WMO/ICSU/IOC
WORLD CLIMATE
RESEARCH PROGRAMME

WMO
WEATHER PREDICTION
RESEARCH PROGRAMMES

CAS/JSC WORKING GROUP
ON NUMERICAL EXPERIMENTATION

RESEARCH ACTIVITIES IN ATMOSPHERIC AND OCEANIC MODELLING

Edited by J. Côté

Report No. 33

APRIL 2003
WMO/TD-No. 1161

Editorial note : This report has been produced without editorial revision by the WMO secretariat. It is not an official WMO publication and its distribution in this form does not imply endorsement by the Organisation of the ideas expressed.

From the Editor

There is considerable international activity in the development of numerical models for the purpose of climate simulation and for forecasting on various timescales. This publication is an attempt to foster an early interchange of information among workers in these areas. The material in the publication is the response to a "call for contributions" sent to approximately 650 scientists worldwide. Contributions obtained in response to this call are included if they are related to the CAS/JSC numerical experimentation programme, if they give new results, and if they are of suitable length and size. Reports that do not meet these criteria, have been previously published, or are purely theoretical may be rejected. Contributors do not routinely receive any correspondence concerning the contributions.

The most appropriate reports give results of new numerical experiments in the form of a succinct explanation accompanied by suitable tables and figures. The contributions are collected into subject groupings as appropriate. The range of subjects is expected to vary with time and depends on the submissions received. The large number of contributions from around the world indicates the wide scope of activities in numerical experimentation, and the valuable addition that this type of report makes to the refereed journals. Comments and suggestions for improvement to the publication are welcomed. To facilitate location of specific contributions, they are ordered alphabetically by author in the various subject areas. An overall index by author is also included.

This year's report contains as an annex the report made to WGNE by Philippe Bougeault on verification. To reflect on the many submissions on short-range ensemble forecasting, they have been placed under regional model development. We have also consolidated the electronic production of this report. Most contributions were submitted through the web site www.cmc.ec.gc.ca/rpn/wgne, few still as an attachment to an e-mail message. Overall the electronic submissions worked well, thanks to Yves Chartier and Inès Ng Kam Chan, and made possible the production of this report on the web site. About 200 copies have also been printed in black and white and mailed directly from Montreal.

As mentioned in this year's call letter, I have become editor following Harold Ritchie who has occupied this position for the last four years. I thank Hal for his period of service marked by the initiation of the electronic publication. I look forward to your continued support.

J. Côté

Jean Côté
Recherche en prévision numérique
Meteorological Service of Canada
2121, route Transcanadienne
Dorval, Québec H9P 1J3
Canada

ACTIVITIES OF THE CAS/JSC WORKING GROUP ON NUMERICAL EXPERIMENTATION (WGNE)

The JSC/CAS Working Group on Numerical Experimentation (WGNE) has the central responsibility in the WCRP for the development of the atmospheric component of climate models and, together with the Working Group on Coupled Modelling (WGCM), lies at the core of the climate modelling effort in WCRP. Close co-ordination is duly maintained between WGNE and WGCM. Furthermore, WGNE works in conjunction with the WCRP Global Energy and Water Cycle Experiment (GEWEX) in the development of atmospheric model parameterizations and, in this respect, WGNE sessions are held jointly with the "GEWEX Modelling and Prediction Panel". Liaison is also maintained with the SPARC "GRIPS" project (focussed on the intercomparison of model stratospheric simulations) and will be developed with the new SPARC initiative on stratospheric data assimilation.

WGNE additionally has an important role in support of the WMO Commission for Atmospheric Sciences (CAS) in reviewing the development of atmospheric models for use in weather prediction and climate studies on all timescales. The close relationship between WGNE and operational (NWP) centres by virtue of the CAS connection underpins many aspects of WGNE work and provides a strong impetus for the refinement of the atmospheric component of climate models. WGNE sessions duly include reviews of progress at operational centres in topics such as data assimilation, numerics, physical parameterizations, ensemble predictions, seasonal prediction, forecasting tropical cyclone tracks, and the verification of precipitation forecasts. WGNE maintains close co-ordination with the CAS World Weather Research Programme. A particularly strong area of collaboration is in the planning and development of THORPEX: A Global Atmospheric Research Programme.

The following paragraphs review the main activities of WGNE in support of WCRP objectives, including especially items of interest and recommendations arising at the eighteenth session of the group kindly hosted by Meteo-France, Toulouse, France, 18-22 November 2002.

1. MODEL INTERCOMPARISON PROJECTS

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, is a series of model intercomparison exercises.

Atmospheric Model Intercomparison Project

The most important and far-reaching of the WGNE-sposored 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. AMIP, 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, 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

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 a companion project CCPP-ARM parameterization Testbed (CAPT) being undertaken by PCMDI and NCAR 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.

Snow Models Intercomparison Project (SNOWMIP)

SNOWMIP is being 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 parameterisation (e.g. albedo, water retention…) the results are highly variable and some show discrepancies between observations and models. For instance an albedo parameterisation 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/.

Comparisons of stratospheric analyses and predictive skill in the stratosphere

In the past two or three years, there has been growing interest in the representation of and prediction in the stratosphere and several major global operational centres have significantly increased the vertical extent and resolution of their models and associated data assimilation and predictions in the stratosphere and into the mesosphere. 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 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 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.

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 occuring 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.

2. STANDARD CLIMATE MODEL DIAGNOSTICS

The 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 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/index.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.

3. DEVELOPMENTS IN NUMERICAL APPROXIMATIONS

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 isosahedral, 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 nonlinear 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 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.

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-plant 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

4. 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 to which are added the climate change from CGCM simulations. RCMs are usually nested in AGCM simulated atmospheric states rather than CGCM atmospheric fields in order to reduce systematic biases.

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 scales 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.

5. 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 (eg. 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.

6. ATMOSPHERIC MODEL PARAMETERIZATIONS

The GEWEX "modelling and prediction" thrust, with which WGNE works in close association, is devoting efforts to the refinement of atmospheric model parameterizations, notably those of clouds and radiation, land surface processes and soil moisture, and the atmospheric boundary layer.

Clouds

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.

GCSS is composed of five working groups, relating to boundary-layer cloud systems, cirrus cloud systems, extratropical layer cloud systems, precipitating convective cloud systems, 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 inhomogeniety due to the combination of CRMs, cloud radar observations, and faster methods of calculating radiative fluxes for arbitary cloud configurations;
- steady progress in the understanding and representation of cloud microphysical, formation, and dissipation processes due to integrated use of LES (large-eddy simulations) 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.

Land-surface processes

The GEWEX Global Land-Atmosphere Study (GLASS) project 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, a new local study including carbon fluxes was initiated over a forested land in the Netherlands, 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 on the role surface moisture and temperature states play in the evolution of weather and the generation of precipitation, a new study called GLACÉ will address the problem of the relative role of land-surfaces in the variability of the climate system. This will be based on a series of GCM experiments using coupled free and forced land-surface schemes.

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. GABLS aims to provide a platform in which scientists working on boundary layers at different scales will interact.

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.

7. GEWEX 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. John Roads made a presentation on the current status of CEOP and also responded to concerns that had 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 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 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.

8. REANALYSES

ECMWF

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 carried out in collaboration with NCEP and NCAR is complete. 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 data base. EUMETSAT is also reprocessing wind products from METEOSAT-2 from 1983-1988. The collection of observations is almost complete and the observational 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 system that was operational from June 2001 to January 2002 and includes 3DVAR analysis, T159 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

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)

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 3DVAR 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.

9. PERFORMANCE OF THE MAIN GLOBAL OPERATIONAL MODELS

As 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. 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.

10. INTERCOMPARISON OF TYPHOON TRACK FORECASTS

An intercomparison of forecasts of typhoon tracks in the western North Pacific has been conducted by the Japan Meteorological Agency on behalf of WGNE for a number of years. The intercomparison has now been extended to cover also the North Atlantic and eastern North Pacific regions. The operational centres submitting forecasts now include ECMWF, the Met Office, the Canadian Meteorological Center, Deutscher Wetterdienst, and the Japan Meteorological Agency. Results show continuing improvements in the track forecasts although there is considerable variability from year to year and from basin to basin. A report summarizing the results of the intercomparisons over the period 1991-2000 has being prepared by the Japan Meteorological Agency and published as a WMO report.

11. VERIFICATION AND COMPARISON 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.

12. THORPEX: A GLOBAL ATMOSPHERIC RESEARCH PROGRAMME

At the invitation of WGNE, Prof Alan 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.

13. REVIEW ON 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 (eq 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 parameterisation?
- How far can we go with a given physical parameterisation?
- 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?

14. VERIFICATION TECHNIQUES FOR MESO-SCALE 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. Philippe Bougeault on behalf of WGNE and is included in the current issue. 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.

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

Kamal Puri

(Kamal Puri)
Chairman, WGNE

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 and WGNE have recently initiated a joint 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 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 event 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 up-scaling 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 varying 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 up-scaling 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.

Table of Contents

01. Atmospheric data assimilation schemes, analysis and initialization, data impact studies, observing system experiments

Author	Paper Title	Country	Pages
Oleg Alduchov and Vladimir Gordin	Anisotropy of Meteorological Fields	Russia	01-01
Oleg Alduchov, Irina Chernykh and Robert Eskridge	Errors in detecting of cloud boundaries for detailed radiosonde profiles	Russia	01-03
Vladimir Bayev and Vladimir Gordin	On Impact of Radiosondes' Shift into Objective Analysis	Russia	01-05
Teruyuki Kato, Yoshiaki Sato and X-BAIU-01 observation group	Importance of vertical water vapor analysis - A problem in assimilating TPW with 4DVar	Japan	01-07
Michael Keil	Assimilating long-duration balloon data using a Partial OSSE	UK	01-09
Ekaterina Klimova	A suboptimal Kalman filter algorithm with the simplified models for calculation of the forecast error covariances	Russia	01-11
Ching-Chi Lam	Case Study on the Impact of Radar-derived TREC Winds on Model Forecast of Heavy Rain Associated with Landfalling Tropical Cyclone	China	01-13
Rolf Langland	Estimation of Observation Value Using the NAVDAS Adjoint System	U.S.	01-15
Amos Lawless and Nancy Nichols	The convergence of incremental 4D-Var using non-tangent linear models	U.K.	01-17
S. Mark Leidner, Ross Hoffman and Mark C. Cerniglia	Rain Quality Control for SeaWinds Near Real-time Data	USA	01-19
Ying Lin, Geoff DiMego, Brad Ferrier, Jim Jung, Dennis Keyser and Eric Rogers	Assimilation of GOES Cloud Top Pressure Data into the Eta Model	USA	01-21
Jean-Francois Mahfouf and Stephane Laroche	On the lower boundary condition of tangent-linear models	Canada	01-23
C. Jesse Meng	Intercompatison of Global Land Surface Solar Insolation in Operations GCMs	USA	01-25
Kazuya Nojima and Yoshiyuki Nakamura	On the use of BUFR-formatted AMVs from METEOSAT satellites	Japan	01-27
Kozo Okamoto and Masahiro Kazumori	The experiments of direct assimilation of ATOVS radicances in the JMA 3D-Var system	Japan	01-28
Andreas Rhodin	Modeling of Covariance Matrices by Generalized Diffusion Operators - A Hierarchical Approach	Germany	01-30
Ian Roulstone	Data Assimilation for Convective Scale NWP	UK	01-32
Hiromu Seko, Hajime Nakamura, Ko Koizumi and Tetsuya Iwabuchi	Data Assimilation Experiments in Nerima Heavy Rainfall - How Useful are Doppler Radar Radial Wind and GPS-Derived Water Vapor? -	Japan	01-34
Monique Tanguay	Adjoint of the non-hydrostatic GEMDM LAM	Canada	01-36

Table of Contents

Masami Tokuno and Yasuaki Ohhashi	Assimilation of QuikSCAT/SeaWinds Ocean Surface Wind Data into the Global Data Assimilation System at JMA	Japan	01-37
Mamadou Tounkara, P. Caille and J. Pailleux	The impact of tropical radiosonde data on the ARPEGE analyses and forecasts	France	01-39

02. Data sets, diagnostic and dynamical investigations, statistical post-processing , multi-year reanalyses and associated studies

Author	Paper Title	Country	Pages
Irina Chernykh, Oleg Alduchov and Victor E. Lagun	About Cloudiness Low Boundary over Antarctic Peninsula	Russia	02-01
Alexey V. Eliseev, M. V. Guserva, Igor I. Mokhov and K. G. Rubinstein	Intercomparison of NCEP and ERA reanalyses in terms of characteristics of surface aire temperature annual cycle	Russian	02-03
M. Goeber and C.A. Wilson	WGNE assessment of Quantitative Precipitation Forecasts from Operational Numerical Weather Prediction Models over the U.K.	UK	02-05
Igor Mokhov and Alexandre Chernokulsky	Global cloudiness: Tendencies of change from ISCCP data	Russia	02-07
Igor Mokhov and Mirseid Akperov	Intense Arctic and Antarctic mesocyclones (polar lows) and their variability	Russia	02-09
Kazutoshi Onogi , Hiroshi Koide, Masami Sakamoto, Nobutaka Manoji, Nobuo Yamazaki, Hirotaka Kamahori, Kiyotoshi Takahashi, Jun'ichi Tsutsui and Hiroaki Hatsushika	Preliminary Experiments for Japanese 25-year Reanalysis Project	Japan	02-11
E.V. Perekhodtseva	Statistical Models for Operative Forecast of Dangerous Convective Phenomena like Rainfalls, Squalls and Tornadoes in the European Part of Russia and in Europe	Russia	02-13
Harun Rashid and Ian Simmonds	Optimal nonmodal growth as a contributing cause to the Southern Hemisphere annular mode variability	Australia	02-15
Gdaly Rivin and Erdmann Heise	Deutscher Wetterdienst (DWD) meteorological data set for the development of a flood forecasting system	Germany	02-17
Gregory Roff	GRIPS GCM simulated polar vortex intercomparison	Australia	02-19
Ryan Sharp, Mark Bourassa and James O'Brien	Tropical Cyclogenesis Detection in the East Pacific Hurricane Basin	United States	02-21
Ian Simmonds	Origins of northern Antarctic Peninsula air masses	Australia	02-23

Table of Contents

Shawn Smith, Mark Bourassa and Rosario Romero	Tropical Pacific wind comparisons: objective FSU versus NCEP reanalysis products	USA	02-25

03. Computational studies including new techniques, the effect of varying model resolution, parallel processing

Author	Paper Title	Country	Pages
Pierre Bénard	Horizontal diffusion operators in models with a stretched horizontal geometry	France	03-01
Pierre Bénard, Jozef Vivoda, Petra Smolikova and Jan Masek	Stable algorithms for the NH version of the LAM "Aladin"	France	03-03
Luca Bonaventura , M. Giorgetta, I. Kirchner, L. Kornblueh, D. Majewski, A. Muller, A. Rhodin, B. Ritter and E. Roeckner	ICON : Development of an icosahedral, non-hydrostatic global model	Germany	03-05
Michael Cullen	Preserving long-term accuracy in semi-implicit schemes for slightly compressible flow	U.K.	03-07
Tsukasa Fujita	Higher order finite difference schemes for advection of NHM	Japan	03-09
Keiichi Katayama, Hiromasa Yoshimura and Takayuki Matsumura	Development of a 20 km mesh global NWP model on the Earth Simulator	Japan	03-11
William Merryfield, Norman McFarlane and M. Lazare	A generalized hybrid transformation for tracer advection	Canada	03-13
Kazuo Saito	Time-splitting of advection in the JMA Nonhydrostatic Model	Japan	03-15
Mikhail Tolstykh	Variable resolution version of the SL-AV global NWP model	Russia	03-17
Hiromasa Yoshimura and Takayuki Matsumura	A Semi-Lagrangian Scheme Conservative in the Vertical Direction	Japan	03-19
Jorge Zavala-Hidalgo , Peng Yu, Steven L. Morey, Mark Bourassa and James J. Obrien	A new interpolation method for high frequency forcing fields	USA	03-21

Table of Contents

04. Parameterization of important atmospheric and surface processes, effects of different parameterizations

Author	Paper Title	Country	Pages
I.V. Akimov	Verification of precipitation calculation method based on parameterization of distribution function evolution	Russia	04-01
Guenther Dom, Detlef Majewski, A. Muller and B. Ritter	A Prognostic Cloud Ice Scheme for NWP-Models	Germany	04-03
Pierre Etchevers, E. Martin, R. Brown, C. Fierz, Y. Lejeune, E. Bazile, A. Boone, Y.-J. Dai, R. Essery,A. Fernandez , Y. Gusev, R. Jordan, V. Koren, E. Kowalczyk, N. O. Nasonova, R. D. Pyles, A. Schlosser, A. B.Shmakin, T. G. Smirnova, U. Strasser, D. Verseghy, T. Yamazaki, Z.-L. Yang	The SNOWMIP project : validation of albedo and short wave radiation budget simulated by several snow models.	France	04-05
Masayuki Hirai and Mitsuo Oh'izumi	Development of a New Land-surface Model for JMA-GSM	Japan	04-07
Yu-Tai Hou, S. Moorthi, K. Campana and M. Iacono	Application of RRTM longwave radiation to NCEP GFS	US	04-09
Hideaki Kawai	Impact of a Cloud Ice Fall Scheme Based on an Analytically Integrated Solution	Japan	04-11
Dmitrii Mironov and Bodo Ritter	A first version of the ice model for the global NWP system GME of the German Weather Service	Germany	04-13
Akihiko Murata and Mitsuru Ueno	Numerical simulations of a tropical cyclone using the Arakawa-Schubert scheme with vertically variable entrainment rate	Japan	04-15
Masayuki Nakagawa	Development of a Cumulus Parameterization Scheme of the Operational Global Model at JMA	Japan	04-17
Jon Radakovich , Guiling Wanf, Shian-Jiann Lin, Sharon Nebuda and Bo-wen Shen	Implementation of The NCAR Community Land Model (CLM) in the NASA/NCAR Finite-Volume Global Climate Model (fvGCM)	USA	04-19
Bodo Ritter , Reinhold Schrodin, Erdmann Heise, Martin Lange and Aurelia Mueller	Implementation and testing of a multi-layer soil model in the NWP models of DWD	Germany	04-21
Z. Sun and L. Rikus	Implementation of ARM derived ice water content in NWP model	Australia	04-23

Table of Contents

Mikhail Tolstykh	The use of combined closure in convection parameterization scheme	Russia	04-25
Mitsuru Ueno	An objective method to diagnose the contribution of steering flow at each level to tropical cyclone motion	Japan	04-27
Syoukichi Yabu	The Doppler Absorption in the Thermal Infrared Radiation Parameterization of JMA Global Spectrum Model	Japan	04-29

05. Development of and studies with regional and smaller-scale atmospheric models

Author	Paper Title	Country	Pages
L.V. Berkovich and Yu.V. Tkacheva	The hydrodynamic short-range of the local weather forecasting	Russia	05-01
Francois Bouttier	Arome; a new operational mesoscale NWP system for Meteo-France	France	05-03
Neill Bowler and Clive Pierce	Development of a short range QPF algorithm based upon Optic Flow techniques	United Kingdom	05-05
Ramon de Elia and Rene Laprise	Predictability of Limited-Area Models: Twin and Big-Brother Experiments	Canada	05-07
Jun Du, Geoff DiMego, Steve Tracton and Binbin Zhou	NCEP Short-Range Ensemble Forecasting (SREF) System : Multi-IC, Multi-Model and Multi-Physics Approach	USA	05-09
William Jr. Gallus and Isidora Jankov	Convective system rainfall forecast accurage as a function of large-scale forcing	USA	05-12
T. Haiden	NWP research in Austria	Australia	05-14
Dale Hess , Martin Cope, S. Lee, P. Manins, Kamal Puri, K. Tory, Manuela Burgers and Margaret Young	The Australian Air Quality Forecasting System	Australia	05-16
Zavisa Janjic , Tom Black, E. Rogers, H. Chuang and Geoff DiMego	The NCEP Nonhydrostatic Mesoscale Forecasting Model	USA	05-18
Xudong Liang and Johnny Chan	Convective Asymmetries Associated with Tropical Cyclone Landfall. Part I: F-Plane Simulations	China	05-20

Table of Contents

Wataru Mashiko and Chiashi Muroi	Development of a Two-Way Multiply-Nested Movable Mesh Typhoon Model Using the Cloud Resolving Nonhydrostatic Model	Japan	05-22
T. Paccagnella , D. Cesari., C. Marsigli, F. Nerozzi and S. Tibaldi	The Limited-Area Ensemble Prediction System COSMO-LEPS	Italy	05-24
A.E. Pokhil and A.D. Naumov	Analysis of the evolution and motion of tropical cyclones on the basis of calculations using the ETA model and satellite data	Russia	05-26
Kai Sattler and Henrik Feddersen	Study on limited-area short-range ensemble approaches targeted for heavy rain in Europe	Denmark	05-28
Hiromu Seko and Hajime Nakamura	Numerical Study of the Shapes and Maintenance Mechanisms of MESO-β Scale Line-Shaped Precipitation System in the Middle-Latitudes	Japan	05-30
S. Senotrusova and V. Svinuhov	Model of real classification of meteorological parameters for regional prognosis	Russia	05-32
Johnny Chan and Xiaqiong Zhou	Convective Asymmetries Associated with Tropical Cyclone Landfall. Part I: F-Plane Simulations	China	05-34

06. Developments in global forecast models, case studies, predictability investigations, ensemble forecasting monthly and seasonal forecasting

Author	**Paper Title**	**Country**	**Pages**
B. K. Basu	Summer Monsoon precipitation over India - a comparison of model predicted and observed values	India	06-01
Sandrine Edouard and Bernard Dugas	The use of a hybrid vertical coordinate in the CMC-MRB Global Environmental Multiscale (GEM) model	Canada	06-03
Hann-Ming Henry Juang	A Reduced Spherical Transform for the NCEP Seasonal Forecast Global Spectral Atmospheric Model	USA	06-05
Xudong Liang and Yihong Duan	Optimal-controlled Ensemble Prediction Technique and its Appilcation in ENSO Prediction	China	06-07
Wataru Ohfuchi and AFES Working Team	Global atmospheric simulations with the T1279L9 6 resolution	JAPAN	06-09
Tim Palmer, Francisco Doblas-Reyes and R. Hagedorn	Development of a European Multi-Model Ensemble System for Seasonal to Inter-Annual Prediction (DEMETER)	UK	06-11
Dong-Wook Shin, Steve Cocke, T. E. LaRow and J. J. O’Brien	Ensemble Configurations for Typhoon-related Rainfall Forecasts	USA	06-13

Table of Contents

G.V. Sourkova	Changes of the zonal atmospheric circulation at the Last Glacial Maximum as it is simulated by atmospheric global circulation models in PMIP	Russia	06-15

07. Global and regional climate models, sensitivity and impact experiments, response to external forcing

Author	Paper Title	Country	Pages
Vivek Arora and George Boer	The temporal variability of soil moisture and surface hydrological quantities	Canada	07-01
Michel Déqué, Florence Sevault and Samuel Somot	Impact of an IPCC-B2 scenario on Mediterranean sea temperature	France	07-03
Irina Chernykh and Oleg Alduchov	Joint Analys is of Climatic Changes of Surface Temperature and Cloudiness Vertical Structure in Antarctic Region on base CARDS	Russia	07-05
Michael Fox-Rabinovitz and Lawrence Takacs	A 12-year (1987-1998) ensemble simulation of the U.S. climate with a variable-resolution stretched-grid GCM	U.S.A.	07-07
Nikolaus Groll, Martin Widmann and Julie M. Jones	Teleconnections and upscaling relationships for the Eemian	Germany	07-09
R. S. Harwood, R. M. Doherty, I. A. Mackenzie and D. S. Stevenson	Chemistry and Climate Studies with the Met Office Unified Model	UK	07-11
Hann-Ming Henry Juang, Yucheng Song and Kingtse Mo	Regional Climate Simulations of the 1990 North American Summer Monsoon	USA	07-13
Igor Mokhov and Andrey Karpenko	Model projections of extreme runoff (floods) in Siberian rivers basins	Russia	07-15
David Noone	Sensitivity of the vertical coupling of Northern wintertime annular modes to anomalous high-latitude surface heating	USA	07-17
V.P. Parkhomenko	Sea Ice Cover Sensitivity analysis in Global Climate Model	Russia	07-19
Sebastian Wagner	Atmosphere-Sea Ice interaction during the Dalton Minimum within a historical simulation with ECHO-G	Germany	07-21
Richard Wardle and Ian Simmonds	Late Quaternary Southern Hemisphere extratropical cyclone characteristics	Australia	07-23

Table of Contents

08. Development of and advances in ocean modelling and data assimilation, sea-ice modelling, wave modelling

Author	Paper Title	Country	Pages
L. CavalerI, L. Bertotti and M. Sclavo	Long-Term Calibration of Wind And Waves in the Mediterranean Sea	Italy	08-01
A.A. Deo, D.W. Ganer and P. S. Salvekar	Ocean mixed layer response to a Bay of Bengal cyclone : A Case Study	India	08-05
Masayoshi Ishii, Kenji Sakamoto and Masahide Kimoto	Ocean data assimilation system using temperature, salinity, and sea surface height observations, and its impact on El Nino prediction	Japan	08-07
Erik Kvaleberg, Steven L. Morey and James J. O'Brien	Modeling frontal instabilities in the Gulf of Mexico	USA	08-08
Arkadiy S. Safray and Igor V. Tkatchenko	Numerical Investigation of Different Types of Open Boundary Conditions by Modelling of the Internal-Wave Transformation at the Edge of the Shelf by M2 Tidal Forcing	Russia	08-10
Hideharu Sasaki, Nobumasa Komori, Keiko Takahashi, Yukio Masumoto and Hirohumi Sakuma	Fifty years Time-integration of Global Eddy-resolving Simulation	Japan	08-12
Bulusu Subramanyam, James O'Brien, Lian Xie and John Morrison	Examination of Heat Transport in the Indian Ocean using altimetry and MICOM	USA	08-14
Akiyoshi Wada	Improvement of wind induced mixing and entrainment in MRI mixed layer model	Japan	08-16
Peng Yu, Jorge Zavala-Hidalgo and James J. O'Brien	A new gridding method for satellite altimeter data	U.S.A.	08-18
Alexander Zelenko and Yurii Resnyansky	Water exchange between the Arctic Ocean and adjacent basins: Essential role of short-term variations	Russia	08-20

09. Development of and studies with coupled ocean-atmosphere models

Author	Paper Title	Country	Pages
Kenneth R. Sperber , Julia M. Slingo, Peter M. Inness, Silvio Gualdi, Wei Li, Peter J. Gleckler and Charles Doutriaux	The Madden-Julian Oscillation in GCMs	USA	09-01

Table of Contents

Keiko Takahashi , Nobumasa Komori, Kenji Komine, Hirofumi Sakuma and Tetsuya Sato	Developing coupled ocean-atmosphere global climate model for the Earth Simulator	Japan	09-03
Akiyoshi Wada	Validation of typhoon intensity prediction by MRI typhoon-ocean coupled model	Japan	09-05
Akiyoshi Wada	Typhoon-ocean coupled model with upgraded mixed layer model	Japan	09-07

Index of Authors

AFES WORKING TEAM	06-09
Akimov, I.V.	04-01
Alduchov, Oleg	01-01
Alduchov, Oleg	01-03
Alduchov, Oleg	02-01
Alduchov, Oleg	07-05
Apkerov, Mirseid	02-09
Arora, Vivek	07-01
Basu, B. K.	06-01
Bayev, Vladimir	01-05
Bazile, E.	04-05
Bénard, Pierre	03-01
Bénard, Pierre	03-03
Berkovich, L.V.	05-01
Bertotti, L.	08-01
Black, Tom	05-18
Boer, George	07-01
Bonaventura, Luca	03-05
Boone, A.	04-05
Bourassa, Mark	02-21
Bourassa, Mark	02-25
Bourassa, Mark	03-21
Bouttier, Francois	05-03
Bowler, Neill	05-05
Brown, R.	04-05
Burgers, Manuela	05-16
Caille, Philippe	01-39
Campana, Ken	04-09
Cavaleri, L.	08-01
Cerniglia, Mark	01-19
Cesari, D.	05-24
Chan, Johnny	05-20
Chan, Johnny	05-34
Chernokulsky, Alexandre	02-07
Chernykh, Irina	01-03
Chernykh, Irina	02-01
Chernykh, Irina	07-05
Chuang, Hui-Ya	05-18
Cocke, Steve	06-13
Cope, Martin	05-16
Cullen, Michael	03-07
Dai, Y.-J.	04-05
de Elia, Ramon	05-07
Deo, A.A.	08-05
Déqué, Michel	07-03
DiMego, Geoff	01-21
DiMego, Geoff	05-09
DiMego, Geoff	05-18
Doblas-Reyes, Francisco	06-11
Doherty, Ruth	07-11
Doms, Guenther	04-03
Doutriaux, Charles	09-01
Du, John	05-09
Duan, Yihong	06-07
Dugas, Bernard	06-03
Edouard, Sandrine	06-03
Eliseev, Alexey V.	02-03
Eskridge, Robert	01-03
Essery, R.	04-05
Etchevers, Pierre	04-05
Feddersen, Henrik	05-28
Fernandez, A.	04-05
Ferrier, Brad	01-21
Fierz, C.	04-05
Fox-Rabinovitz, Michael	07-07
Fujita, Tsukasa	03-09
Gallus, William Jr.	05-12
Ganer , D.W.	08-05
Giorgetta, M.	03-05
Gleckler , Peter J.	09-01
Goeber, Martin	02-05
Gordin, Vladimir	01-05
Gordin, Vladimir A.	01-01
Groll, Nikolaus	07-09
Gualdi , Silvio	09-01
Guserva, M. V.	02-03
Gusev, Y.	04-05
Hagedorn, R.	06-11
Haiden, T.	05-14
Harwood, Robert	07-11
Hatsushika, Hiroaki	02-11
Heise, Erdmann	02-17
Heise, Erdmann	04-21
Hess, Dale	05-16
Hirai, Masayuki	04-07
Hirofumi, Sakuma	09-03
Hoffman, Ross	01-19
Hou, Yu-Tai	04-09
Iacono, Mike	04-09
Inness, Peter M.	09-01

Index of Authors

Ishii, Masayoshi	08-07
Iwabushi, Tetsuya	01-34
Janjic, Zavisa	05-18
Jankov, Isidora	05-12
Jones, Julie M.	07-09
Jordan, R.	04-05
Juang, Hann-Ming Henry	06-05
Juang, Hann-Ming Henry	07-13
Jung, Jim	01-21
Kamahori, Hirotaka	02-11
Karpenko, Andrey	07-15
Katayama, Keiichi	03-11
Kato, Teruyuki	01-07
Kawai, Hideaki	04-11
Kazumori, Masahiro	01-28
Keil, Michael	01-09
Keyser, Dennis	01-21
Kimoto, Masahide	08-07
Kirchner, I.	03-05
Klimova, Ekaterina	01-11
Koide, Hiroshi	02-11
Koisumi, Ko	01-34
Komine, Kenji	09-03
Komori, Nobumasa	08-12
Komori, Nobumasa	09-03
Koren, V.	04-05
Kornblueh, L.	03-05
Kowalczyk, E.	04-05
Kvaleberg, Erik	08-08
Lagun, Victor E.	02-01
Lam, Ching-chi	01-13
Lange, Martin	04-21
Langland, Rolf	01-15
Laprise, René	05-07
Laroche, Stéphane	01-23
LaRow, T. E.	06-13
Lawless, Amos	01-17
Lazare, Mike	03-13
Leem S.	05-16
Leidner, S. Mark	01-19
Lejeune, Y.	04-05
Li, Wei	09-01
Liang, Xudong	05-20
Liang, Xudong	06-07
Lin, Shian-Jiann	04-19
Lin, Ying	01-21
MacKenzie, Ian	07-11
Mahfouf, Jean-Francois	01-23
Majewski, D.	03-05
Majewski, Detlef	04-03
Manins, P.	05-16
Manoji, Nobutaka	02-11
Marsigli, C.	05-24
Martin, E.	04-05
Masek, Jan	03-03
Mashiko, Wataru	05-22
Masumoto, Yukio	08-12
Matsumura, Takayuki	03-11
Matsumura, Takayuki	03-19
McFarlane, Norman	03-13
Meng, Jesse	01-25
Merryfield, William	03-13
Mironov, Dmitrii	04-13
Mo, Kingtse	07-13
Mokhov, Igor	02-03
Mokhov, Igor	02-07
Mokhov, Igor	02-09
Mokhov, Igor	07-15
Montani, A.	05-24
Moorthi, S.	04-09
Morey, Steven L.	03-21
Morey, Steven L.	08-08
Morrison, John	08-14
Mueller, Aurelia	04-21
Muller, A.	03-05
Muller, A.	04-03
Murata, Akihiko	04-15
Muroi, Chiashi	05-22
Nakagawa, Masayuki	04-17
Nakamura, Hajime	01-34
Nakamura, Hajime	05-30
Nakamura, Yoshiyuki	01-27
Nasonova, N. O.	04-05
Naumov, A.D.	05-26
Nebuda, Sharon	04-19
Nerozzi, F.	05-24
Nichols, Nancy	01-17
Nojima, Kazuya	01-27
Noone, David	07-17
O'Brien, James J.	02-21
O'Brien, James J.	03-21

Index of Authors

O’Brien, James J.	06-13
O’Brien, James J.	08-08
O’Brien, James J.	08-14
O’Brien, James J.	08-18
Oh’izumi, Mitsuo	04-07
Ohfuchi, Wataru	06-09
Ohhashi, Yasuaki	01-37
Okamoto, Kozo	01-28
Onogi, Kazutoshi	02-11
Paccagnella, T.	05-24
Pailleux, Jean	01-39
Palmer, Tim	06-11
Parkhomenko, V.P.	07-19
Perekhodtseva, E.V.	02-13
Pierce, Clive	05-05
Pokhil, A.E.	05-26
Puri, Kamal	05-16
Pyles, R. D.	04-05
Radakovich, Jon	04-19
Rashid, Harun	02-15
Resnyansky, Yurii	08-20
Rhodin, A.	03-05
Rhodin, Andreas	01-30
Rikus, L.	04-23
Ritter, Bodo	03-05
Ritter, Bodo	04-03
Ritter, Bodo	04-13
Ritter, Bodo	04-21
Rivin, Gdaly	02-17
Roeckner, E.	03-05
Roff, Gregory	02-19
Rogers, Eric	01-21
Rogers, Eric	05-18
Romero, Rosario	02-25
Roulstone, Ian	01-32
Rubinstein, K. G.	02-03
Safray, Arkadiy S.	08-10
Saito, Kazuo	03-15
Sakamoto, Kenji	08-07
Sakamoto, Masami	02-11
Sakuma, Hirohumi	08-12
Salkevar, P. S.	08-05
Sasaki, Hideharu	08-12
Sato, Tetsuya	09-03
Sato, Yoshiaki	01-07
Sattler, Kai	05-28
Schlosser, A.	04-05
Schrodin, Reinhold	04-21
Sclavo, M.	08-01
Seko, Hiromu	01-34
Seko, Hiromu	05-30
Senotrusova, S.	05-32
Sevault, Florence	07-03
Sharp, Ryan	02-21
Shen, Bo-wen	04-19
Shin, Dong-Wook	06-13
Shmakin, A. B.	04-05
Simmonds, Ian	02-15
Simmonds, Ian	02-23
Simmonds, Ian	07-23
Slingo , Julia M.	09-01
Smirnova, T. G.	04-05
Smith, Shawn	02-25
Smolikova, Petra	03-03
Somot, Samuel	07-03
Song, Yucheng	07-13
Sourkova, G.V.	06-15
Sperber, Kenneth R.	09-01
Stevenson, David	07-11
Strasser, U.	04-05
Subramanyam, Bulusu	08-14
Sun, Z.	04-23
Svinuhov, V.	05-32
Takacs, Lawrence	07-07
Takahashi, Keiko	08-12
Takahashi, Keiko	09-03
Takahashi, Kiyotoshi	02-11
Tanguay, Monique	01-36
Tibaldi, S.	05-24
Tkacheva, Yu.V.	05-01
Tkatchenko, Igor V.	08-10
Tokuno, Masami	01-37
Tolstykh, Mikhail	03-17
Tolstykh, Mikhail	04-25
Tory, K.	05-16
Tounkara, Mamadou	01-39
Tracton, Steve	05-09
Tsutsui, Jun'ichi	02-11
Ueno, Mitsuru	04-15
Ueno, Mitsuru	04-27
Verseghy, D.	04-05
Vivoda, Jozef	03-03

Index of Authors

Wada, Akiyoshi	08-16
Wada, Akiyoshi	09-05
Wada, Akiyoshi	09-07
Wagner, Sebastian	07-21
Wanf, Guiling	04-19
Wardle, Richard	07-23
Widmann, Martin	07-09
Wilson, Clive	02-05
X-BAIU-01 observation group	01-07
Xie, Lian	08-14
Yabu, Syoukichi	04-29
Yamazaki, Nobuo	02-11
Yamazaki, T.	04-05
Yang, Z.-L.	04-05
Yoshimura, Hiromasa	03-11
Yoshimura, Hiromasa	03-19
Young, Margaret	05-16
Yu, Peng	03-21
Yu, Peng	08-18
Zavala-Hidalgo, Jorge	03-21
Zavala-Hidalgo, Jorge	08-18
Zelenko, Alexander	08-20
Zhou, Bin bin	05-09
Zhou, Xiaqiong	05-34