VALIDATION AND DIAGNOSIS OF ATMOSPHERIC MODELS

A. Hollingsworth
European Centre for Medium-Range Weather Forecasts
Shinfield Park, Reading, UK

ABSTRACT

We discuss the problems of validating numerical weather prediction models and general circulation models. The methods used for validating general circulation models are used extensively to validate forecast models.

Operational centres make daily forecasts from accurate initial conditions and can verify the forecasts against good quality observations and analyses. Forecast centres have exploited these scientific resources to devise new and effective ways to validate their models and to diagnose errors in the model formulations. For example, operational assimilation systems provide opportunities to study the dependence of forecast errors on particular physical processes. Examples of these new methods of model validation are discussed.

In a separate development, operational forecast data are used to drive a range of other models such as ocean circulation models, ocean wave models, and long-range transport models, each of which can be validated in turn against appropriate oceanographic or chemical data. This work provides striking and valuable indirect validations of the forecast models.

1. INTRODUCTION

This talk discusses methods for validation of atmospheric models. The word 'valid' has a number of definitions including (i) efficacious, or, able to produce the result intended; (ii) well-grounded on principles or evidence; (iii) able to withstand criticism or objection. The word 'validation' means making valid, i.e. confirming the validity of. However meteorologists frequently use the word validation in the wider sense of determining the fitness or suitability for the intended application. In the case of an atmospheric model this involves identifying the strengths and weaknesses of the model, and so identifying those features of its performance which are good and robust for the intended application, and those features where performance is poor for the intended application. Validation of an atmospheric model involves comparisons of model results with observations, analyses, field data, and with results from more refined and accurate models. Each of these sources of information in turn has to validated, so we find ourselves in an elaborate process of cross-validation. Validation of an atmospheric model is essential if one is to base serious decisions on the results of its simulations, whether they are weather forecasts or climate forecasts.

Since the work of Charney and Phillips with filtered models in the 1950s, and the work of Smagorinsky and Miyakoda with primitive-equation models in the 1960s, there have been close relations between the models used for forecasting and for general circulation modelling. The first

GCM was a development of a forecast model. The first primitive equation general circulation models were used as forecast models, with noteworthy success. Some current forecast models have in turn become the basis of successful general circulation models. New formulations in numerics and in the parametrizations of convection, orographic effects, radiation, and clouds are regularly exchanged between the NWP and GCM communities. The current generation of NWP models will contribute much to the next generation of GCMs, and vice versa.

Gates et al. (1990) provide an extensive discussion of the validation exercises undertaken for atmospheric general circulation models (AGCM), in terms of simulations of means and variances of climate parameters. Current medium-range numerical weather prediction models (NWPM), and AGCMs are very closely related to each other in terms of the global domain and the numerical techniques used for the integrations, and in terms of the complexity of the physical parametrizations. The main differences between the two types of models are the different spatial resolutions, and the different lengths of the integration periods used in applications. The standard methods used for validating general circulation models are extensively used to validate forecast models. Through the application of different validation and diagnostic techniques, we are gradually constructing a secure appreciation of the strengths of current models, and of what must be done to improve them.

Model validation has an important diagnostic purpose, namely to relate the deficiencies in performance to different components of model formulation. The most serious simulation errors probably arise from the physical parametrizations, where every aspect of model formulation has limitations. These limitations lead to errors in the simulations. The errors in the simulations interact, sometimes to cancel each other, but frequently to exacerbate each other. This non-linearity can make it very difficult to relate the errors in a long climate run to the aspects of the physical parametrizations which caused the problems. There are many advantages to studying the growth of model errors at an early stage in the forecast when the rotational component of the large scale flow is still accurately modelled.

2. VALIDATION OF NUMERICAL FORMULATIONS

The validation of numerical formulations is usually more straight-forward than the validation of other components of a model; although this is not invariably so. One must have accuracy and stability of the numerical scheme, and these properties can frequently be determined by analytical means. Other aspects of the numerical schemes are also important. Conservation principles have played a major role in the development of numerical methods. Mass and energy conservation is most desirable for GCMs. Small but systematic violations of conservation principles can corrupt long integrations. In fact, strict conservation principles are not satisfied by the many spectral models used as GCMs and NWPMs. This is tolerable because in any time-step the violations of the principles are small and are not systematic.

Economy of a numerical scheme is a major issue. There have been considerable strides in this area in recent times with the introduction of semi-implicit schemes in the late 1960s (*Robert*, 1968), and of semi-Lagrangian methods during the 1980s (*Robert*, 1981). The semi-Lagrangian schemes so far published have not satisfied an energy conservation principle. There is empirical evidence that in some semi-Lagrangian formulations there can be small systematic biases in the energy conversion terms which may cause problems even in 10-day integrations (Simmons, pers comm 1992). However new developments (*Machenhauer* 1992, pers comm) indicate the possibility of implementing conservation principles in a semi-Lagrangian framework.

Hoskins et al. (1985) have discussed the central role played by potential vorticity anomalies in many aspects of atmospheric flow. In a recent synoptic study, *Appenzeller and Davies* (1992) emphasize the importance of potential vorticity dynamics in the interaction of the stratosphere and troposphere near tropopause breaks, and the consequences of these interactions for new tropospheric developments. The treatment of dynamically active and highly structured quantities such as potential vorticity must place heavy demands on the accuracy of a numerical scheme. Little work has been undertaken to design schemes for accurate treatment of potential vorticity advection. Indeed, little is known about the performance of current numerical schemes in this regard. There is a real need to validate the performance of current numerical schemes in this area.

3. THE COMPLEXITY OF THE FORCING AND DISSIPATION FUNCTIONS OF THE ATMOSPHERE

Physical parametrizations are expressions of conceptual models of the effect on the resolved scales of unresolved turbulent, radiative, convective, condensational and cloud processes, where the computed physical effect depends parametrically on the resolved flow and on specified constants. Ideally the functional form of the dependencies, and the numerical values of the constants, should be determined from field data. Unfortunately, there are many uncertainties in the formulation of the physical parametrizations of all models, arising from the paucity of field data and from the uncertainties of the representativity of the field data.

The ultimate driving force for the atmosphere arises from the marked pole-equator gradient in incoming solar radiation, which is not balanced locally by the latitudinally more uniform outgoing thermal radiation. Radiative energy exchange between the atmosphere and space is strongly mediated by clouds both in the visible and in the thermal frequencies. Most of the incoming solar energy is absorbed by the oceans, and then passed to the atmosphere in the form of latent energy. Energy exchange between the atmosphere and the ocean or land surface is affected by mesoscale processes in the planetary boundary layer. Latent energy can be transported horizontally over large distances before being converted to sensible and potential energy by lifting and condensation. Sensible and potential energy can be advected over large distances before being converted into kinetic energy.

Much of the dissipation of atmospheric kinetic energy takes place in the atmospheric boundary layer. We have good field observations and effective theories for the planetary boundary layer in cases where baroclinity and/or mesoscale processes are unimportant. We have neither adequate observations nor adequate theories for the more common situation where these processes are important. Internal dissipation in the atmosphere involves poorly understood processes such as enstrophy cascades (*Fjortoft*, 1953; *Appenzeller and Davies*, 1992).

The interaction of the atmosphere with orography is subtle and pervasive on many scales of motion, from the planetary scales down through lee cyclogenesis, to valley winds. The state of our understanding is rather poor, especially for time-dependent interactions.

The parametrization problem is inherently dependent on the resolution of the forecast model. As the resolution changes, processes previously unresolved become crudely resolved and then adequately resolved. The problems in this progression are exemplified by the need to modify the parametrization of orographic gravity wave drag as the resolution of the model increases. The representation of orographic forcing, surface fluxes and stratiform precipitation should become simpler as the horizontal resolution increases while the treatment of vertical mixing (especially near the tropopause), boundary layer clouds, and radiative transfer should improve with increases in vertical resolution.

The atmosphere controls its own forcing mechanisms and boundary conditions through a variety of highly non-linear processes. It is difficult to model correctly the magnitudes of the forcing and dissipation processes, and even more difficult to model the feedback loops between them. This complexity makes the validation of physical parametrizations a difficult and demanding task.

4. VALIDATION OF PHYSICAL PARAMETRIZATIONS

Many model validations have been based on general circulation statistics for the mean state, for second order moments and for the energy cycle. Parametrizations in forecast models can be validated, and their weaknesses diagnosed, in a number of new ways including validation using detailed operational data, study of the forecast errors as they grow on an accurate basic state, and exploitation of the balance requirements to diagnose forecast errors in terms of errors of individual physical processes. Forecast researchers also use validation against field data and satellite data, validation using highly resolved models, diagnosis using sensitivity analyses, and validation using ocean and other models. It goes without saying that one of the most important research resources in the validation and diagnosis of model performance is the insight and imagination of the researchers themselves.

4.1 The importance of parametrizations for forecast performance and for general circulation simulations

For the forecaster, the accuracy of the forecast is the main consideration behind his research. For the climate scientist the realism of the statistics of the integration is the main consideration. A forecast model that consistently fails to predict blocking phenomena is as seriously deficient for its forecast purpose as a climate model that produces the same type of flow every winter and fails to produce a realistic range of variability.

Tibaldi and Molteni (1990) discuss the operational prediction of winter-time blocking in the ECMWF model in the period 1980-87. They found that the blocking frequency was severely under-estimated in medium-range forecasts, that the model is reasonably skilful if the initial conditions are blocked, but that blocking onset is poorly represented if it occurs more than a few days into the forecast. The inability to enter the blocking regime has a substantial impact on the systematic error of the model. These results raised the question: Was the failure of the blocking forecasts due to limitations in the initial data or due to inherent limitations in the model?

The underlying dynamical problems have been studied in terms of regimes and regime transitions (*Molteni et al.*, 1990; *Ghil and Mo*, 1991; *Vautard*, 1990). *Molteni and Tibaldi* (1990) define weather regimes as clusters of atmospheric states in a low dimensional phase space generated by the leading empirical orthogonal functions (EOFs) of hemispheric eddy geopotential fields. They found that the probability distributions of both the hemispheric forecast error and the amplitude of planetary waves are bi-modal in the region of phase space where the fields have negative or small projection onto an EOF resembling the Pacific North American pattern. In addition to a true systematic error they found a regime-dependent forecast error whose pattern is opposite to the anomaly in the observed regime. They found that the earlier versions of the forecast model tended to relax the forecast flow towards the model's most densely populated regime. Both *Molteni and Tibaldi* (1990) and *Palmer et al.* (1990) use operational forecasts and longer term integrations to show that errors in the forecasts of regimes and regime transitions are sensitive to model formulations, and that the errors have been considerably reduced as a result of model developments, particularly in the area of physical parametrizations.

The capability of a model to simulate the correct probability distribution of weather regimes and the correct distribution of the transitions between weather regimes is just as critical for the forecaster as it is for scientists concerned with accurate simulation of low frequency variability in the general circulation. Both groups are faced with the difficulty that it is easier to document short-comings in model performance than it is to determine how to improve model performance. Forecasters have successfully exploited a number of special aspects of their work to address this problem.

4.2 Validation using detailed operational data

Because NWPMs are used to produce forecasts every day, a range of useful validation techniques are readily available to NWP researchers. Some of these techniques depend on exploiting the results of the data assimilation system.

Data assimilation is a systematic development of the synoptician's analysis method based on time-continuity. The aim of the procedure is to provide the best possible analyses of current and earlier observations as initial data for the forecast. Objective methods have been developed to quality control the observations and to infer as much as possible from the observations using accurate forecasts as an essential element in the process of inference. The forecasts are also a powerful aid in quality control in data rich as well as data sparse areas, and have become the basis of effective global data monitoring under the aegis of WMO.

Apart from its value in improving the forecast, data assimilation is a marvellous learning laboratory. The analyses represent the best synthesis we can make of all available data. Data assimilation provides a continuous confrontation between our understanding (expressed in the model) and the observations; a confrontation which forces us to see and explain the data errors, the model errors, and the weaknesses in our use of the data. If exploited systematically, the results of the confrontation in data assimilation between the observations and our understanding can accelerate growth of the latter.

Because each forecast begins from a rather accurate description of the atmosphere on the resolved scales, each forecast enables one to observe forecast errors growing through a linear phase on an essentially correct basic state. This frequently enables one to isolate particular errors or complexes of errors. The daily verification of parameters such as temperature at 2m, wind at 10m, cloud, or rain, is a powerful tool for exposing shortcomings in parametrization schemes.

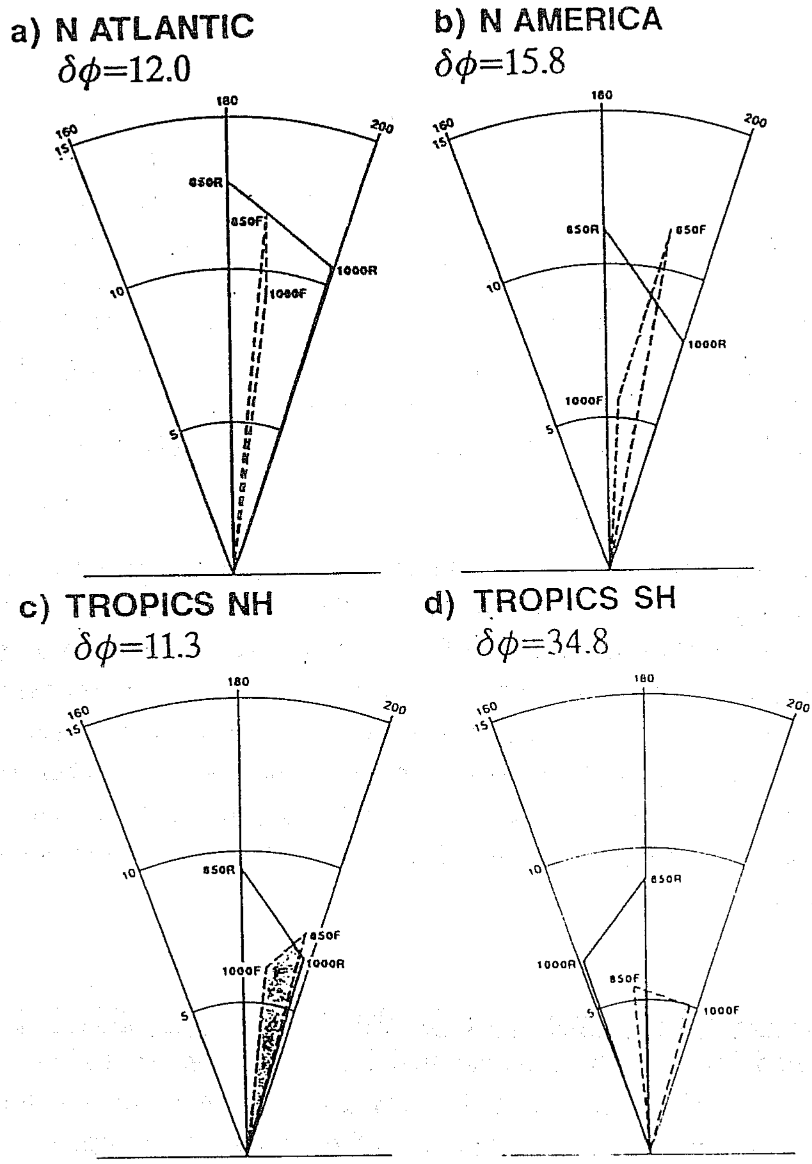
One such investigation concerned the ability of the short-range forecasts to simulate the observed baroclinity of the boundary layer. Figure 1a from a forthcoming study shows a comparison of composited observations and short range forecasts of wind shear across the boundary layer (between the surface and 850mb), in cases where the observed wind backed with height. The composite was made for a set of five radiosonde stations over the North Atlantic for three winter months. The figure shows that the forecast model does not reproduce the observed backing wind shear. Similar results were found in the forecasts over the Great Plains (Fig 1b) of North America, and in the Pacific trade wind zones, both north (Fig 1c) and south (Fig 1d) of the equator. In regimes of veering wind, the model performance is much better.

This example illustrates what may well be a common problem in many models. Much is known from observations and theory about the neutral boundary layer. Results from *Lenschow* (1965) and from more recent experiments (*Lenschow and Agee*, 1976; *Lenschow et al.*, 1980) have indicated serious shortcomings of the simple K-theory used to represent the baroclinic boundary layer in many atmospheric models. Even now, little is understood about the structure of the wind field in the planetary boundary layer under the baroclinic and convective conditions described in Fig 1.

Clearly it is a serious matter requiring further investigation if the model has systematic errors in its ability to represent boundary layer convergence/divergence behind a cold front. This little study is just one of numerous examples where the daily verification of a forecast model against analyses and observations highlights deficiencies in the performance of the parametrization schemes.

4.3 Diagnosis from balance requirements

Investigations of the growth of systematic model errors in the first few hours or days of the forecast has the advantage that one can examine the regional growth of errors while the large scale flow is still fairly accurately modelled. This approach is little used by developers of GCMs,



a) N ATLANTIC $\delta\phi=12.0$

b) N AMERICA $\delta\phi=15.8$

c) TROPICS NH $\delta\phi=11.3$

d) TROPICS SH $\delta\phi=34.8$

Figure 1 Composites (based on a three month sampling period) of mean surface and 850mb winds as observed (labelled 1000R and 850R) and forecast (labelled 1000F and 850F) for four groups of selected radiosonde stations [a, North Atlantic; b, Great Plains of N America; c, the tropical Pacific north of the Equator; d, the tropical Pacific south of the Equator]. The composites are made when the observed surface wind exceeds 5m/s and the observed backing (veering in the Southern Hemisphere) of the wind is between 12 and 27 degrees. Shading indicates that the wind in the forecast veers in the mean. The error in the surface wind direction is given in degrees.

who tend to study model problems when the model's climate statistics are stable, i.e. when the errors are fully developed and non-linear.

The diagnosis of the growth of forecast error arising from errors in orographic forcing (*Wallace et al.*, 1983) is a well-known example of the value of studying the linear growth of forecast errors on an essentially correct basic state. Their proposal for an envelope orography has been widely used. This was just the first of several useful studies of orographic effects using short range forecasts.

The systematic exploitation of balance requirements is a powerful technique for studying the performance of individual parametrizations. Operational data assimilation fields satisfy realistic dynamical constraints. The method of balance requirements uses the near-balance of operational data assimilation fields to estimate the time-mean forcing functions from the adiabatic terms. These forcing functions should be balanced by the diabatic forcing terms. If there is an imbalance between the adiabatic and diabatic terms, as there usually is, one can identify the individual processes contributing most to the imbalance (*Klinker and Sardeshmukh*, 1991). Great care is needed in interpreting the results as the assimilation system can sometimes mask errors in physical forcing. Nevertheless, the technique has been extremely useful in identifying many errors in our forecast model.

Studies of the westerly bias in forecast models in the early 1980s led to the formulation of parametrizations for gravity wave drag (*Boer et al.*, 1984; *Palmer et al.*, 1986). Following this work, there was an implementation in the ECMWF model of a parametrization of gravity wave drag which put more drag in the lower stratosphere than in the lower troposphere. *Klinker and Sardeshmukh* (1991) exploited the balance requirements to discuss the relative magnitudes of gravity wave drag on the lower troposphere and lower stratosphere. Their results (suggesting that the predominant effect was in the lower troposphere) were consistent with other studies with very detailed models. Their work led to a beneficial revision of the parametrization along the lines they suggested.

4.4 Validation using field data and satellite data (TOVS, ERBE, ISCCP, SSM/I)

Comparison of model results with observational data for individual processes is a basic method for developing parametrizations. Since field experiments are of limited duration and spatial extent, are difficult to mount, and are expensive, there are many aspects of model formulation which are impossible to validate in this way. Satellite data give a global view of model performance for a few parameters. This is valuable as a survey tool. However there are numerous cases in which the satellite data can identify a problem which can only be resolved with more detailed in-situ data.

As discussed in the IPCC report (*Houghton et al.* 1990), the treatment of clouds and radiation is a central concern for general circulation modelling. Accurate treatment of clouds and radiation are also of great importance in forecasting research (*Morcrette*, 1990). Two important space-based cloud/radiation datasets have been made generally available in recent years, the ERBE data

(*Barkstrom and Smith*, 1986) and the ISCCP data (*Rossow et al.*, 1988). These global data sets have been used to validate the performance of both GCMs and NWPMs. For example, *Vesperini et al.* (1991) have used the ERBE data to diagnose a variety of problems in one version of our forecast model. Similarly, *Morcrette* (1991) has used the ISCCP radiance data to document problems in the same model's treatment of the diurnal variability of skin temperature and brightness temperature. In this latter paper, the approach of comparing observed radiances with simulated radiances (the 'model to satellite' approach) is shown to be particularly valuable in permitting a penetrating diagnosis and validation while avoiding some of the difficulties involved in using, for validation, retrieved quantities such as cloud amount.

Deficiencies have been identified in the ECMWF humidity analyses by *Eymard et al.* (1989) using SMMR data, and by *Liu and Tang* (1992) using SSM/I data. *Phalippou* (1992) used SSM/I data and field data to study the problem in more detail. He confirmed that there are extensive relative biases between the analysed total precipitable water (TPW) vapour and the estimates from SSM/I data. In the dry sub-tropical highs the analyses have 30-50% more TPW than the SSM/I estimates, while in the moist tropical regions the analyses have 10-20% less TPW than the SSM/I estimates.

To check these results in the dry sub-tropics, Phalippou was able to access FIFE sonde data taken at San Nicolas Island off the California coast during the 1987 FIRE experiment. Fig 2 shows a typical comparison by Phalippou of the sonde humidity ascent at San Nicolas with the corresponding ascent in a research assimilation; the San Nicolas sonde was not made available to the assimilation. It is clear that the assimilation fails to describe the observed very dry mid-tropospheric air above the shallow moist marine boundary layer.

Eyre (1992) compared measured radiances in TOVS channel 11 (a humidity channel whose weighting function peaks around 700mb) with expected radiances based on radiative calculations from short range forecasts. He found results consistent with Phalippou's inferences from SSM/I data, that the mid-tropospheric analyses and short range forecasts are too dry in the tropics and too moist in the sub-tropics. Both *Van de Berg and Schmetz* (1991), and *Schmetz and Van de Berg* (1992) have found similar results using data from the water vapour channel on METEOSAT, which is sensitive to upper tropospheric humidity.

This work clearly documents important problems in modelling the humidity field. It is not clear where the model problem lies. It could be that the model has insufficient vertical resolution in the PBL, or that there are important defects in the PBL parametrization. Equally, the problem could lie in the convective and radiative forcing of the Hadley circulation, resulting in subsidence in the model's sub-tropics which is weaker than observed. Further work is clearly needed.

One might comment that studies of this kind are validations of the analyses rather than the model. However no humidity data other than sonde data was used in the assimilation, so the analysed humidity fields over the ocean are largely controlled by the parametrization schemes. The validations of the analysed fields thus provide a powerful diagnosis of the performance of the parametrizations.

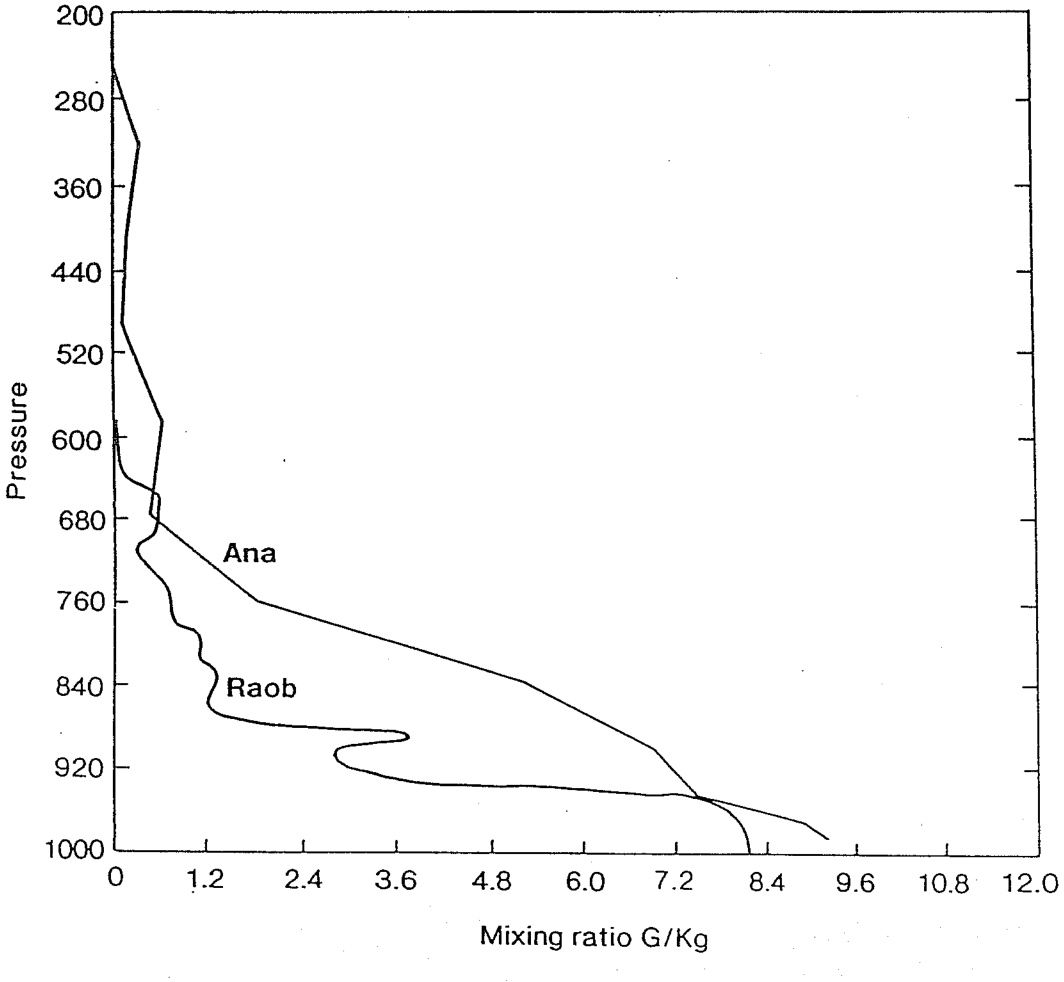


Figure 2 Vertical profiles of humidity mixing ratio (g/kg) at 1200 UTC on 3 July 1987 at San Nicolas Island off the California coast for the radiosonde, and for a research assimilation which did not use the sonde.

4.5 Model intercomparisons

Valuable insight has been derived from controlled inter-comparisons of model components and of particular aspects of over-all model behaviour. Such intercomparisons can provide an effective means of determining the uncertainties in some physical formulations, and of the resulting uncertainties in model results. Radiation parametrizations have been studied with considerable success (*Ellingson et al.*, 1991, *Fouquart et al.*, 1991), because it is possible to study some aspects of the formulations (e.g the clear-sky calculation) in isolation from the rest of the model. So far there has been little successful inter-comparison of convective or boundary layer parametrizations, for example, because these formulations interact strongly with many other formulations in the model.

There have been a number of successful studies of individual aspects of model performance (*Cess et al.*, 1991, *Randall et al.*, 1992). However it is not always easy to separate cause and effect in these studies. A very comprehensive exercise, AMIP (Atmospheric Model Intercomparison Project) is currently under way under the auspices of WGNE. In this project, many modelling groups will produce integrations for the period 1979-1988 using a common prescription for sea-surface temperature. This should provide us with an indication of current capabilities in simulation of the annual cycle and of interannual variability (such as ENSO) originating from fluctuations in sea surface temperature.

4.6 Validation using highly resolved models

The capabilities of large eddy simulation models, and of very high resolution meso-scale models, are increasingly important in the understanding of atmospheric processes, in the interpretation of field data, in the formulation of coarser scale models, and in the validation of the coarser scale models. In many cases the details of processes that must be parametrized are only accessible by simulation. For example, models of this kind have been extensively used in the interpretation of measurements near mountains and in the development and validation of orographic formulations for coarser models (*Bougeault*, 1992).

High resolution cloud-scale (1km) models covering large areas (300-500km) are under development in a number of centres. In addition, large eddy simulation models can be adapted to study baroclinic and convective boundary layers. These models can be validated statistically using field experiment data and will provide a powerful new tool to verify and develop parametrization schemes under various synoptic conditions. Recent reviews (*ECMWF*, 1991; *Wyngaard*, 1992) provide examples of the value of this approach for understanding processes and for developing and validating parametrizations.

4.7 Sensitivity analyses

Sensitivity studies are a useful way to understand the performance of a model. Such studies can be quite informal or rigorously formal. An informal sensitivity study may begin with the observation that the model is mis-behaving in a specific way, and one asks 'can it be made to behave better if it is given a push in a particular direction'? An instructive experiment of this kind is reported by *Miller et al.* (1992). These authors were concerned about the lack of intensity in forecasts of major features of the tropical general circulation. They found marked sensitivity of the tropical features to changes in tropical sea surface temperature. Since the tropical sea surface temperature is accurately known, they interpreted the result as a marked sensitivity to latent heat fluxes, especially in areas of weak winds. Using the sensitivity experiments, and earlier theoretical studies, they developed a revision of the surface latent heat flux over oceans at low wind speeds which gave a substantial improvement in the tropical simulations. There was a frisson of excitement at the TOGA meeting in Hawaii in 1990 when a presentation of the model based inferences about the latent flux was closely followed by quite independent presentations of field measurements confirming the model-based inferences (*Bradley et al.*, 1991; *Khalsa*, pers comm, 1990).

More formal sensitivity analyses will be possible with the adjoints of atmospheric models. Development of the adjoints of the adiabatic components of models is underway at a number of forecast centres. The adjoints of the dynamics have been developed first, while the adjoints of the physical parametrizations will take longer. Software to generate the adjoint of a FORTRAN subroutine will accelerate these developments. The adjoints have been developed initially to calculate the gradients of cost functions appropriate for data assimilation problems. At ECMWF we have evidence that four dimensional assimilation can extract plausible wind information from the TOVS humidity channels by using the humidity as a tracer (*Thépaut*, pers comm 1992). A second application of adjoint methods is to generate the 'optimal modes' for generating efficient perturbations for ensemble forecasts. We plan to use such optimal perturbations of a primitive equation model in a real-time quasi operational trial of ensemble forecasting in the near future (*Palmer et al.*, 1992).

The adjoint of a model can also be used to provide an efficient and detailed analysis of the sensitivity of the model performance to the initial state and to the parameters of the model. Such studies are in their infancy for 'real-size' problems, but we expect progress to be rapid. Potential applications of the sensitivity analyses are manifold. One such application is parameter estimation (*Zou et al.*, 1992). Such parameter estimates are likely to be of considerable value in providing systematic validation of model formulation. The method cannot identify missing processes, but can probably give valuable indications of their absence.

4.8 Implicit validation with oceanic, chemical and hydrological data

Ocean circulation models, ocean surface wave models, ozone transport models, models for long range transport of aerosols & pollutants, and hydrological models, are all driven by output from atmospheric models, either with one-way interaction (slaved mode), or with two-way interaction (fully-coupled). These specialised models are validated against appropriate data (XBT data, wave-height data, ozone data, aerosol data, run-off data). Since many of the specialised models are rather sensitive to the atmospheric forcing, such validation exercises frequently provide implicit validation of the atmospheric forcing functions.

Significant wave-height data from the ERS-1 satellite have been used to validate the wind fields provided by atmospheric models as forcing functions for ocean wave models (*Günther et al.*, 1992). *Anderson* (1992) reviews the validations undertaken with ocean circulation models driven by energy and momentum fluxes from atmospheric models; these studies have identified several problem areas for the atmospheric modellers. *Pyle* (1992) provides many examples where the interplay between scientists involved in ozone field measurement campaigns and those providing meteorological analyses and forecasts have provided useful cross-validation for both disciplines. *Swap et al.* (1992) report validation of NWP wind fields for long range transport calculations, using aerosol and chemical measurements. *Viterbo and Illari* (1993) explore the use of run-off data for the validation of forecast rainfall amounts.

These studies of the analyses and forecasts from an NWP system provide a diverse and broad validation of many aspects of the parametrization schemes that would have been unobtainable a decade ago.

5. DISCUSSION

In a sense, atmospheric models are a statement of the sum of our applicable knowledge and understanding of atmospheric flow. The examples cited in this talk give a flavour of the novel ways in which forecast models are validated and diagnosed. For reasons of time and space we have said little about extensive on-going activities in verification and synoptic evaluation. The current generation of models simulate a very wide range of atmospheric behaviour. Model changes and developments must be evaluated over a similarly wide range of behaviours.

Forecast models have benefitted greatly from studies of field data, satellite data, operational data, and data assimilation fields. Important new trends are developing in the working relations between forecast centres, field experimenters, and satellite agencies. Field data is invaluable to modellers. Forecast centres, operating in real time, can provide valuable information to field experimenters for mission planning. In addition, the value of the NWP analysed fields for analysis of the experimental data can be greatly enhanced if some components of the field data are made available for real-time assimilation. This pattern of data exchange is happening for a growing number of field campaigns, to the benefit of all concerned.

It is also becoming clear that forecast centres, operating in real time, can provide valuable resources for calibration and validation of new and existing satellite data (*Günther et al.*, 1992; *Stoffelen and Anderson*, 1992). In the light of experience with a number of missions, one can now confidently say that if a satellite mission produces data which can be assimilated, then the mission will benefit from real-time validation of the data at NWP centres. The NWP centres cannot provide precision calibration or validation which can only come from detailed and painstaking analysis of the satellite data and in-situ field data. However since the NWP centres have access to vast quantities of data and have sophisticated assimilation systems, they can provide a quick validation which identifies critical areas where further detailed research efforts may usefully be concentrated. In turn, the forecast centres benefit from rapid access to the satellite data. There have been cases where rapid access of this kind would have accelerated progress by several years, because the 'main story' for modellers was evident from a first look at the data, and did not need to wait on the years of work needed to produce superbly polished research-quality datasets.

Validation efforts at forecast centres are driven by the need to improve forecast quality. I have tried to show in this talk that forecast validation efforts have a broad scope, that forecasters have exploited their operational activities to improve the models in imaginative ways, that the forecasters' research benefits from the efforts of many different research communities, and that the forecasters can in turn contribute much to the efforts of those communities.

REFERENCES

Anderson, D.L.T., 1992: Use of ocean models to validate surface flux parametrizations; to appear in Proc ECMWF 1992 Seminar on Validation of Atmospheric Models.

Appenzeller, C., and H.C. Davies, 1992: Structure of stratospheric intrusions into the stratosphere. Nature, 358, 570-572.

Barkstrom, B.R., and G.L. Smith, 1986: The Earth Radiation Budget Experiment: Science and Implementation. Rev.Geophys., 24, 379-390

Boer, G.J., N.A. McFarlane, R. Laprise, J.D. Henderson, J.P. Blanchet, 1984: The Canadian Climate Centre, spectral atmospheric general circulation model. Atmos.-Ocean., 22, 397-429.

Bougeault, P., 1992: Validation of meteorological models and parametrizations with observations of the PYREX field experiment. In: Proc ECMWF 1992 Seminar on validation of Atmospheric models; in press.

Bradley, F.F., P.A. Coppin, J.S. Godfrey, 1991: Measurements of sensible and latent heat flux in the western equatorial Pacific Ocean. J.Geophys.Res., 96, 3375-3390.

Cess, R.D., G.L. Potter, J.-P. Blanchet, G.J. Boer, S.J. Ghan, J.T. Kiehl, H., Le Treut, Z.X. Li, X.-Z. Liang, J.F.B. Mitchell, J.-J. Morcrette, D.A. Randall, M.R. Riches, E. Roeckner, U. Schlese, A. Slingo, K.E. Taylor, W.M. Washington, R.T. Wetherald, and I. Yagai, 1989: Interpretation of cloud-climate feedback as produced by 14 atmospheric general circulation models. Science, 245, 513-516.

ECMWF, 1991: Proc of ECMWF Workshop on fine-scale modelling and the development of parametrization schemes; available from ECMWF

Ellingson, R.G., J. Ellis, S. Fels, 1991: The intercomparison of radiation codes used in climate models: Longwave results. J.Geophys.Res., 96, 8929-8954.

Eymard, L., C. Klapiscz, R. Bernard, 1989: Comparison between Nimbus-7 SMMR and ECMWF model analyses: The problem of the surface latent heat flux. J.Atmos.Oceanic Technol., 6, 866-911.

Eyre, J.R., 1992: A bias correction scheme for simulated TOVS brightness temperatures. ECMWF Technical Memorandum 186, available from ECMWF.

Fouquart, Y., B. Bonnel, V. Ramaswamy, 1991: Intercomparing shortwave radiation codes for climate studies. J.Geophys.Res., 96, 8955-8968.

Fjortoft, R., 1953: On the changes in the spectral distribution of kinetic energy for two-dimensional non-divergent flow. Tellus, 5, 225-235

Gates, W.L., P.R. Rowntree, Q-C. Zeng, 1990: Validation of Climate Models. In: Climate Change - The IPCC Scientific Assessment, ed. J.T. Houghton, G.J. Jenkins, and J.J. Ephraums. Cambridge University Press, 93-130.

Ghil, M., K.C. Mo, 1991: Intraseasonal oscillations in the global atmosphere - Part I: Northern hemisphere and Tropics. J.Atmos.Sci., 48, 752-779.

Günther, H., P. Lionello, P.A.E.M. Janssen, L. Bertotti, C. Brüning, J.C. Carretero, L. Cavaleri, A. Guillaume, B. Hansen, S. Hasselmann, K. Hasselmann, M. de las Heras, A. Hollingsworth, M. Holt, J.M. Lefevre, R. Portz, 1992: Implementation of a third-generation ocean wave model at the European Centre for Medium-Range Weather Forecasts. ECMWF Technical Report 68, available from ECMWF.

Hoskins, B.J., M.E. McIntyre and A W Robertson, 1985: On the use and significance of isentropic potential vorticity maps. Quart.J.Roy.Meteor.Soc., 111, 877-946.

Houghton, J.T., G.J. Jenkins, and J.J. Ephraums (eds), 1990: Climate Change - The IPCC Scientific Assessment. Cambridge University Press, xxxix + 1-365.

Klinker, E., and P.D. Sardeshmukh, 1991: The diagnosis of mechanical dissipation in the atmosphere from larg-scale balance requirements. J.Atmos.Sci., 49, 608-627.

Lenschow, D.H., 1965: Airborne measurements of atmospheric boundary layer structure. Dept. of Meteorology, Univ. Wisconsin, Final Report 1965, Contract DA-36-039-AMC-00878(E): Studies of the effect of variations in boundary conditions on the atmospheric boundary layer, ed. H.H. Lettau, 146-208.

Lenschow, D.H., E.M. Agee, 1976: Preliminary results from the air-mass transformation experiment (AMTEX). Bull.Amer.Meteor.Soc., 57, 1346-1358.

Lenschow, D.H., J.C. Wyngaard and W.T. Pennell, 1980: Mean-field and second moment budgets in a baroclinic convective boundary layer. J.Atmos Sci., 37, 1313-26.

Liu, W.T., and W. Tang, 1992: Precipitable water and surface humidity over global oceans from Special Sensor Microwave Imager and European Centre for Medium Range Weather Forecasts. J.Geophys.Res., 97, C2, 2251-2264.

Miller, M.J., A.C.M. Beljaars, and T.N. Palmer, 1992: The sensitivity of the ECMWF model to the parametrization of evaporation from the tropical ocean. J.Climate, 5, 418-434.

Molteni, F. and S. Tibaldi, 1990: Regimes in the wintertime circulation over northern extratropics II: Implications for Predictability. Quart.J.Roy.Meteor.Soc, 116, 1263-1288.

Molteni, F., S. Tibaldi, and T.N. Palmer, 1990: Regimes in the wintertime circulation over northern extratropics I: Observational Evidence. Quart.J.Roy.Meteor.Soc, 116, 31-68.

Morcrette J-J., 1990: Impact of changes to the radiation transfer parametrizations plus cloud optical properties in the ECMWF model. Mon.Wea.Rev., 118, 847-873.

Morcrette, J-J., 1991: Evaluation of model-generated cloudiness: Satellite-observed and model-generated diurnal variability of brightness temperature. Mon.Wea.Rev., 119, 1205-1224.

Palmer, T.N., G.J. Shutts, and R. Swinbank, 1986: Alleviation of a systematic westerly bias in general circulation and numerical weather prediction models therough an orographic gravity wave drag parametrization. Quart.J.Roy.Meteor.Soc, 112, 1001-1039.

Palmer, T.N., Č. Branković, F. Molteni, and S. Tibaldi, 1990: Extended-range predictions with ECMWF models: Interannual variability in operational model integrations. Quart.J.Roy.Meteor. Soc, 116, 799-834.

Palmer, T.N., F. Molteni, R. Mureau, R. Buizza, P. Chapelet, J. Tribbia, 1992: Ensemble Prediction, ECMWF Technical Memorandum No. 188.

Phalippou, L., 1992: Comparisons between SSM/I and ECMWF total precipitable water. To appear in Proceedings of the Conference on "Microwave radiometry and remote sensing" Boulder, CO; 14-16 January 1992, ed. R Westwater; pp 22-26

Pyle, J., 1992: Validation of models using EASOE data. To appear in: Proceedings of the ECMWF 1992 Seminar on Validation of Atmospheric Models.

Randall, D.A., R.D. Cess, J.-P. Blanchet, G.J. Boer, D.A. Dazlich, A.D. Del Genio, M. Deque, V. Dymnikov, V. Galin, S.J. Ghan, A.A. Lacis, H. Le Treut, Z.-X. Li, X.-Z. Liang, B.J. McAvaney, V.P. Meleshko, J.F.B. Mitchell, J.-J. Morcrette, G.L. Potter, L. Rikus, E. Roeckner, J.-F. Royer, U. Schlese, D.A. Sheinin, J. Slingo, A.P. Sokolov, K.E. Taylor, W.M. Washington, R.T. Wetherald, I. Yagai, and M.-H. Zhang, 1992: Intercomparison and interpretation of surface energy fluxes in atmospheric general circulation models. J.Geophys.Res., 97, No.D4, 3711-3724.

Robert, A., 1968: The integration of a low-order spectral form of the primitive meteorological equations. J.Meteor.Soc.Japan, 44, 237-244.

Robert, A., 1981: A stable numerical integration scheme for the primitive meteorological equations. Atmos.-Ocean, 19, 35-46.

Rossow, W.B., L.C. Garder, P.J. Lu, and A. Walker, 1988: International Satellite Cloud Climatology Project (ISCCP). Documentation of cloud data. WMO/TD-266. World Meteorological organisation, Geneva Switzerland, 123pp.

Schmetz, J., L. Van de Berg, 1992: On the relationship between observed upper tropospheric moisture and cloud-radiation fields. To appear in Proceedings of the International Radiation Symposium, Tallinn, 1992.

Swap, R., M. Garstang, S. Greco, R. Talbot, and P. Kållberg, 1992: Saharan dust in the Amazon Basin. Tellus, 44B, 133-149.

Stoffelen, A. and D.L.T. Anderson, 1992: ERS-1 scatterometer calibration and validation activities at ECMWF: A. the quality and characteristics of the radar backscatter measurements. Proc of Central Symposium of the International Space Year Conference, Münich 1992, pub ESA SP-341.

Tibaldi, S. and F. Molteni, 1990: On the operational predictability of blocking. Tellus, 42A, 343-365.

Van de Berg, L. and J. Schmetz, 1991: Meteosat monitors humidity in the Earth's troposphere. ESA Bulletin, No.67, 84-88.

Vautard, R., 1990: Multiple weather regimes over the north atlantic. Analysis of precursors and successors. Mon.Wea.Rev., 118, 2056-2081.

Vesperini, M., J-J. Morcrette, and Y. Fouquart, 1991: Simulation of cloud radiative forcing with the ECMWF model. Dyn.Atmos & Oceans, 16, 85-109.

Viterbo, P., and L. Illari, 1993: The impact of runoff changes on surface and near surface parameters in a general circulation model. J.Hydrol, in press.

Wallace, J.M., S. Tibaldi, and A.J. Simmons, 1983: Reduction of systematic forecast errors in the ECMWF model through the introduction of an envelope orography. Quart.J.Roy.Meteor.Soc., 109, 683-718.

Wyngaard, J.C., 1992: Atmospheric Turbulence. Annual Rev.Fluid Mech., 24, 205-233.

Zou, X., I.M. Navon, F.X. LeDimet, 1992: An optimal nudging data assimilation scheme using parameter estimation. Quart.J.Roy.Meteor.Soc., accepted for publication.