubini, T., A. Ghelli and F. Lalaurette, 2001: Verification of precipitation forecasts over the Alpine region using a high density observing network. ECMWF Technical Memorandum, n°340, 18pp.

Davis, C., and F. Carr, 2000: Summary of the 1998 Workshop on mesoscale model verification. Bull. Amer. Meteor. Soc., 81, 809-819.

De Elia, R., and R. Laprise, 2002: Distribution-oriented verification of limited-area model forecasts in a perfect-model framework. Mon. Wea. Rev., submitted.

Ebert, E.E., and J.L. McBride, 2000: Verification of precipitation in weather systems: determination of systematic errors. J. Hydrology,239, 179-202.

Ebert, E.E., U. Damrath, W. Wergen, and M. Baldwin, 2002: The WGNE assessment of short-term quantitative precipitation forecasts from operational NWP models. Bull. Amer. Meteor. Soc., to appear.

Glahn, H.R., 1976: Forecast evaluation at techniques development laboratory. In “Weather Forecasting and Weather Forecasts: Models, Systems and Users”. NCAR Colloquium, Summer 1976, pages 831-838.

Goeber, M., and S. Milton, 2002: Verifying precipitation events. NWP Gazette, March 2002, Met Office, 9-11.

Hamill, T.M., 1997: Reliability diagrams for multi-category probabilistic forecasts. Wea. Forecasting, 12, 736-741.

Hamill, T.M., 1999: Hypothesis tests for evaluating numerical precipitation forecasts. Wea. Forecasting, 14, 155-167.

Hoffman, R.N. Z. Liu, J.F. Louis, and C. Grassoti, 1995: Distortion representation of forecast errors. Mon. Wea. Rev., 123, 2758-2770.

Lalaurette, F., 2002: Early detection of abnormal weather using a probabilistic Extreme Forecast Index. ECMWF Technical Memorandum, n°373, 27pp.

Mass, C.F., D.Ovens, K. Westrick and B.A. Colle, 2002: Does increasing horizontal resolution produce more skillful forecasts? Bull. Amer. Meteor. Soc., March 2002, 407-430.

Murphy, A.H., 1991: Forecast verification: its complexity and dimensionality. Mon. Wea. Rev., 119, 1590-1601.

Murphy, A.H., 1993: What is a good forecast? An essay on the nature of goodness in weather forecasting. Weather and Forecasting, 8, 281-293.

Murphy, A.H., and R.L. Winkler, 1987: A general framework for forecast verification. Mon. Wea. Rev., 115, 1330-1338.

Richardson, D.S., 2000: Skill and economic value of the ECMWF Ensemble Prediction System. Quart. J. Royal Meteor. Soc., 126, 649-668.

Stanski, H.R., L.J. Wilson, and and W.R. Burrows, 1989: Survey of common verification methods in meteorology, WMO/WWW Tech. Rep. n° 8.

Stephenson, D.B., 2000: Use of the “Odds Ratio” for diagnosing forecast skill. Weather and Forecasting, 15, 221-232.

Toth, Z., O. Talagrand, G. Candille, and Y. Zhu, 2003: Probability and ensemble forecasts. In: Environmental Forecast Verification: A practitioner's guide in atmospheric science. Ed.: I. T. Jolliffe and D. B. Stephenson. Wiley, in print.

Ward, M.N., and C.K. Folland, 1991: Prediction of seasonal rainfall in the north nordeste of brazil using eigenvectors of sea surface temperature. Int. J. Climatology, 11, 711-743.

Wilks, D.S., 1995: Statistical Methods in the Atmospheric Sciences. Academic Press, 467pp.

Zepeda-Arce, J., E. Foufoula-Georgiou and K.K. Droegemeier, 2000: Space-time rainfall organization and its role in validating quantitative precipitation forecasts. J. Geophys. Res., 105, 10129-10146.