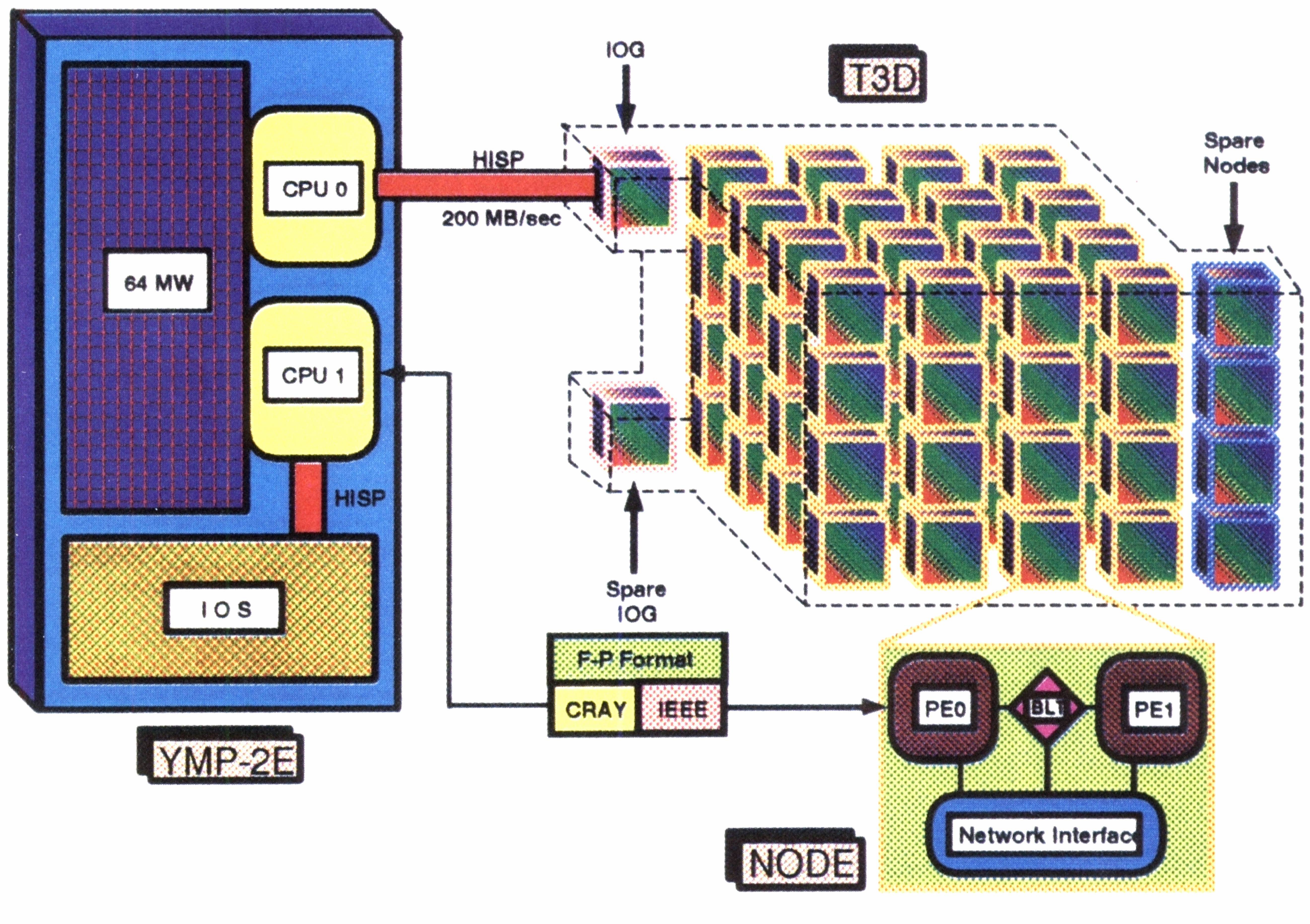
ECMWF Newsletter

Number 67 - Autumn 1994



ECMWF Fig 1 T3D & YMP-2E Configuration Diagram

Shinfield Park, Reading, Berkshire, RG2 9AX, England. Telephone: U.K. (01734) 499000
International: (+44 1734) 499000, Telex: 847908 ECMWF G, Fax: (01734) 869450

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen terme

IN THIS ISSUE

Editorial 1

METEOROLOGICAL

Changes to the operational forecasting system 2
Weather and wave prediction over the oceans 3
Additional data sets available from ECMWF 16

COMPUTING

The CRAY T3D at ECMWF 17
Security considerations of Member State access to the Centre's UNIX services 27

COMPUTER USER INFORMATION

StorageTek upgrade to 36-track tapes 34
The use of 8mm data cartridges at ECMWF 36
Computer user training course 40

GENERAL

ECMWF calendar 1994 42
ECMWF publications 42
Index of still valid Newsletter articles 44
Useful names and 'phone numbers within ECMWF 47

COVER: The CRAY T3D and YMP-2E configuration: see article page 17

This Newsletter is edited and produced by User Support.

The next issue will appear in Winter 1994-95.

Telephone No.: CHANGE OF AREA CODE

Please note that the telephone number for ECMWF is now

National	01734 499000
International	+44 1734 499000

Readers will find in this issue an introduction to the CRAY T3D newly-installed at the Centre. This article deals with the hardware of the new configuration, and an article treating the software aspects will follow in a later issue of the Newsletter.

Considerable thought has been devoted to the security of the Centre's computer system, and the results of the researches into security systems available, and the Technical Advisory Committee's deliberations, are presented here.

A report is made on the progress of work on the prediction of ocean waves, and their interaction with weather. This work is now being carried out as an Optional Project based at ECMWF.

Computer users will also find some information on the recent upgrade by StorageTek to 36mm tapes, and the use of 8mm data cartridges.

* * * * * * * * * * * *

CHANGES TO THE OPERATIONAL FORECASTING SYSTEM

Recent changes

On 4 July 1994 the model soil humidity was reset to field capacity to correct for an excessive drying of the soil over parts of Europe and Asia.

On 12 July two data types were introduced into the data assimilation: temperature observations from aircraft, and winds derived from the METEOSAT water vapour channel.

The post-processing of 10m winds over land was changed on 23 August, resulting in a more realistic representation of model winds at observing sites. The negative bias of the wind speed has been significantly reduced, in particular during daytime.

\- Bernard Strauss

* * * * * * * * * * * * *

WEATHER AND WAVE PREDICTION OVER THE OCEANS

Introduction

An important application of the Centre's medium-range forecast is that associated with marine activities. Shipping, fisheries and offshore operations, for example, are all strongly dependent on weather and require marine weather forecasting extending to the limit of the medium-range forecasting period. Of course, an important element of the marine weather is the sea state and in this article we address the present status of ocean wave modelling at ECMWF. Apart from the direct practical benefits, we will present some examples which show that ocean wave information can also give benefits for atmospheric modelling and data assimilation.

Ocean wave prediction

Interest in wave prediction started during the Second World War because of the practical need for knowledge of the sea state during landing operations. First attempts to predict ocean waves were based on empirical relations between wave height, wind speed and duration. In the fifties, some important steps towards a better understanding of wave evolution were taken. It was recognised that the fundamental quantity to describe waves was the wave spectrum, which was governed by the energy balance equation. The net source function governing the rate of change of the wave spectrum was in principle known after the work of Phillips (1957) and Miles (1957) on wave generation by wind and the work of Hasselmann (1962) on the nonlinear transfer between ocean waves. In deep water the physics of wave evolution was therefore determined by wind input, nonlinear transfer and dissipation by white capping. Experimental campaigns during the sixties and seventies (notably Jonswap (1973), Mitsuyasu et al (1968, 1969) and Snyder et al (1981)) confirmed this picture of wave evolution.

The earliest wave models, however, did not compute the wave spectrum from the full energy balance equation because the limited computer power in those days precluded the use of the nonlinear transfer. This changed with the introduction of vector machines. Furthermore, it became clear that the first wave models were unable to model wave evolution properly in cases of rapidly varying winds. Thus, in the early eighties, a group of mainly European scientists (who called themselves the WAM (WAve Model) group) started working on the development of a new wave prediction model that is based on an explicit description of the physical processes governing wave evolution, including wave-wave interactions. Additional stimulus was provided by the advent of ocean observing satellites such as GEOSAT and ERS-1, which gave high quality wave height observations on a global scale.

Development of the WAM model proceeded rapidly and work on ocean wave modelling at ECMWF began in 1986 in the context of a special project. Routine wave forecasting, using ECMWF forecasts of the low-level wind, started in March 1987. Subsequently, there has been a dedicated effort to optimise the wave model and to verify wave model products against satellite and buoy data, while technical problems, such as archiving in MARS, have also been addressed. In the acknowledgement is a list of people who have contributed to the successful completion of the WAM model development here at ECMWF.

Results of the quasi-real time experiments were found to be most encouraging and in 1991 it was decided to start an optional wave project (supported by 14 Member States) with the object of producing 10-day forecasts for the globe (with a 3° resolution) and 5-day forecasts for the Mediterranean (with a 50 km resolution). Operational forecasting started in June 1992 and a high quality wave product is produced that verifies well against buoy and satellite observations. Since then the wave analysis has been improved by assimilating altimeter wave heights by means of the Optimum Interpolation method, and the spatial resolution of the global model has been increased by a factor of 2 (from 3° to 1.5°). An example of wave model output, the wave height map of the North-east Atlantic at the time of the disaster with the ferry Estonia, is shown in Fig. 1. Altimeter wave heights from ERS-1 are presented as well. Although the resolution for the Baltic is rather crude we observe that at midnight of 27 September the significant wave height was only two metres in the disaster area.

The quality of the wave analysis is continuously monitored by comparing significant wave heights with observed wave height for a number of buoys in the open ocean. These buoys provide independent information since they have not been used in the wave data assimilation scheme. Fig. 2 (from Hansen et al, 1994) shows for the period 19 April - 19 May 1994 the comparison between the modelled wave height and the wave height observed from a number of buoys around the United Kingdom. Also, the number of observations and some statistical parameters are given. This comparison suggests that the WAM model driven by ECMWF 10m winds is performing well as does the comparison of first-guess wave heights with Altimeter wave heights from ERS-1 for the whole globe which is shown in Fig. 3. The small bias of about 20 cm and a standard deviation of only 50 cm suggest that we now deliver a high-quality product which will be of value in providing good marine weather forecasts.

These are but a few examples of the performance of the WAM model. During the course of time the WAM group has studied results of the WAM model in many different circumstances, e.g. extra-tropical storms, hurricanes, and also swell propagation over the Pacific. A complete account of this work may be found in the recently published book by Komen et al, where the theoretical background of wave modelling and the assimilation of satellite data is also addressed.

Apart from the practical advantages already mentioned, ocean wave information can give benefits for atmospheric modelling and data assimilation. This is not surprising as from the physical point of view there is a strong interaction between wind and waves. In particular, wave results have been used to diagnose planetary boundary layer problems, to obtain a consistent energy balance at the ocean surface, and to interpret satellite data. Finally, wave information may also be of value for the atmospheric state when assimilated into a coupled ocean-wave, atmosphere model. Some of these benefits are discussed in the following sections.

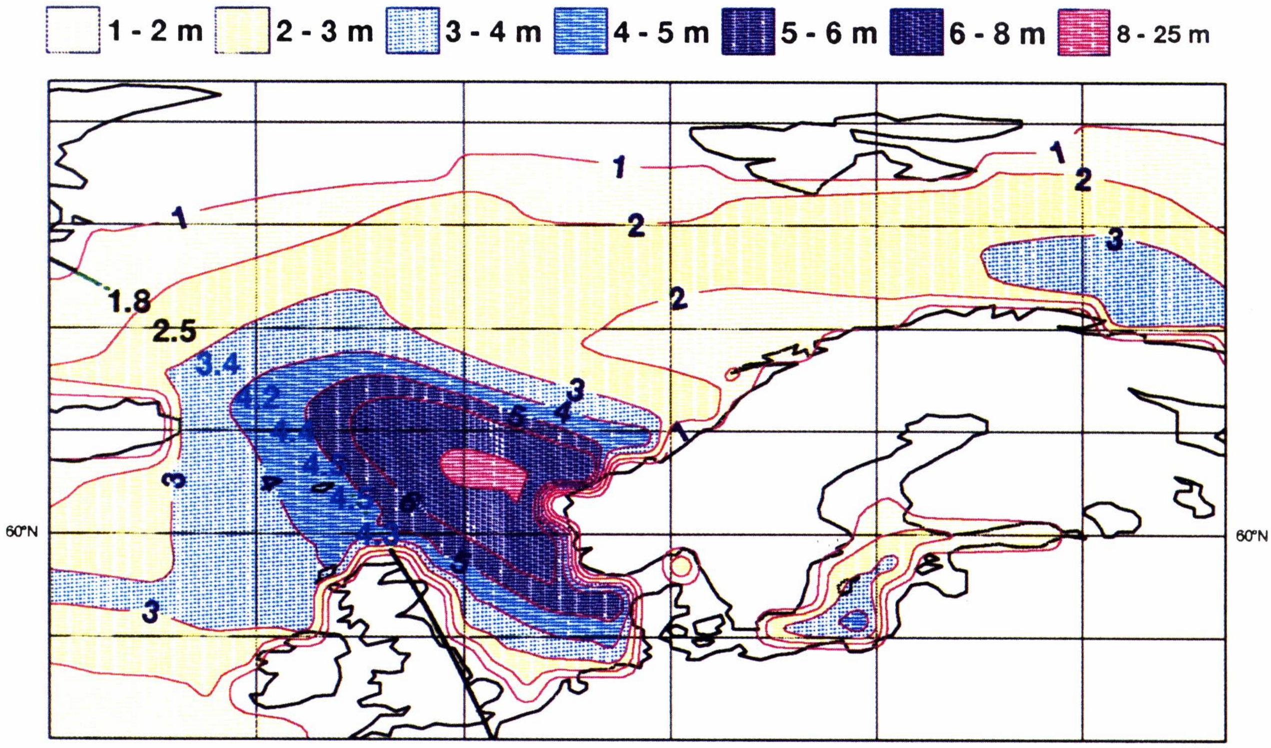


Fig. 1: **Analysed wave height map for North Eastern Atlantic at 28 September 1994 midnight. The ERS-1 satellite track and Altimeter wave height observations around that time are shown as well.**

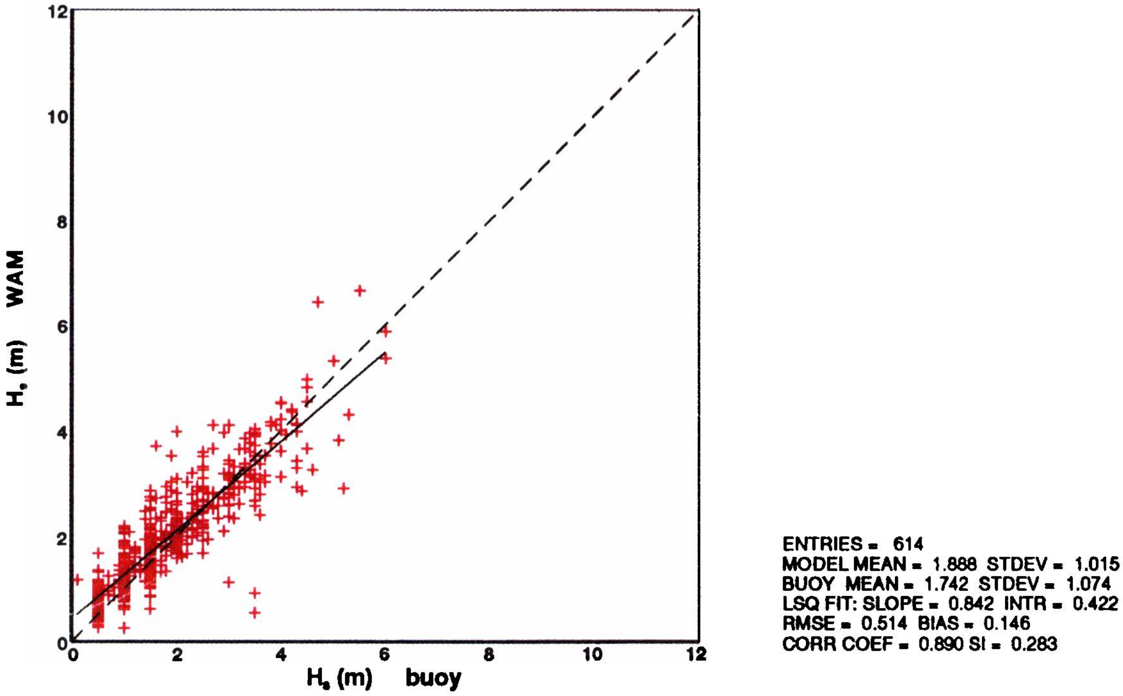


Fig. 2: **Collocation of modelled and observed buoy wave heights for period of 19 April - 19 May 1994. Area around U.K.**

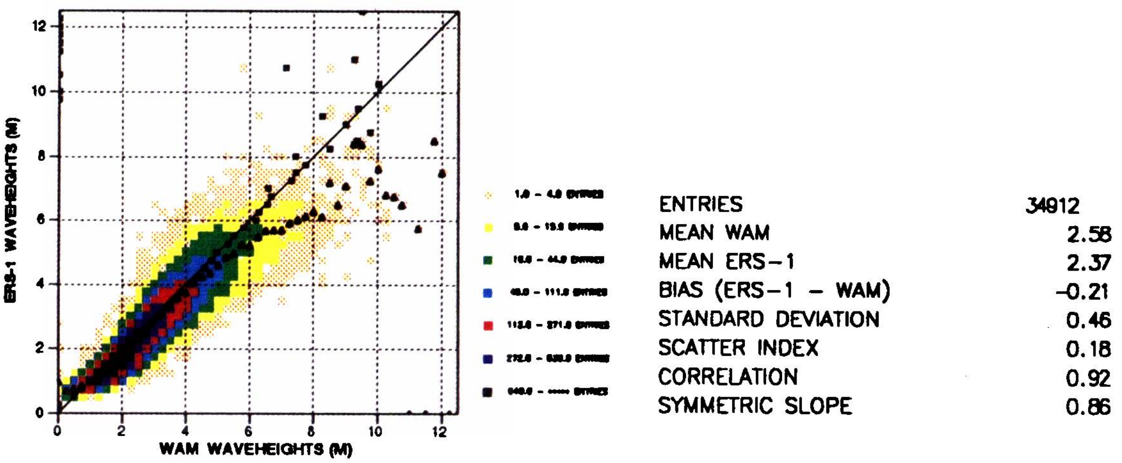


Fig. 3: **Comparison of ERS-1 Altimeter wave heights versus first-guess modelled wave heights for the whole globe. Period same as Fig. 2.**

Planetary boundary problems

If one were to ask a wave modeller to name one of the key problems in ocean wave prediction, the reply would immediately be the driving wind field. This does not mean that all the problems of wave forecasting have been completely resolved (consider dissipation due to wave breaking) but certainly a major source of error in wave heights is the error in the wind field. We know from observations that in a steady state wave height is proportional to the square of the surface wind field. Thus, an accurate wave prediction requires an accurate wind speed prediction. It is mainly because of this that waves may contain relevant information about the weather over the oceans.

Fig. 4 gives an example of this, showing the forecast skill of the WAM model for the period 19 April - 19 May 1994. Here, Southern Hemisphere forecasts are compared to the verifying analysis and the mean error and standard deviation of the forecast, the standard deviation of the persistence and the mean of the verifying analysis are presented. Comparing the standard deviation of the forecast with the persistence, it can be seen that the forecast skill of waves is rather good, but that the wave forecast shows a considerable bias of about 60 cm at day 8. Compared to the verification results with buoy data and altimeter data, this bias is large. This suggests that the forecast winds show a marked bias when compared to the analysed winds. The positive forecast error might be explained by too weak an analysis and a correct atmospheric model or by a correct analysis and too active an atmospheric model. However, since the analysed wave fields compare well with observations, it can be concluded that the wind analysis is correct and hence that the atmospheric model seems to be too active in the later stages of the forecast, especially in the Southern Hemisphere.

Wind-wave interaction

When ocean waves are generated by wind, a considerable amount of momentum and energy is transferred from the atmosphere to the ocean. In meteorology this momentum transfer has traditionally been parametrised in terms of the local windspeed only. There is now ample empirical evidence (Donelan et al (1993), Hexos (1992)) that this picture of momentum transfer is not adequate. In general, the air-sea momentum transfer depends on the history of the wind and the wave field. For example, when ocean waves are only generated by wind (we call this 'young' wind sea) the ocean waves are much steeper than in the case of 'old' wind sea. Steeper ocean waves provide a rougher surface and therefore give rise to a bigger momentum transfer. In other words, the air-sea momentum transfer is sea-state dependent. Note that close to the coast for offshore wind conditions ocean waves may be regarded as 'young' as well and therefore in coastal areas the ocean surface is on average rougher than in the open seas.

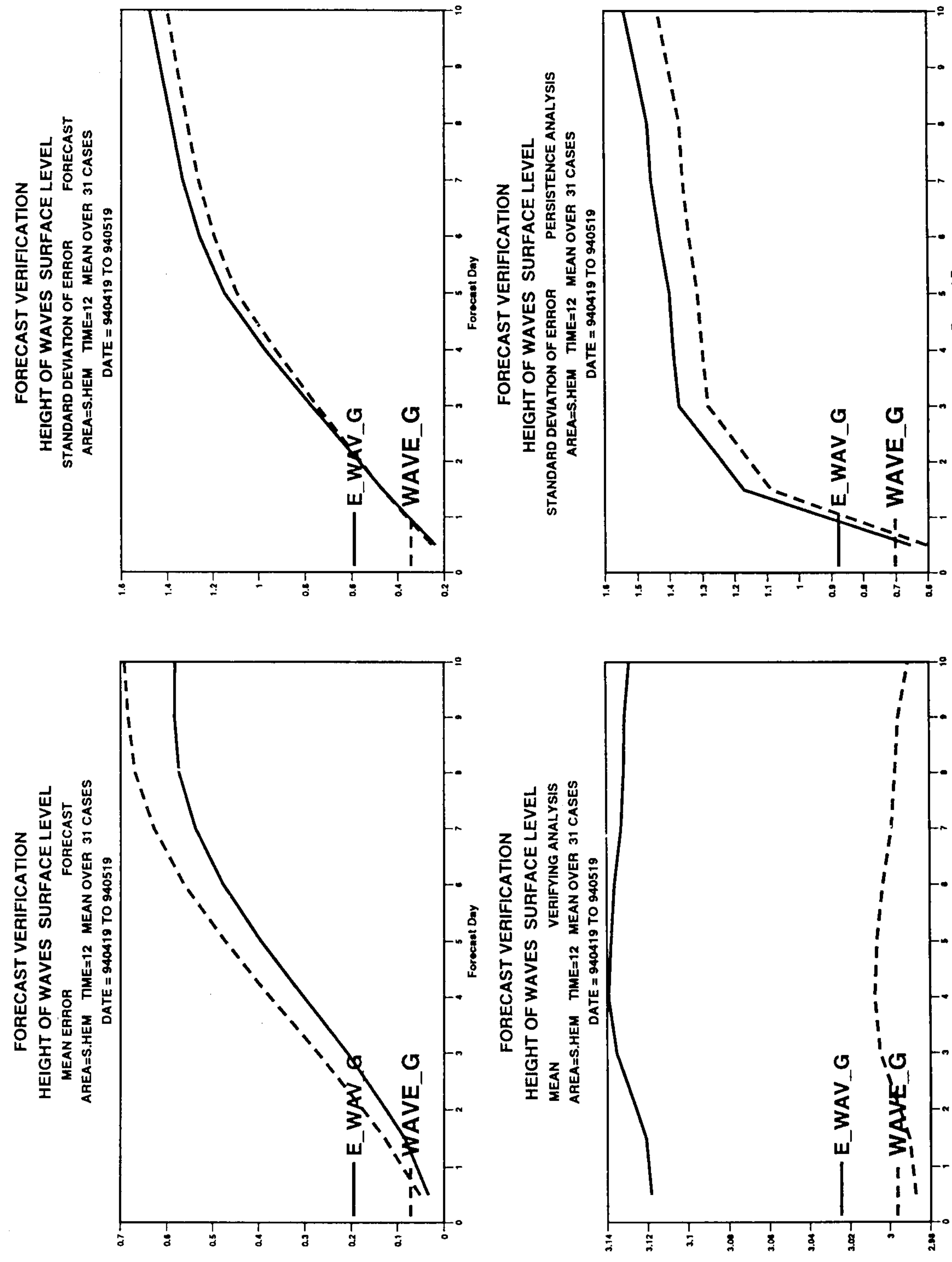


Fig. 4: Forecast verification against verifying analysis for Southern Hemisphere.

The sea-state dependent momentum transfer may have important consequences for the evolution of weather systems, especially in extreme conditions. In order to investigate this issue, Doyle (1994) coupled a high-resolution meso-scale atmospheric model with the WAM model. This enabled him to obtain a consistent momentum balance over the ocean surface as the momentum loss from atmosphere to ocean waves may be determined using the WAM model. The consequences of the sea-state dependent momentum transfer are shown in Fig. 5. The evolution of the central pressure of a synthetic depression for the coupled model is compared with the pressure trace from a conventional model with a momentum transfer that only depends on windspeed (called control). It is clear that the differences in surface pressure are not inconsiderable (± 7 mb) and that they already occur after 2½ days. Apparently, the enhanced roughness of the ocean surface gives rise to an increased surface friction, leading to an increase of the cross-isobar flow which results in enhanced filling up of the low.

Of course, this is only one example where ocean waves may have a considerably impact on weather evolution, but it turns out that the sea-state dependent momentum transfer also has a systematic impact on the climate of the atmospheric model. To demonstrate this we have performed a number of extended-range forecasts for the Northern Hemisphere winter with the coupled WAM-ECMWF model, and we compared the ensemble mean of the 90-day average of 500 mb height with results from the standard ECMWF model. Results are displayed in Fig. 6 where, for reference, we also present the mean of the analysis over the years 1986-1990. In the difference plot we have denoted areas which have a difference which is twice as large or larger than the local variance by red shading. There are considerable changes in the climate over the Pacific (+ 60m), Europe (+ 40m) and Siberia (- 50m) and, in fact, the coupled simulation produces a climate which is in better agreement with the analysis than the control simulation.

Satellites and data assimilation

Research on, and development of, assimilation schemes for wave data are in a relatively early stage, certainly when compared to atmospheric data assimilation. There are two reasons for this. In contrast to the atmospheric problem, ocean waves are more sensitive to forcing (by wind) than to the initial conditions. Thus, if one forces a wave model by means of analysis winds from an atmospheric model, a reasonably reliable 'analysed' wave field is already obtained. The second reason for the slow development of wave data assimilation schemes has been the lack of global wave height data. This situation has changed dramatically in recent years with the launch of satellites such as GEOSAT and ERS-1. For example, ERS-1 now provides a wealth of data on the sea-state (wave heights and winds (Altimeter) and wave spectra (SAR)) and the low-level wind (scatterometer).

It should be realised, however, that the instruments on board these satellites are not directly measuring the geophysical parameters of interest (except perhaps the Altimeter wave heights) but are measuring related quantities. The scatterometer, for example, measures the state of the high-frequency ocean-wave spectrum, and although these waves are certainly related to the geophysical

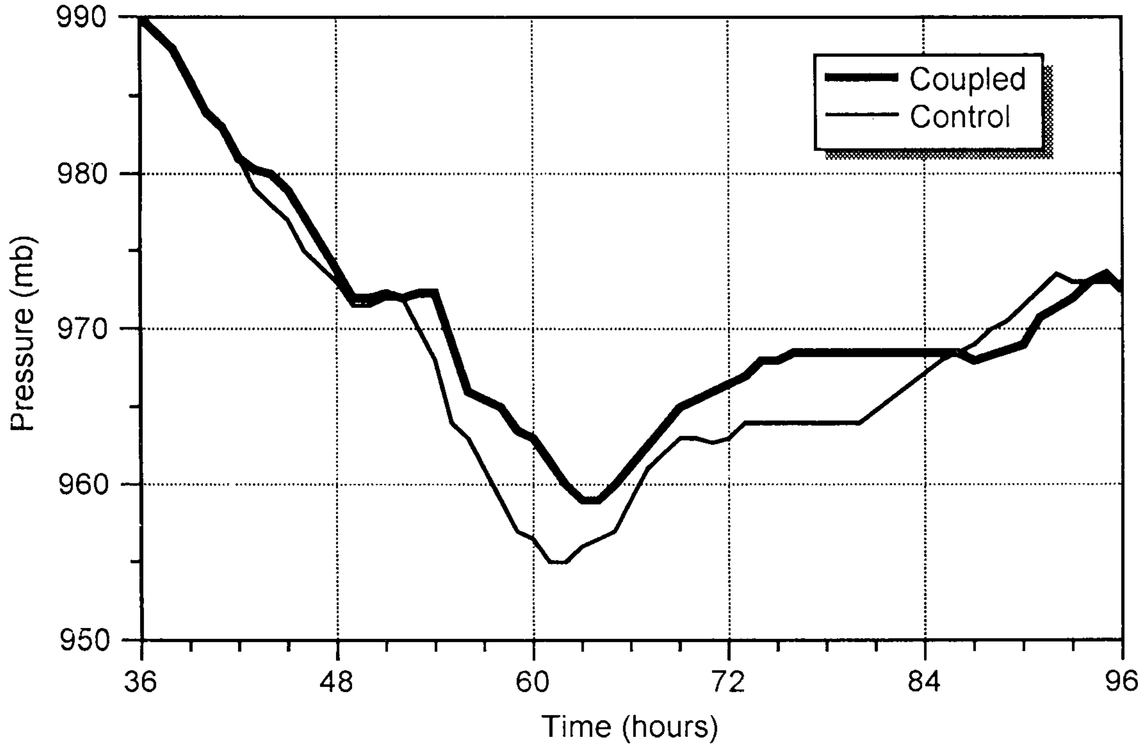


Fig. 5: **Comparison of time evolution of central pressure of a 'synthetic' cyclone for coupled and control simulation (from Doyle (1994) with permission).**

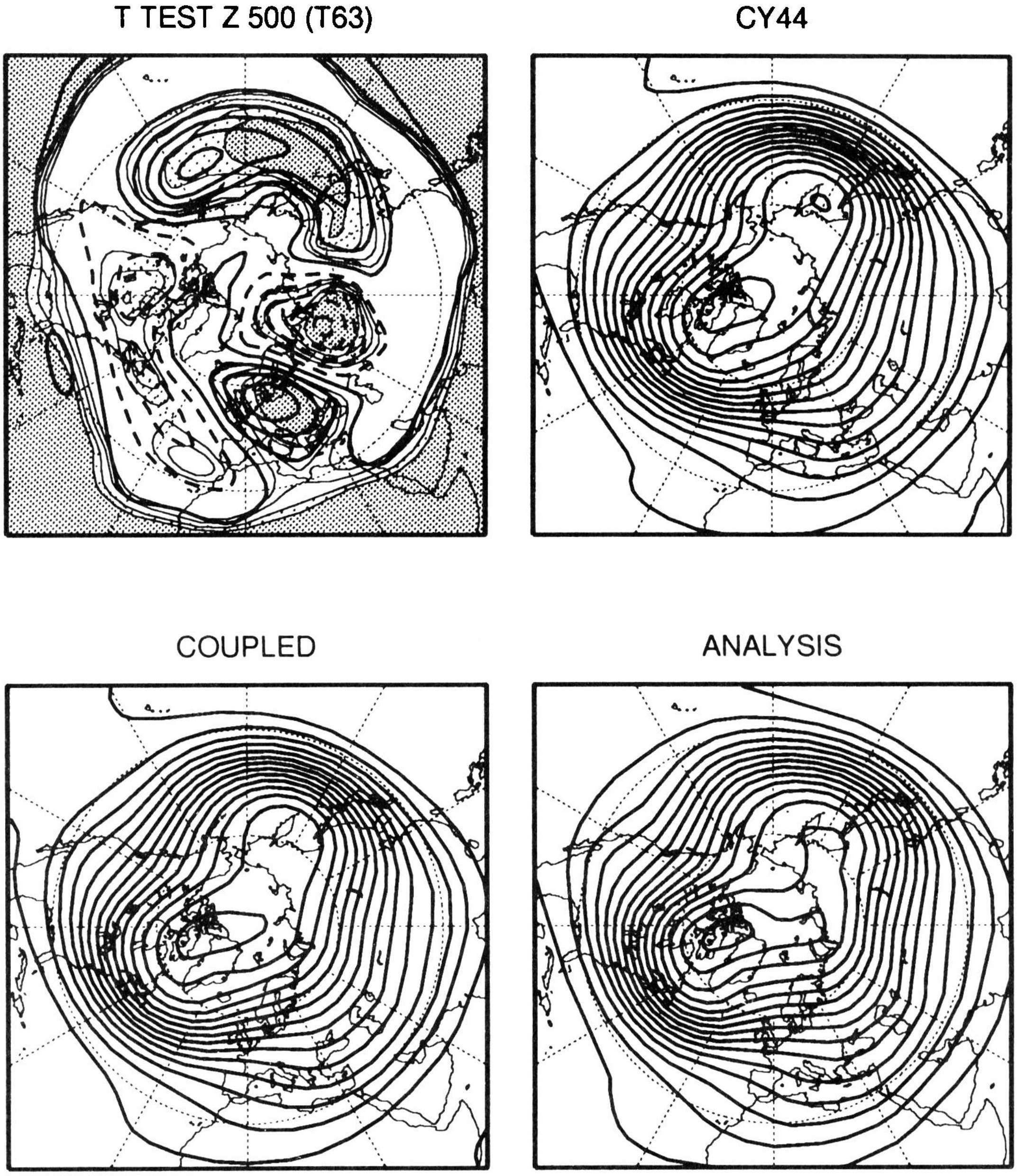


Fig. 6: Ensemble mean of 500 mb geopotential height of coupled and control run and their differences. For comparison also the mean analysed winter climate over the period 86-90 is shown. Period is winter 1990 and area is Northern Hemisphere.

quantity of interest, namely the surface wind vector, other factors, such as the age of the ocean waves, will determine the state of the high-frequency waves as well (Viers (1993)). Similarly, a nonlinear relation exists between the measured SAR image spectrum and the ocean wave spectrum. The wave spectrum can therefore only be obtained from the SAR image spectrum by means of an iteration process where the modelled wave spectrum is used as a first-guess (Hasselmann and Hasselmann (1991)).

Let us assume for the moment that (with help of modelled wave data if needed) we have observed wave data available. Can we use these wave data to obtain information about the weather over the oceans? The answer to this question is yes and thus far the most convincing support of this claim has been given by the work of de las Heras (1994), who has made a study of variational wave data assimilation in the relatively simply case of the generation of ocean waves by wind. As control variable the windspeed at 10m height was taken. The cost function was quadratic and penalised deviations between observed and analysed wind field under the constraints imposed by the WAM model (this, of course, required its adjoint). True and first-guess wave spectra were obtained by summing the single grid point version of the WAM model for one day with a windspeed of 18 and 12 m/s respectively. Wave heights of the 'true' model run were regarded as observations which were assimilated every three hours. No wind data were assimilated. Fig. 7 shows the resulting analysis and the comparison with the first-guess and reality for a number of parameters. The agreement is impressive for windspeed, friction velocity and the mean frequency. The approach was equally successful for turning wind sea cases (provided the mean wave direction was used in the assimilation) and for cases of simulated observed wave heights not generated by the WAM model itself. The work of de las Heras (1994) therefore shows that a variational approach seems promising. Clearly, wave data alone may provide information on the surface windspeed and stress. Combined with the recent results of Thépaut et al (1993), it follows that wave data may even give beneficial information on the atmospheric state throughout the troposphere.

Conclusion

An important element of the weather over the oceans is the sea state. Ocean wave prediction has enormous practical benefits for a number of human activities at sea. In addition, we have seen that wave information may provide important information on the weather of the oceans. Wave results are useful in the analysis of planetary boundary layer problems. Also, the associated momentum transfer to ocean waves may be important for the evolution of (extreme) weather systems, while it also plays a role in reducing systematic errors of the climate. Finally, in the context of 4DVar, observed wave data may even give beneficial information on the atmospheric state throughout the troposphere, while modelled wave data may be of help in interpreting satellite data from ERS-1, ERS-2, ...

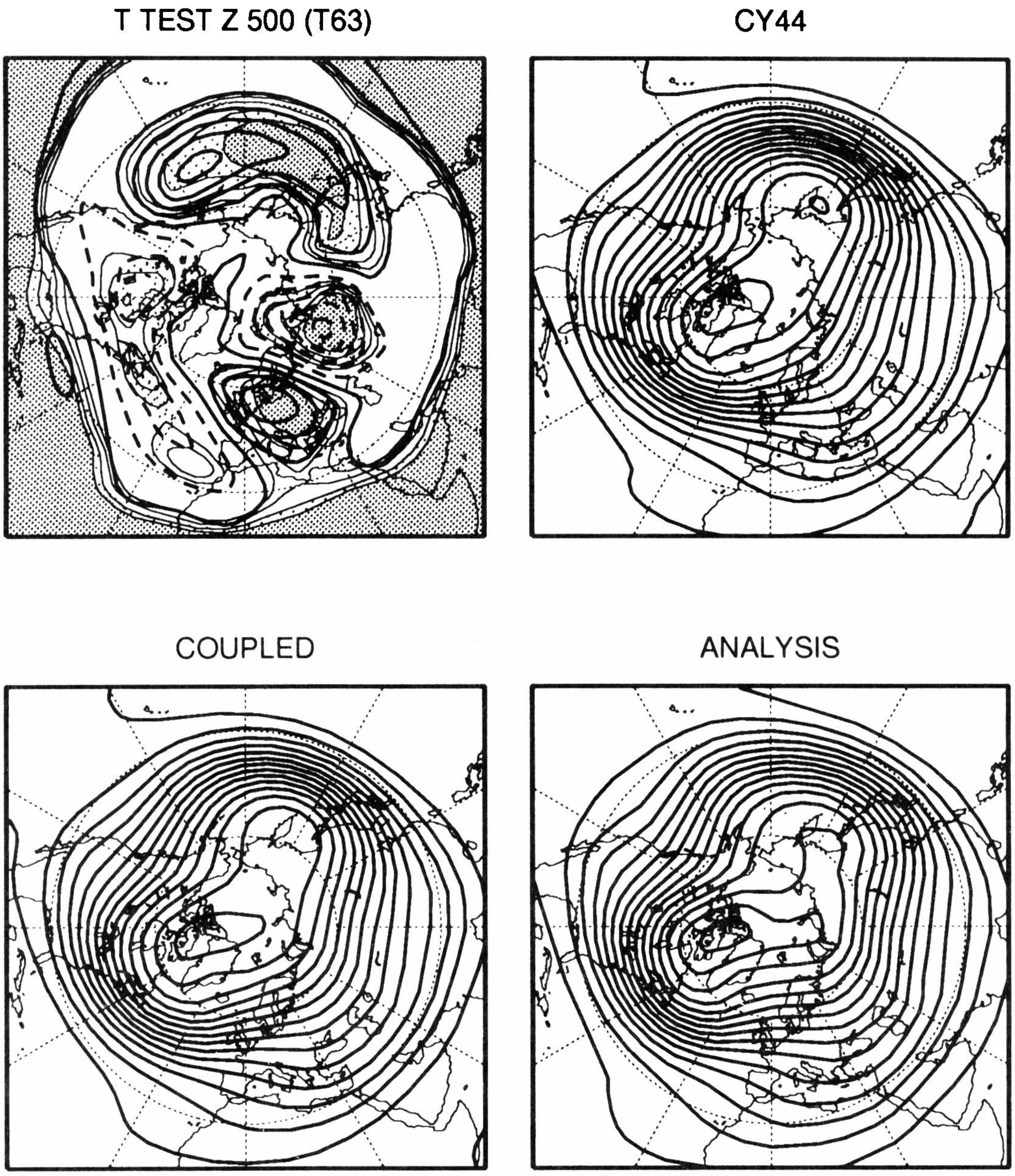


Fig. 6: Ensemble mean of 500 mb geopotential height of coupled and control run and their differences. For comparison also the mean analysed winter climate over the period 86-90 is shown. Period is winter 1990 and area is Northern Hemisphere.

quantity of interest, namely the surface wind vector, other factors, such as the age of the ocean waves, will determine the state of the high-frequency waves as well (Viers (1993)). Similarly, a nonlinear relation exists between the measured SAR image spectrum and the ocean wave spectrum. The wave spectrum can therefore only be obtained from the SAR image spectrum by means of an iteration process where the modelled wave spectrum is used as a first-guess (Hasselmann and Hasselmann (1991)).

Let us assume for the moment that (with help of modelled wave data if needed) we have observed wave data available. Can we use these wave data to obtain information about the weather over the oceans? The answer to this question is yes and thus far the most convincing support of this claim has been given by the work of de las Heras (1994), who has made a study of variational wave data assimilation in the relatively simply case of the generation of ocean waves by wind. As control variable the windspeed at 10m height was taken. The cost function was quadratic and penalised deviations between observed and analysed wind field under the constraints imposed by the WAM model (this, of course, required its adjoint). True and first-guess wave spectra were obtained by summing the single grid point version of the WAM model for one day with a windspeed of 18 and 12 m/s respectively. Wave heights of the 'true' model run were regarded as observations which were assimilated every three hours. No wind data were assimilated. Fig. 7 shows the resulting analysis and the comparison with the first-guess and reality for a number of parameters. The agreement is impressive for windspeed, friction velocity and the mean frequency. The approach was equally successful for turning wind sea cases (provided the mean wave direction was used in the assimilation) and for cases of simulated observed wave heights not generated by the WAM model itself. The work of de las Heras (1994) therefore shows that a variational approach seems promising. Clearly, wave data alone may provide information on the surface windspeed and stress. Combined with the recent results of Thépaut et al (1993), it follows that wave data may even give beneficial information on the atmospheric state throughout the troposphere.

Conclusion

An important element of the weather over the oceans is the sea state. Ocean wave prediction has enormous practical benefits for a number of human activities at sea. In addition, we have seen that wave information may provide important information on the weather of the oceans. Wave results are useful in the analysis of planetary boundary layer problems. Also, the associated momentum transfer to ocean waves may be important for the evolution of (extreme) weather systems, while it also plays a role in reducing systematic errors of the climate. Finally, in the context of 4DVar, observed wave data may even give beneficial information on the atmospheric state throughout the troposphere, while modelled wave data may be of help in interpreting satellite data from ERS-1, ERS-2, ...

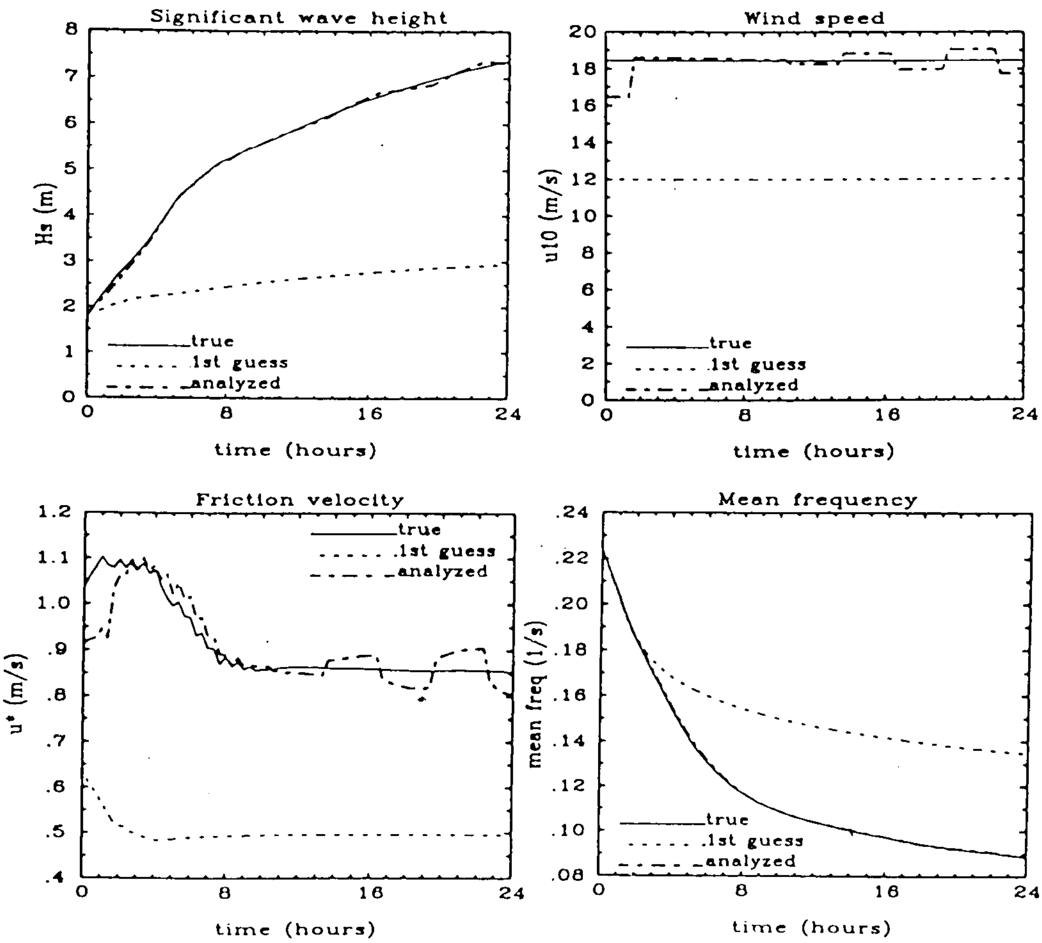


Fig. 7: **Comparison of time, first-guess and analysed wave height, windspeed, friction velocity and mean frequency using the adjoint method (from de las Heras, 1994).**

In view of all this it is to be hoped that in the near future one will be able to make optimal use of observed and modelled wave results for specifying the weather over the oceans by running a coupled atmosphere, ocean-wave model. Such a coupled system seems to be the optimal framework to bring about the above-mentioned advantages.

Acknowledgements

In the past eight years we have seen a rapid development of ocean wave modelling here at ECMWF. This work would not have been possible without the devoted contributions of Liane Zambresky and Heinz Günther. Besides carrying out their own work, they assisted the many visitors that came to use the model or to contribute to its development: Eva Bauer, Luciana Bertotti, Claus Brüning, Gerrit Burgers, Luigi Cavaleri, Vince Cardone, Juan Carlos Carretero, Spiros Christopoulos, John Ewing, Arthur Greenwood, Anne Guillaume, Björn Hansen, Miriam de las Heras, Jean-Michel Lefèvre, Piero Lionello, Henrique Oliveira-Pires, Renate Portz, Magnar Reistad, Claire Ross, Rachel Stratton, Bechara Toulany and Gerbrant van Vledder.

- Peter Janssen

References

Donelan, M.A., F.W. Dobson, S.D. Smith, R.J. Anderson, 1993: On the dependence of sea surface roughness on wave development. Y. Phys. Oceanogr. 23, 2143-2149.

Doyle, J.D., 1994: Air-sea interaction during marine cyclogenesis. Proc of the Life Cycles of Extratropical Cyclones, Bergen.

Hansen, B., P. Janssen and A. Guillaume, 1994: Validation and verification of the WAModel implemented on a global 1.5° (high resolution) grid, ECMWF Memo 940707.

Hasselmann, K., 1962: On the non-linear energy transfer in a gravity-wave spectrum, part 1: general theory. J. Fluid Mech 12, 481.

Hasselmann, K. et 15 al, 1973: Measurements of wind-wave growth and swell decay during the JOint North Sea WAve Project (JONSWAP). Dtsch Hydrors Z, Suppl A, 8 (12).

Hasselmann, K., and S. Hasselmann, 1991: On the nonlinear mapping of an ocean wave spectrum into a SAR image spectrum and its inversion. Y. Geophys. Res. Cq6 10713-10729.

de las Heras, M.M., 1994: On variational data assimilation in ocean wave models. PhD Thesis Utrecht University.

Komen, G.J., L. Cavaleri, M. Donelan, K. Hasselmann, S. Hasselmann, P.A.E.M. Janssen, 1994: Dynamics and modelling of ocean waves, Cambridge University Press.

Miles, J.W., 1957: On the generation of surface waves by shear flows. J. Fluid Mech. 3, 185-204.

Mitsuyasu, H., 1968: On the growth of the spectrum of wind-generated waves. 1. Rep Res Inst Appl Mech, Kyushu Univ, 16, 459-465.

Mitsuyasu, H., 1959: On the growth of the spectrum of wind-generated waves. 2. Rep Res Inst Appl Mech, Kyushu Univ, 17, 235-243.

Phillips, O.M. 1957: On the generation of waves by turbulent wind. J. Fluid Mech. 2, 417-445.

Smith, S.D., R.J. Anderson, W.A. Oost, C. Kraan, N. Maat, J. De Cosmo, K.B. Katsaros, K. Davidson, K. Bumke, L. Hasse, H.M. Chadwick, 1992: Sea surface wind stress and drag coefficients: The HEXOS RESULTS, Bound-Layer Meteor, 60, 109-142.

Snyder, R.L., F.W. Dobson, J.A. Elliott and R.B. Long, 1981: Array measurements of atmospheric pressure fluctuations above surface gravity waves. J. Fluid Mech. 102, 1-59.

Thépaut, J.-N., R.N. Hoffman and P. Courtier, 1993: Interaction of dynamics and observations in a four-dimensional variational assimilation. Mon Wea Rev, 121, 3393-3414.

VIERS-1 group: Snoeij, P., E. Van Halsema, J. Vogelzang, S. Waas, S. Zecchetto, H. Janssen, W. Oost, B. Jähne and Ch. Calkoen. VIERS-1 Final report Phase 3, BCRS report 92-94.

* * * * * * * * * * * * *

ADDITIONAL DATA SETS AVAILABLE FROM ECMWF

There have been important additions recently to the ECMWF archives. These are monthly mean fields for analysis and forecast products. The production of these archives is done at the beginning of each month, using the daily archives from the previous month. The monthly mean fields are in GRIB code.

The monthly means have been calculated for surface, model levels and pressure levels from the daily fields of four analysis cycles and one 10-day forecast. The resolution, levels and parameters are the same as for the daily archive data for the corresponding period. These are described in the MARS User Guide. Where changes to resolution, levels or parameters occurred during a month, the corresponding 'monthly' means are means of the days available; the number of days actually used is given in the GRIB header. Only the monthly means of analysis data on pressure levels and at the surface are being made available through ECMWF Data Services. Member States may also access this data plus forecast and model level data through MARS. To access the data from MARS it is necessary only to use 00 for the day in the required date and set STREAM=MONTHLY, and set other values as in requests for the Daily Archive. Analysis data are available from July 1985 and forecast data from August 1985.

- John Hennessy

* * * * * * * * * * * *

THE CRAY T3D COMPUTER AT ECMWF

Introduction

The Centre recently took delivery of a CRAY T3D computer which, at the time of writing, is in Final Acceptance. This article will describe the T3D system. It will not describe how to program the T3D; this will be discussed in a later edition of the Newsletter. It will be confined mainly to describing the hardware, explaining how jobs can use the T3D, and the method of communication between the CRAY T3D and its host CRAY YMP-2E system.

The hardware configuration

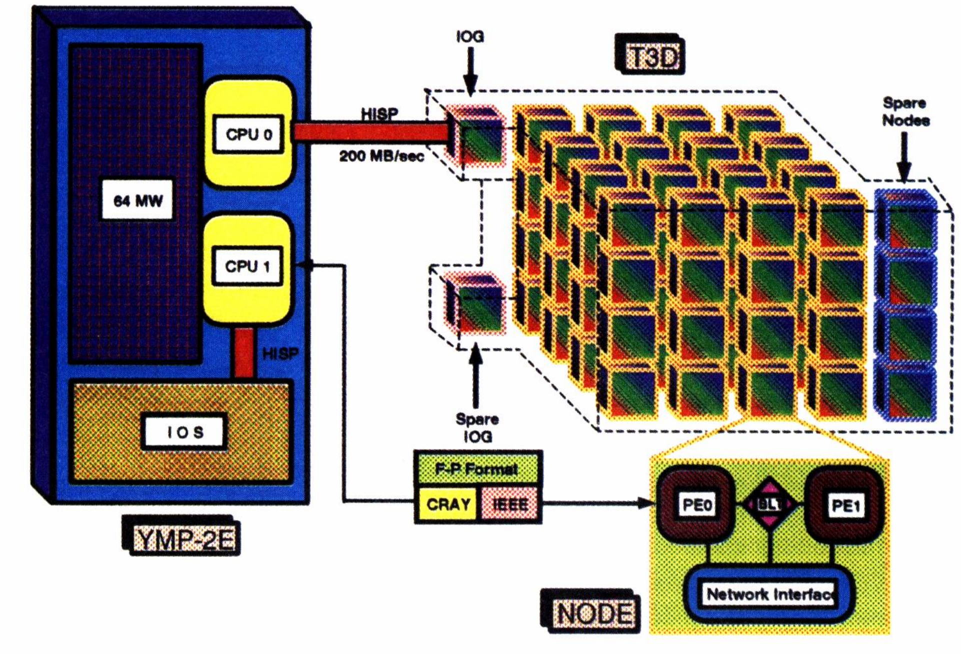
Fig. 1 is a schematic diagram of the hardware configuration of the T3D and its YMP-2E host.

The YMP-2E consists of: two CPUs; 64 MWords (512 MBytes) of memory and an IO subsystem; on which reside: an FEI3 network interface connected to the Operator WorkStation (OWS); an FDDI network interface connected to ECMWF's Local Area Network; two HIPPI network interfaces connected to a Network Systems PS32 HIPPI switch, and 16 DD-62 disk drives with a total capacity of 40 GBytes.

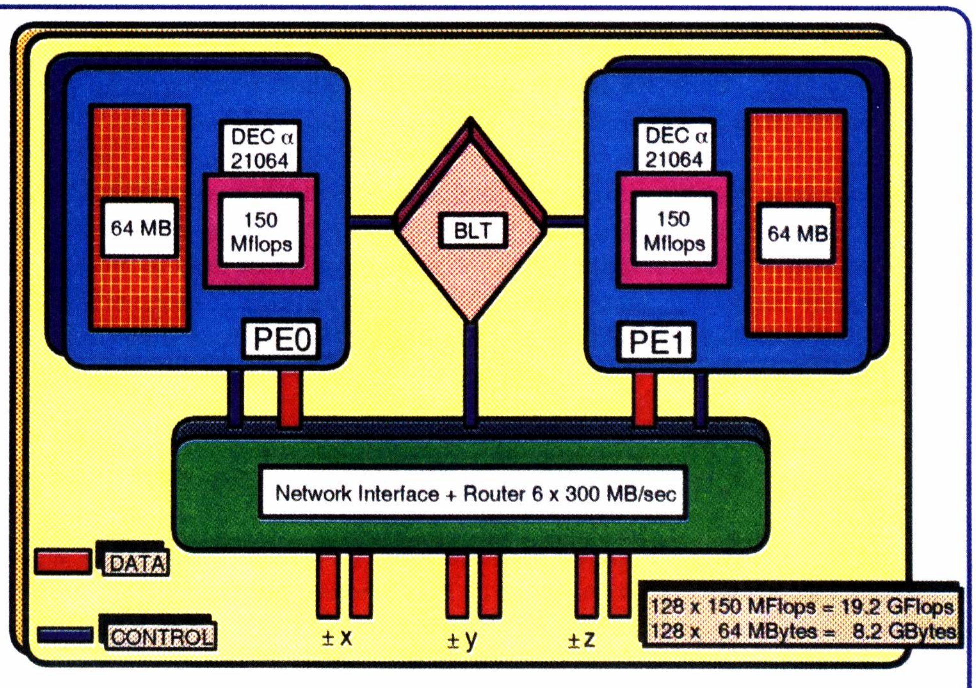
The T3D consists of: 64 nodes, each comprising two Processing Elements (PEs); four spare nodes; an IO Gateway (IOG) connected to the YMP-2E host via a 200 MBytes/sec high speed channel (HISP) and a spare IOG. Each PE is linked to its six nearest neighbours in the x-, y- and z-dimensions, as shown later in Fig. 3.

The IOG is a special type of node, similar to the compute nodes, but requiring different hardware to connect to the HISP. If a compute node were to malfunction it is a fairly simple exercise to replace it with one of the spare compute nodes. The operator needs only to mark the node bad in the T3D configuration file and a *reboot* of the T3D will automatically cause one of the spare PEs to logically replace the failed one. The process is not quite as simple if the IOG fails. In this instance it requires an engineer to replace the failed IOG physically with the spare before *rebooting* the T3D system.

The YMP-2E is classed in Cray terminology as a Parallel Vector Processor (PVP) as is the C90, while the T3D is classed as a Massively Parallel Processor (MPP). However, the term MPP is a misnomer. What really distinguishes the MPP from the PVP is not the number of processors, in fact Cray can offer T3D systems with as few as 32 PEs in them, but the memory architecture. The YMP-2E has a single memory, which is accessible from either CPU. The time taken to access the memory is very short and the same for each CPU. The T3D has a distributed memory architecture, where each PE has its own local memory, which it can access very quickly and which can be accessed from any other PE, but with a time penalty which is large in relation to the time taken for local memory accesses.



ECMWF Fig 1 T3D & YMP-2E Configuration Diagram



ECMWF Fig 2: Diagram of a Single Node

From Fig. 1 one can see that the T3D cannot function without its YMP-2E host. All the peripherals and network connections are attached to the IOS, which is part of the YMP-2E. Almost all the software including compilers, loaders, NQS and UNIX commands such as "grep", "awk", "ftp", "ecfile", "mars" etc. reside on the IOS disks and execute on the YMP-2E CPUs. The T3D relies totally on its YMP-2E host to perform I/O and provide even the most trivial system functions, such as obtaining the time of day (more about this later).

Starting the system involves the operators logging on to the OWS (the Operator Workstation - a SUN) and *booting* the IOS and the YMP-2E mainframe. Once the system is up and running, the operator then logs into the YMP-2E and *boots* the T3D. If the T3D crashes with a system software problem or hardware failure, this does not affect the running of its YMP-2E host, and the operator would then usually dump the T3D and *reboot* it. If, however, the YMP-2E host crashes, then the T3D will be affected and all jobs using the T3D will either be killed or rerun.

Fig. 2 is a schematic diagram of a single node. Each node consists of two processing elements (PEs), and some hardware shared by each PE, the network interface, router and block transfer engine (BLT). The processing element consists of a DEC α 21064 CPU chip, with a theoretical peak performance of 150 MFlops and 64 MBytes (8 MWords) of local memory. Since ECMWF's T3D contains 128 PEs, the MPP has a theoretical peak performance of 19.2 GFlops, with a total memory size of 8.2 GBytes. Unfortunately, because of the immaturity of the Fortran compiler and the relatively small data and instruction caches of the DEC α chip, the current actual peak performance of the T3D on the Centre's IFS code is more like 12 MFlops/PE or 1.5 GFlops in total. The DEC α chip uses IEEE format for floating-point numbers while the YMP-2E uses a Cray-proprietary format. There is software available to convert from one to the other, either explicitly using calls to library subroutines CRAY2CRI a CRI2CRAY, or implicitly by using the **'-N cray'** or (later this year) **'-N t3d'** parameter on an "assign" command.

It is implicit from Fig. 2 that a PE can access its own local memory very quickly, but to access any other PE's local memory (including that of the other PE in the same node), the PE has to send a request over the T3D's internal network which is an order of magnitude slower. The BLT allows large amounts of data to be sent across the network efficiently, without involving the DEC α CPU. However, it can block the CPU from being able to access local memory and thereby cause a reduction in the MFlop rate.

Fig. 3 is a schematic diagram of the node-network links. This figure shows a node in the centre surrounded by its six neighbours, left & right, up & down, back & front, corresponding to the negative and positive dimensions (x, y, z). There are six links to the centre node, in fact to every node in the T3D, and data can be flowing on all six links simultaneously at 300 MBytes/sec which amounts to a total bandwidth through the node of 1.8 GBytes/sec. One can see that the node in the most negative x-dimension is connected to that in the most positive x-dimension. Similarly nodes in the most negative y- and z-dimensions are connected to their opposite nodes. This

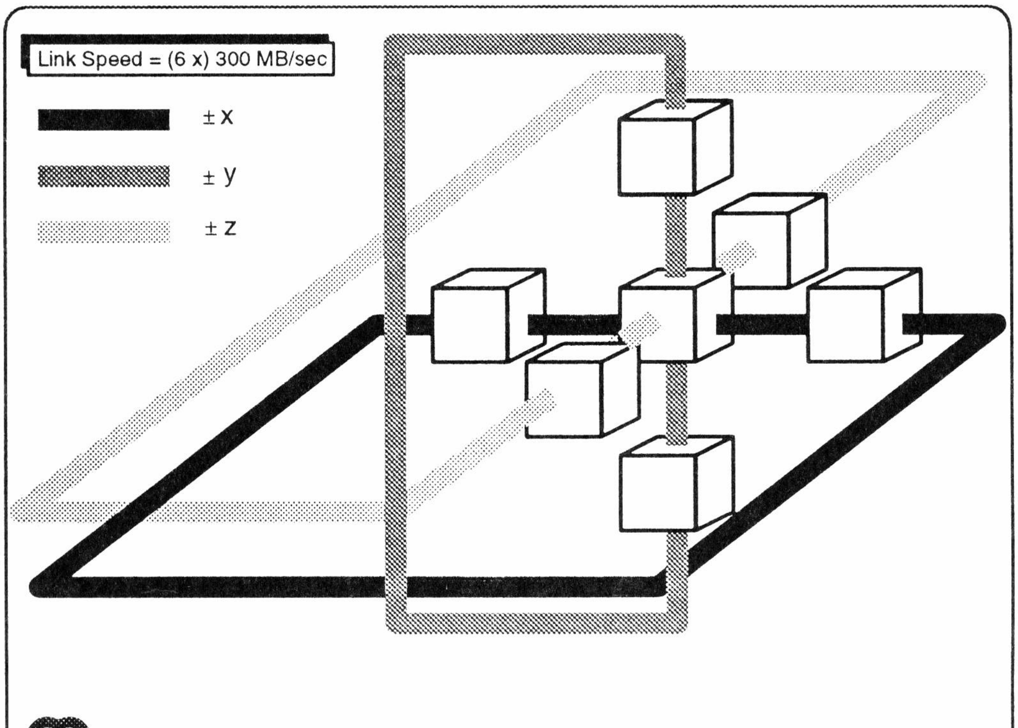


Fig 3: Node-Network Links

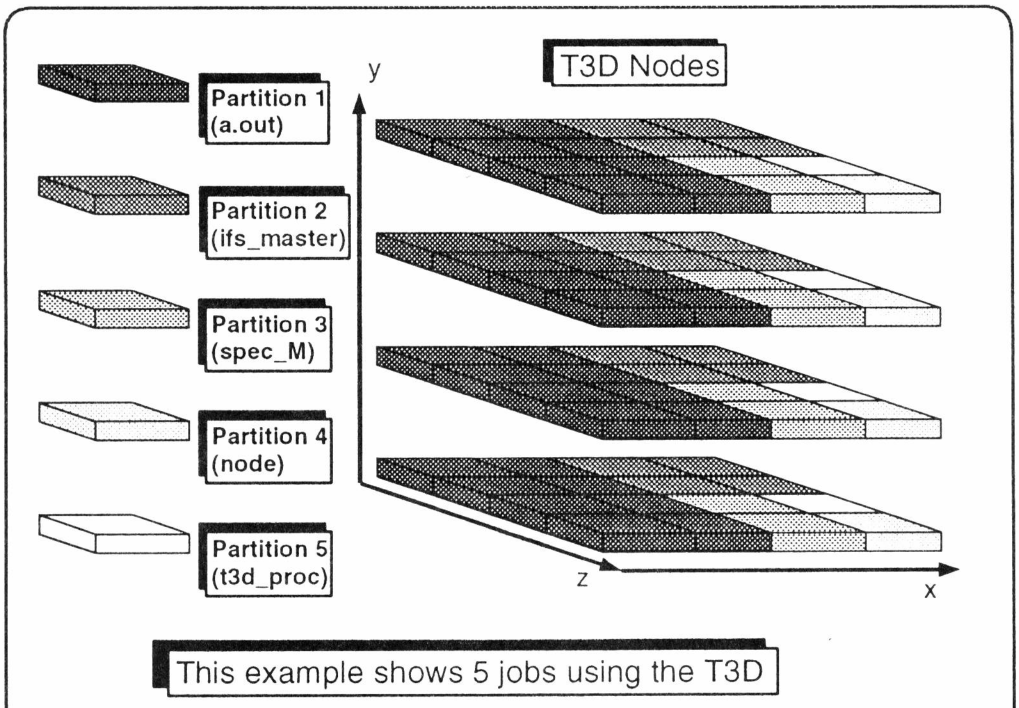


Fig 4: T3D Nodes Allocated to Partitions

topology is the 3-Dimensional Torus which gave the MPP its name of T3D. The IOG is actually *spliced* into the internal network, but is only linked to its four nearest neighbours as the y-dimension links are used to provide the HISP connection.

Allocation and use of T3D resources

Fig. 4 shows T3D resources (nodes) allocated to five different jobs. T3D nodes are allocated in powers of 2, i.e. 1, 2, 4, 8, 16, 32 or 64 nodes. Moreover, they are allocated as a group whose shape is a specific rectangular box. Hence if a job requests 64 PEs, it could happen that there are in fact 64 PEs (32 nodes) free, but the nodes cannot be allocated if they either do not make up a rectangular box, e.g. the 1st and 3rd plane in the x-dimension, or they are the wrong shape, e.g. (4, 2, 4) instead of (2, 4, 4).

The set of nodes allocated to a job are known as the job's T3D partition. The smallest partition is one node and the largest is 64 nodes. The case where a job requests a single PE is handled in a special way. The job will be allocated one node, but one of the PEs will be logically disabled by software. Partitions cannot overlap, i.e. once a node is allocated to a job its resources (PE CPU and memory) cannot be shared by another job. This has some serious consequences for long-running T3D jobs as we shall see later.

Fig. 5 shows a sample T3D NQS job. As was mentioned earlier, the T3D does not run NQS, so a job which needs to use the T3D must be submitted to the YMP-2E host system. This can be done from any of the other Cray computers at ECMWF, or from any of the workstations by submitting it to the "normal" queue on the YMP-2E host (sn1644), via a "**qsub -q sn1644.normal**" command. In Fig. 5 there are two sets of NQS *#QSUB* directives. The first three directives specify resources required on the YMP-2E host system, while the last three specify resources required on the T3D itself. These last three directives are the **only** NQS "T3D" directives available, so it is worth looking at them a little more closely.

The directive:

#QSUB -l mpp_p=64

informs NQS on the YMP-2E host that this job will, at some point during its execution, require a 64 PE partition (32 nodes) on the T3D.

#QSUB -l mpp_t=10

informs NQS on the YMP-2E host that this job will require a maximum of 10 seconds of **elapsed** time on the T3D.

```
#QSUB -eo                          } YMP-2E host system resources
#QSUB -lT 4                        }
#QSUB -lM 2Mw                      }
#QSUB -l   mpp_p=64                } T3D resources
#QSUB -l   mpp_t=10                }
#QSUB -l p_mpp_t=5                 }
cd $TMPDIR
cat >world.f <<'+EOF+'
     PROGRAM   world
     INTRINSIC my_pe
     PRINT *, 'Hello world from PE No.', my_pe()
     STOP
     END
+EOF+
mppcf77 -Wl'-X 64' world.f
a.out
ecfile -p data_file.27 get fort.11
a.out
```

ECMWF Fig 5: Sample NQS T3D Job

```
$ qstat -m
--------------------------------
NQS 80.38 BATCH QUEUE MPP LIMITS
--------------------------------
QUEUE NAME                RUN     QUEUE-PE'S   R-PE'S  R-TIME  P-TIME
                       LIM/CNT     LIM/CNT     LIMIT   LIMIT   LIMIT
---------------------- --- ---  ------ ------  ------  ------  ------
ops                       5/0        --/0        --      --      --
holding                   1/0        --/0        --      --      --
express                   6/0        --/0        --      --      --
normal1                   6/0        --/0        --      --      --
rdarchiv                  6/0        --/0        --      --      --
background                4/0        --/0        --      --      --
sys                       6/0        --/0        --      --      --
problem                   1/0        --/0        **      **      **
t3d_sys                   6/0        --/0       128      **      **
t3d_sa                    6/0        --/0       128    1000    1000
t3d_la                    6/0        --/0         4      **      **
t3d_lb                    4/0        --/0        16      **      **
t3d_lc                    2/0        --/0        64      **      **
t3d_ld                    1/0        --/0       128      **      **
---------------------- --- ---  ------ ------  ------  ------  ------
sn1644                   20/0       128/0
---------------------- --- ---  ------ ------  ------  ------  ------
```

ECMWF Fig 6: NQS Queue Structure

#QSUB -l p_mpp_t=5

informs NQS on the YMP-2E host that any single T3D partition created by this job will be required for a maximum of 5 seconds elapsed time.

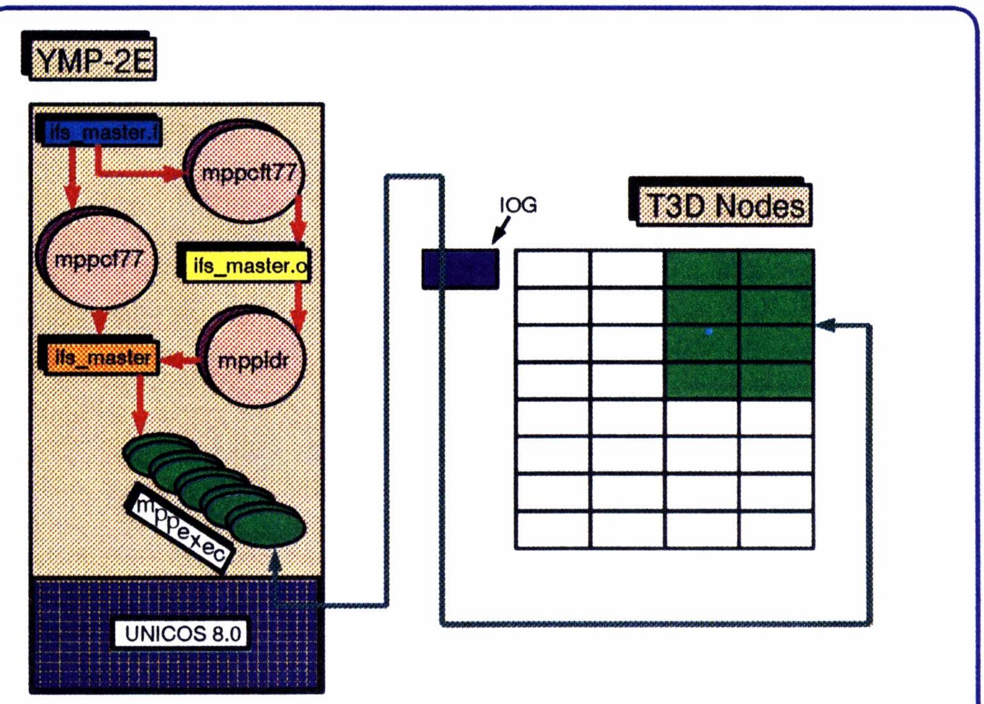
The *#QSUB* directives themselves are interpreted by the **qsub** command on the machine on which it is executed, and passed to NQS on the YMP-2E host. Once the job arrives on the YMP-2E it goes into one of the queues set up to handle jobs which require T3D resources. Eventually it will start to execute **on the YMP-2E**. The first command "cd $TMPDIR" executes on the YMP-2E, as does the next one: "cat >world.f <<'+EOF+'". The third command "mppcf77 -Wl'-X 64' world.f" also executes on the YMP-2E. As you may guess from its name, *"mppcf77"*, this is similar to the *"cf77"* command, but instead of producing an *"a.out"* file which can execute on the YMP-2E as is the case with *"cf77"*, *"mppcf77"* produces an *"a.out"* file which will execute on the T3D. When it is time for the next command *"a.out"* to execute, the UNICOS kernel recognises this as a T3D binary and takes steps, which will be described later, to ensure that it executes on the T3D, not on the YMP-2E host. The *"ecfile"* will, of course, be executed on the YMP-2E host.

The current NQS queue structure on the YMP-2E host system is shown in Fig. 6. This may change in the light of experience with T3D jobs. The queues into which jobs requiring T3D resources are put all have names beginning "t3d_". Any "T3D" jobs requiring less than 1001 seconds of T3D time will be put in the "short" queue "**t3d_sa**", while any job requiring more than 1000 seconds of T3D time will be put in one of the "long" queues, "**t3d_la**", "**t3d_lb**", "**t3d_lc**" or "**t3d_ld**", depending upon the PE limit specified, **4, 16, 64** or **128**. The queues are designed to allow the T3D resources to be managed as much as possible using NQS, by switching queues on and off and by suspending or resuming jobs in these queues.

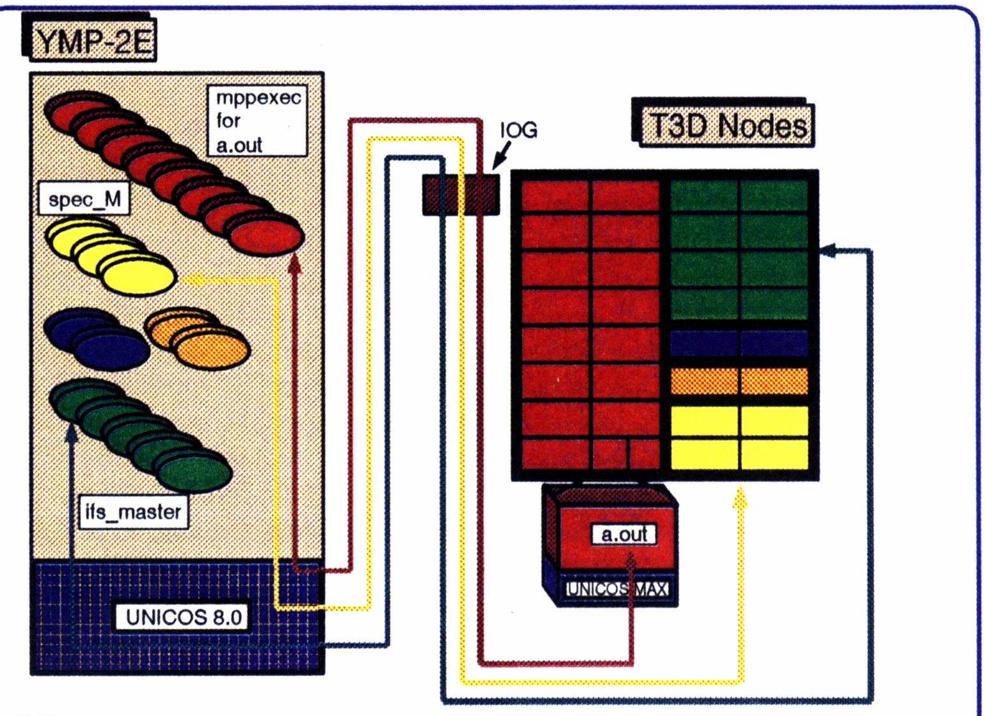
Downloading applications and T3D/YMP-2E software linkages

Fig. 7 shows the steps taken to load a T3D application starting from the Fortran source code. The Fortran source in file "ifs_master.f" is passed through the "mppcf77" command to produce a T3D binary program. Alternative, it can be compiled using the "mppcft77" command and the resulting "ifsmaster.o" can be passed through the "mppldr" command to produce the T3D binary "ifs_master". The first record in the "ifs_master" binary contains the string "#!/mpp/bin/mppexec". When the job tries to execute "ifs_master", the kernel sees this first record and actually loads a program called "mppexec". This program, which is multi-tasked, reads the actual T3D binary portion of "ifs_master" and contact the MPP configuration driver to allocate T3D resources. Once the resources have been allocated it uses the YPE-driver to load the T3D binary onto the allocated PEs and starts them executing.

When the program is executing on the T3D, mppexec acts as a "go-between" between the T3D application and the UNICOS operating system, and is known as the application's agent. This linkage is shown in Fig. 8.



ECMWF Fig 7: Steps Taken to Load a T3D Application



ECMWF Fig 8: T3D - YMP-2E Software Linkages

The agent is multi-tasked and the number of tasks depends upon the number of PEs in the partition according to the formula "nPEs/4 + 3". This in effect means that each task handles requests for four PEs. These 'requests' are actually system calls that the T3D application makes. Looking at the expanded view of the memory of one of the PEs one can see that the application "a.out" uses most of the memory, while UNICOS MAX, a micro-kernel (μ-kernel) operating system, occupies about three or four MBytes. This μ-kernel has very little functionality. It handles PE memory functions, signal interrupts, and very little else. It cannot even handle simple requests from application such as a request for the time of day.

When a T3D application needs an operating system facility, such as I/O or time of day, it has to send the request to the μ-kernel's UNIX System Call Handler (USCH). USCH passes the request, via the IOG and UNICOS, to the application's agent (mppexec) on the YMP-2E host. the agent then passes the request to UNICOS, where the request is satisfied and the response returned to the agent. The agent then passes the response, via UNICOS and the IOG, back to USCH on the originating PE which, in turn, passes it to the application. Since many PEs can be making system requests simultaneously, the agent is multi-tasked so that "mppexec" does not become a *bottleneck* in the chain of events.

Scheduling considerations

As was mentioned earlier, T3D PEs are allocated to partitions in powers of two, in the shape of rectangular boxes. Because of this it may appear that sufficient PEs are available to satisfy a request for resources, but because of their position in the T3D they may not be capable of fulfilling the request.

Nodes cannot be shared between applications, i.e. partitions cannot overlap. Once a T3D application has been allocated a partition it owns all the resources (PEs, BLTs, network interfaces, routers etc.) and cannot share these with any other applications. Once the application terminates, the resources are freed and the configuration manager may then allocate them to another application.

A T3D application cannot be suspended or checkpointed; however, the application's agent (mppexec) can be suspended, with potentially disastrous results since the T3D application will be unable to have any system calls serviced. Once operational work is being run on the T3D it may be necessary to kill or rerun long-running T3D applications in order to enable the operational application to obtain the resources it needs. In a future version of UNICOS MAX, to be released later this year or early next year, Cray plan to provide a "job rolling" facility whereby the application will be "quiesced" then rolled to disk, thus allowing its T3D resources to be re-used in another partition. The job can then be reloaded later on, to carry on from where it left off.

MPP time limits refer to **elapsed** (wall-clock) time, not CPU time. It is therefore not a good idea to use facilities such as "ecfile" or "mars" from within a T3D application since they can take a variable amount of time to complete, depending upon the workload, the state of the software on the data handler and whether or not the data is on magnetic tape. If you need to use these facilities, do so in the job prior to running the T3D application. That way you do not tie up valuable resources that cannot be used.

- Neil Storer

* * * * * * * * * * * *

SECURITY CONSIDERATIONS OF MEMBER STATE ACCESS TO THE CENTRE'S UNIX SERVICES

The following article is an extract from document ECMWF/TAC(94)15 presented by the Centre to the Technical Advisory Committee in September 1994.

Introduction

In the late 1970s and during most of the 1980s, the Centre's Computer Facility was connected to a small number of individual Member State machines directly controlled by the national weather services. The early telecommunication protocols supported little more than file transport and batch job submission. The introduction of the DECnet based system added many new facilities. Most notably, it enabled interactive access from Member State machines. At the request of Member States preferring not to use DECnet, the Council, at its 35th session (December 1991), allowed the use of TCP/IP protocols on the ECMWF wide-area network.

The increased use of UNIX and TCP/IP, both at ECMWF and within many Member States, has led to a complex network of interconnected machines. Increasingly, the Member State end of the telecommunication link to the Centre is now directly accessible to a multitude of machines. Some of these machines are even outside the administrative control of the national weather services, e.g. machines located at affiliated organisations such as climate research centres.

The Centre's computer security is directly affected by these developments: there are now many more hosts from which unauthorised access may originate, the complexity of the services has dramatically increased and the prevalent host operating system, UNIX, was not designed with security in mind. The Centre intends to gradually complement the existing VMS based services available to Member States with equivalent UNIX based services. A Member State Security Representatives' Workshop was held in June 1993 to deal with the security aspects of a UNIX based remote batch service. During that workshop, a system combining the Network Queuing System (NQS) with smart cards was identified as the most desirable approach. An experimental system was constructed and the principle of remote batch job authentication using smart cards was successfully demonstrated. The results of this trial were then reviewed by the Member State Security Representatives during their meeting in May 1994.

Risks and major weaknesses

On most UNIX systems, once an intruder gains 'root' access, i.e. the intruder obtains system administrator privileges, the system as a whole is compromised. The system is then 'root-compromised' and the intruder will have access to all files, storage devices, network connections, etc. Within ECMWF, the various hosts trust each other to varying degrees. Having gained root access on one of the Centre's major systems, a malicious intruder could cause several days of downtime.

More likely, an intruder will be interested either in publicity or in using the identity of ECMWF's hosts for attacks on sites connected to ECMWF. Member State users who store userid/password combinations on ECMWF's machines (e.g. for outgoing file transfers) make it particularly easy for intruders to compromise further systems.

A typical UNIX attack proceeds along the following lines:

i) gain initial access;

ii) use a known UNIX vulnerability to obtain root access;

iii) use the access to the root functions to disguise the presence of an intruder;

iv) use the access to the root functions to compromise further systems (e.g. to search for stored passwords, to monitor local area network traffic for cleartext passwords, etc.).

The information gained in step iv) is then used to repeat the process on other systems.

In ECMWF's case, the likelihood of each of these individual steps to succeed is as follows:

Step i) The initial access to the Centre's machines is protected by simple passwords. These passwords are transmitted in cleartext over local and wide-area networks and the nature of the batch access encourages users to store passwords as part of their jobs.

If a Member State installation is root-compromised, step i) can be trivial.

Step ii) Over the past 12 months ECMWF received about 20 serious security advisories. Many of the weaknesses exposed by these advisories were present in ECMWF's systems. The Centre makes every effort to remove these vulnerabilities as quickly as possible.

No matter how quickly the Centre reacts, UNIX operating system exposures will continue to be present. Some of the recently announced vulnerabilities were known to some groups of hackers for many months. This situation is unlikely to change over the next two years.

Step iii) The Centre has implemented some measures to make this step more difficult. However, at this stage the intruder already has root access and any measures are therefore limited to the detection rather than the prevention of incidents.

Step iv) The likelihood of this step to succeed depends largely on the authentication technique employed at the sites routinely accessed from ECMWF. Sites using reusable cleartext passwords will be very vulnerable. Sites using a strong authentication method will be much more difficult to penetrate.

Several authentication techniques have been developed that address steps i) and iv). Employing one of these techniques will dramatically reduce the chances of intruders gaining initial access to ECMWF. As to step ii), the Centre already devotes considerable resources to maintaining the integrity of its host systems. There is little more to be gained by increasing the efforts in this area.

By far the most effective way of making the Centre's security measures adequate for today's networked environment is to replace the standard, reusable passwords by a stronger authentication method.

Authentication techniques

There are a number of techniques which address the problems of the standard password scheme:

Trusted Hosts

This form of authentication is only suitable within one administrative domain, e.g. within ECMWF or within a national weather service. Typically, it is employed between machines within one physical location. Remote users accessing a system from a trusted host are automatically validated - passwords are not required. If a remote host is root-compromised, all systems trusting that host are compromised as well. A network of trusted hosts, spanning all Member States, administered to a variety of standards, would not be in the interest of computer security.

Kerberos

Kerberos, an authentication and key distribution system invented at the Massachusetts Institute of Technology, allows safe authentication across public networks. The vulnerabilities resulting from the transmission over networks have been resolved but the other problems remain. There are still reusable passwords and cleartext passwords can be stored in files. Kerberos uses a ticket granting system to avoid users having to enter their password each time a new service is required. On root-compromised systems, an intruder would have access to the ticket granting mechanism.

Software based challenge/response systems

Challenge/response systems provide passwords that are only used once (so called one-time passwords). The public domain S/KEY package is probably the best known of these systems. Like Kerberos, these systems mainly address the risks resulting from transmitting authentication information over networks but they provide little protection if hosts are root-compromised.

Physical tokens

These systems provide each user with a unique physical token and a private PIN number to unlock that token. These tokens are commonly called smart cards. There are two major technologies: time synchronised tokens and challenge/response tokens. Both systems can be safely used over public networks and provide a large degree of protection in the event of root-compromised hosts: an intruder operating over networks will not be able to obtain the authentication information stored in the physical tokens. Smart cards also prevent the exposures resulting from users choosing easily guessable passwords.

Of these technologies, only smart cards will provide the required level of protection. Smart cards based on a challenge/response interaction require network access to the target host during authentication and are therefore not suited for a batch queuing system. Time synchronised tokens are in principle suitable for batch access although none of the commercially available systems include batch support.

Secure batch trial

A trial was conducted to evaluate whether it would be possible to adapt a smart card system to authenticate users in a batch job environment. The scope of this trial was to build a TCP/IP based secure batch submission service, using existing standard components where possible. It was decided to use a public domain version of the Network Queuing System (NQS) software as the basis for the batch service and to enhance it to include a strong authentication mechanism.

A number of possible mechanisms were investigated and the SecurID card from Security Dynamics Inc. was selected. SecurID cards are credit card sized, time synchronised tokens which display a new, unpredictable authentication code every minute. The original version of the Security Dynamics software did not lend itself easily for batch authentication, but after discussions with the company, these problems were overcome by the provision of a new feature in the SecurID product.

A detailed technical description of the trial was given at the Member State Security Representatives' Meeting held at ECMWF in May 1994, along with reports from the Member States involved with the trial (Denmark and The Netherlands). The trial proved that smart cards can be used successfully to authenticate batch jobs.

A questionnaire had been distributed before the Security Representatives' Meeting requesting information about the use of NQS within Member States and on the type of batch jobs submitted to ECMWF. The results of this questionnaire showed that about half of the Member State installations are already using NQS and that some Member State users submit batch work to ECMWF from jobs running on Member State machines.

The term 'unattended remote batch' was established to describe batch work submitted by remote programs rather than remote users. There is a fundamental conflict between this type of batch access and a strong authentication based on physical tokens: to enable automated batch submission without human intervention, the authentication information has to be stored on the remote host. A ticket based mechanism can be devised to achieve that, but any such extension will considerably weaken the new authentication system by re-introducing the problems inherent in root-compromised hosts. If provided, the unattended remote batch features should be restricted to as few users as possible.

Proposed strategy for secure access to the Centre's UNIX services

It is proposed that all general purpose access to the Centre's UNIX services will be protected by a strong authentication mechanism based on time synchronised smart cards. A minimum number of machines, probably only two, will be modified to support the selected smart card system. These machines will then validate all external general purpose access to ECMWF. Limiting the smart card system to a small number of gateway hosts avoids costly software licences and ensures that the support of the selected smart card system does not become a requirement for the Centre's major systems.

All other ECMWF hosts with Member State accounts will be accessible from these gateway machines. Authentication from the gateway machines to other ECMWF hosts will be based on the trusted host principle or on passwords, depending on the type of service. Smart cards will only be required for the initial access to the Centre's computing facility.

This system should be complemented by application specific access to ECMWF. For example, an interface allowing solely MARS retrievals poses significantly fewer security risks than general purpose access. Such interfaces could be allowed to bypass the smart card authentication, making them much more suitable for automated processing within Member States.

In detail, the individual UNIX services will be protected in the following way:

telnet — Member State users will first telnet to 'ecserver', the general UNIX server for Member State use. Access to 'ecserver' will be authenticated via smart cards. From 'ecserver', users can then submit batch jobs (without further validation) or telnet to other ECMWF machines (protected by passwords).

batch — ECMWF will distribute 'ecqsub', a client program which will be ported to all major UNIX platforms. The 'ecqsub' utility is designed to run locally within Member State installations and will submit batch jobs to ECMWF. At the time of submission, 'ecqsub' will prompt for authentication information from the user's smart card. Commands for job control and status information will be part of the package.

A ticket based system to facilitate unattended remote batch is required by some Member States. Member State Computing Representatives will be able to control the use of this facility on a userid/hostname basis.

file transfer — Incoming ftp requests require strong authentication. The gateway machines will offer a version of ftp prompting for smart card information rather than passwords. Data destined for other ECMWF hosts will have to be transferred to a gateway machine and then copied or accessed via NFS.

A utility allowing automated file transfers between ECMWF and selected Member State machines is highly desirable. The security aspects of such a utility were discussed during the last Member State Security Representatives' Meeting. Transferring data from ECMWF to Member State machines was considered low risk but this was not the case for traffic in the opposite direction.

Timescales

If approved, a smart card system will be acquired and installed towards the end of 1994. The most important service to be protected is the interactive access to the Centre's UNIX machines. Two months after the telnet gateway is configured and all Member State users registered for that service have received their smart cards, the password based interactive access to 'ecserver' will be discontinued. The new ftp and batch service will be introduced in a similar manner.

With the exception of interactive access, there are no plans to extend the proposed new authentication system to the VMS based services. Interactive access to VMS may be included at a later stage. Interim solutions, such as the ftp based job submission system, will be phased out once the proper replacement services are in place.

The Technical Advisory Committee supported the Centre's proposal that a smart card authentication system should be introduced and will inform the Council accordingly, at the next Council session (December 1994).

- Walter Zwieflhofer

* * * * * * * * * * * *

StorageTek UPGRADE TO 36 TRACK

Developments in this field began in 1991 when IBM announced that they would introduce a new magnetic tape cartridge system, similar to their existing 3480 system, with an increased data capacity achieved by doubling the number of recording tracks from 18 to 36. In 1992 they announced that the new drives could use a cartridge holding 1100' of media compared with the standard 575' length. Previously, in 1989, they had announced support for a hardware data compaction system which could increase capacity by 50% - 400%. Thus IBM led the industry into a new standard for data recording, increasing the original 210 MB capacity to something greater than 800 MB. Naturally other suppliers followed IBM's lead.

The Centre had installed StorageTek tape robots and tape drives beginning in 1990 and StorageTek announced that they too would support compaction, 36 track and 1100' tape media. Unlike IBM, STK also offered a technical upgrade path which allowed existing investment to be preserved. The Centre installed the compaction option in 1991 and the first 36 track upgrades became available in late summer 1993. Thus in September 1993, two drives and one controller were upgraded, followed closely by a further four drives and a second controller. This gave the opportunity to try out the new equipment to ensure that it met our reliability criteria for the computer service.

Initially, the new equipment worked without problems, but as the date to upgrade the remaining equipment approached it was clear that we were experiencing various unexplained problems. The equipment was still very new to STK so they were also in a learning phase. Given the need to gain time to analyze the problems, it was agreed that the major upgrade would be delayed from mid November 1993 until early January 1994. During this delay, STK delivered new microcode, which resolved many of the problems, and ECMWF introduced a systematic test of the new equipment, under carefully controlled conditions, so that we could establish the root cause of the problems.

Most of the remaining equipment was then upgraded early in January 1994, at which time all new data was written in 36 track mode but existing data, and essential backups, remained in 18 track mode. Shortly after this, a further problem occurred when it became clear that the new drives could not read "marginal" 18 track media, though these could be read without problems on the old drives. This was solved by a microcode change from STK. This method of operation continued until late February when 18 track data began to be copied to 36 track media, thus reducing the need for data cartridges and allowing a doubling of data stored in the STK silos.

From early March 1994, copying began in earnest. By the end of September 54000 cartridges had been copied with a further 12000 still to be done. It is expected that all data will be copied to 36 track media by late October 1994. This has released over 30,000 tape cartridges which means that no further cartridges will have to be purchased for several months.

The move to 36 track media was certainly worthwhile since it halves the cost of storage. However, 36 track makes greater demands on the media and so the effort required to support a reliable data library has increased. Each day, cartridges flagged as exceeding certain recoverable error thresholds are checked, and if found to have problems, data is copied from the cartridge to a new cartridge. In the days of 18 track cartridges around 1 - 2 cartridges per month were replaced. This number has now increased to 10 - 15 per month.

- Peter Gray

* * * * * * * * * * * *

THE USE OF 8mm DATA CARTRIDGES AT ECMWF

During the last few months we have seen an increase in the number of requests to transfer data to 8mm Helical Scan tapes, alternatively known as Exatapes or Exabytes. Their huge theoretical capacity, 2.3 to 5 Gigabytes, compared to 100 to 300 Megabytes for traditional tapes, their small physical dimensions, and their ease of use on UNIX workstations are the main reasons why they are becoming so popular.

The Tape Library is now able to provide a service for transferring user's data stored in CFS (ECMWF's archiving system) to these tapes, similar to the existing service for IBM 3480 cartridge tapes. This service only covers the creation (writing) of tapes. Users wanting to import data on Exabyte tapes should contact the User Support Section.

This newly supported medium can also be used to transfer data coming from the meteorological archive via ECMWF's Data Service. Contact Data Services for more information.

This article describes the procedure to follow when you request transfer of data to 8mm tapes. The ECMWF Computer Bulletin "*Supporting incoming/outgoing magnetic tapes at ECMWF*" (B8.2/1) will be updated later.

1. Preliminary remarks

The tapes are written in tar format. They will therefore be readable mainly on UNIX systems, using the tar command.

As single files can reach some hundreds of Megabytes, each archive file on the tape will contain only one data file. You will therefore have to use the "norewind" device name to extract the file and run subsequent "tar" commands, each command reading one file.

If you transfer data coming from the MARS archive (not via Data Services), you are reminded that this data when extracted on the CRAY C90 computer is created as COS Binary Blocked files. These files should therefore be unblocked before they are stored in ecfile (CFS) if you plan to use them on non-Cray systems. To do this use the command uscpblock on the Cray, e.g.

```
% uscpblock -u -fBB file1 file2
```

will "unblock" file1 into file2. Please note that this can not be done by the Tape Librarian when transferring the file to tape.

The transfer of data to 8mm tapes is slower than to ½" tapes. Our tape library may only be able to produce two tapes per working day.

You are asked to provide sufficient tapes to our tape library beforehand to cover the volume of data you wish to have transferred.

A similar service can be provided for QIC ½" cartridges, but you should remember their small capacity - 150 Mbytes. Currently we do not provide a service for DAT data cartridges.

2. Administration

To request the transfer of files from CFS to 8mm tapes, you should write a letter addressed to "The Tape Librarian" or send a mail message to "tape-library@ecmwf.co.uk".

You are asked to provide the following information:

2.1 The ecfile directory(ies)

The ecfile (CFS) path name where the files to be transferred are held, should be provided. In addition to this, you will have to give read permission to our Tape Librarian to the files to be transferred, or to the directories containing those files.

In the following example, suppose that your data is held in the CFS subdirectory /uid/storm, in files day1, ... day 10.

Read access for the Tape Librarian can be given with the ecfile command, e.g.

a) by giving read access for each single file:

```
% ecfile -p storm/day1 -v a/csp/r/-/s,a/csg/r/-/s modify
% ecfile -p storm/day2 -v a/csp/r/-/s,a/csg/r/-/s modify
....
% ecfile -p storm/day10 -v a/csp/r/-/s,a/csg/r/-/s modify
```

These commands give read access to our Tape Librarian (users csp and csg) for each of these ten files.

It can be done more easily by

b) giving read access to the directory

% ecfile -p storm -v a/csp/r/-/s,a/csg/r/-/s modify

2.2 A list of the files

This list should contain the file names, including the size of each file. It can easily be extracted from the output of the ecfaudit command, e.g.

% ecfaudit storm

will list all the files (filenames, size...) in the subdirectory "storm".

2.3 The tapes to use

Please give a list of the tapes to be used. As noted above, you are asked to provide the Tape Librarian with sufficient tapes in advance.

2.4 The density to use

Data can be written on 8 mm tapes in one of two formats:

* low density with a maximum capacity of 2.3 Gigabytes
* high density allowing the storage of 5 Gigabytes per tape.

As not all drives support both densities, you should check in advance which format you will be able to read and select the appropriate one.

In addition to choosing the density, some drives are able to run a compression tool on your transferred data, which theoretically provides a maximum capacity of 10 Gbytes when you select the high density format. If you have this type of drive available you can also select compression.

However, remember that the amount of compression depends on the characteristics of the data. In various tests and transfers that we ran the capacity reached was close to the theoretical maximum capacity when compression was **not** used, eg. more than 2 GBytes for low density and around 4.5 GBytes for high density. However, we could "only" transfer some 6 GBytes to one tape when selecting high density and compression.

will "unblock" file1 into file2. Please note that this can not be done by the Tape Librarian when transferring the file to tape.

The transfer of data to 8mm tapes is slower than to ½" tapes. Our tape library may only be able to produce two tapes per working day.

You are asked to provide sufficient tapes to our tape library beforehand to cover the volume of data you wish to have transferred.

A similar service can be provided for QIC ½" cartridges, but you should remember their small capacity - 150 Mbytes. Currently we do not provide a service for DAT data cartridges.

2. Administration

To request the transfer of files from CFS to 8mm tapes, you should write a letter addressed to "The Tape Librarian" or send a mail message to "tape-library@ecmwf.co.uk".

You are asked to provide the following information:

2.1 The ecfile directory(ies)

The ecfile (CFS) path name where the files to be transferred are held, should be provided. In addition to this, you will have to give read permission to our Tape Librarian to the files to be transferred, or to the directories containing those files.

In the following example, suppose that your data is held in the CFS subdirectory /uid/storm, in files day1, ... day 10.

Read access for the Tape Librarian can be given with the ecfile command, e.g.

a) by giving read access for each single file:

```
% ecfile -p storm/day1 -v a/csp/r/-/s,a/csg/r/-/s modify
% ecfile -p storm/day2 -v a/csp/r/-/s,a/csg/r/-/s modify
....
% ecfile -p storm/day10 -v a/csp/r/-/s,a/csg/r/-/s modify
```

These commands give read access to our Tape Librarian (users csp and csg) for each of these ten files.

It can be done more easily by

b) giving read access to the directory

% ecfile -p storm -v a/csp/r/-/s,a/csg/r/-/s modify

2.2 A list of the files

This list should contain the file names, including the size of each file. It can easily be extracted from the output of the ecfaudit command, e.g.

% ecfaudit storm

will list all the files (filenames, size...) in the subdirectory "storm".

2.3 The tapes to use

Please give a list of the tapes to be used. As noted above, you are asked to provide the Tape Librarian with sufficient tapes in advance.

2.4 The density to use

Data can be written on 8 mm tapes in one of two formats:

* low density with a maximum capacity of 2.3 Gigabytes
* high density allowing the storage of 5 Gigabytes per tape.

As not all drives support both densities, you should check in advance which format you will be able to read and select the appropriate one.

In addition to choosing the density, some drives are able to run a compression tool on your transferred data, which theoretically provides a maximum capacity of 10 Gbytes when you select the high density format. If you have this type of drive available you can also select compression.

However, remember that the amount of compression depends on the characteristics of the data. In various tests and transfers that we ran the capacity reached was close to the theoretical maximum capacity when compression was **not** used, eg. more than 2 GBytes for low density and around 4.5 GBytes for high density. However, we could "only" transfer some 6 GBytes to one tape when selecting high density and compression.

2.5 Your address

... of course. The tapes will be sent by normal mail.

3. Additional information

Please consult the ecfile and ecfaudit man pages on ECMWF systems, and the tar man page on any UNIX system for further information.

Please contact User Support if you have any additional questions.

- Dominique Lucas

* * * * * * * * * * * *

COMPUTER USER TRAINING COURSE

The Centre is offering a computer user training course for Member States' personnel and ECMWF staff from 6 - 24 March 1995. Full information and a request for nominations has been sent to all Member State meteorological services.

The course is divided into three one-week modules. Attendees may register for separate modules.

Week 1: UNICOS extensions to UNIX, ECMWF utilities, the ECFILE file system, MARS.

Week 2: MAGICS, Advanced UNICOS.

Week 3: MPP usage.

Each week will consist partly of lectures and partly of practicals. In more detail, the three modules are:

MODULE 1 (6 - 10 March 1995): ECMWF'S UNICOS SERVICE

System and hardware overview

UNICOS batch jobs

FORTRAN

ECFILE file storage system

MARS data retrieval

File transmission services, including **sendtm**

Those attending module 1 are expected to know basic UNIX commands and be able to use the *vi* editor.

MODULE 2 (13 - 17 March 1995): MAGICS & Advanced UNICOS

MAGICS

Introduction and overview

Concepts

Parameters

- Subroutines
- Action and pseudo-action routines
- Data input

Plotting features.

ADVANCED UNICOS

Program maintenance tools

Introduction to vectorising

Specialist I/O techniques.

Those attending module 2 are expected to know basic UNIX commands, be able to use the *vi* editor and to be able to submit jobs to UNICOS.

MODULE 3 (20 - 24 March 1995): MPP USAGE

Overview of parallel processing

Architecture of the Cray T3D

Message passing

Running a program on the Cray T3D

Optimisation & debugging tools.

Those attending module 3 are expected to know basic UNIX commands, be able to use the *vi* editor and to be able to submit jobs to UNICOS.

- Andrew Lea

* * * * * * * * * * * *

ECMWF CALENDAR 1994

14 - 16 November	Workshop - Modelling and assimilation of clouds
21 - 25 November	Workshop - Parallel processing in meteorology
1 - 2 December	Council, 41st session
23 - 27 December	*ECMWF HOLIDAY*

* * * * * * * * * * * *

ECMWF PUBLICATIONS

TECHNICAL MEMORANDA	NO. 199	Assimilation of radio occultation measurements into a numerical weather prediction system
	NO. 200	Recommendations on the verification of local weather forecasts
	NO. 201	Error growth and predictability estimates for the ECMWF forecasting system
	NO. 202	The ECMWF ensemble prediction system: methodology and validation
	NO. 203	Sensitivity of two-day forecast errors over the Northern hemisphere to initial conditions
	NO. 204	Recent studies of semi-Lagrangian advection at ECMWF

ECMWF PUBLICATIONS (continued)

TECHNICAL REPORTS	NO. 71	Results with a coupled wind wave model
	NO. 72	Implementation of the semi-Lagrangian method in a high resolution version of the ECMWF forecast model
ECMWF Workshop Proceedings		Fourth Workshop on Meteorological Operational Systems, 22-26 November 1993
ECMWF Seminar Proceedings		Developments in the use of satellite data in numerical weather prediction, 6-10 September 1993

Forecast and verification charts to 30 September 1994

* * * * * * * * * * * *

INDEX OF STILL VALID NEWSLETTER ARTICLES

This is an index of the major articles published in the ECMWF Newsletter series. As one goes back in time, some points in these articles may have been superseded. When in doubt, contact the author or User Support.

		No	Newsletter Date	Page
GENERAL				
ECMWF publications - range of		26	June 84	16
Technical Advisory Committee and Computing Representatives, Meteorological Contact Points		61	Mar 93	46
COMPUTING				
Data handling system		57	Mar 92	16
ECMWF Documentation				
- Current Computer Bulletins		64	Dec 93	46
- On line		64	Dec 93	44
Fortran 90 features in CF77		60	Dec 92	23
	&	61	Mar 93	29
	&	62	June 93	26
	&	63	Sept 93	33
	&	65	Spring 94	20
	&	66	Summer 94	34
MAGICS - the ECMWF graphics package		62	June 93	15
Massively parallel computing - ECMWF's current investigations		61	Mar 93	15
Multitasking ECMWF spectral model		60	Dec 92	3
Networks				
- ECMWF's internal network		54	June 91	20
- New LANs at ECMWF		59	Sept 92	20
POSIX shell under UNICOS 8		66	Summer 94	17
Supervisor Monitor Scheduler (SMS)		59	Sept 92	13
Telecommunications				
- The ECNET system		54	June 91	16
- Digital links to Member States		57	Mar 92	19

		No	Newsletter Date	Page
COMPUTING (continued)				
Workstations at ECMWF		55	Sept 91	7
	&	61	Mar 93	21
Y-MP/C90: An introduction		57	Mar 92	10
Y-MP/C90: I/O optimisation techniques		64	Dec 93	21
METEOROLOGY				
Comparison between SSM/I and ECMWF total precipitable water		57	Mar 92	3
Data acquisition - ECMWF's new system		46	June 89	21
Development of the operational 31-level T213 version of the ECMWF forecast model		56	Dec 91	3
Envelope orography - discussion of its effects		33	June 86	2
ECMWF Analysis				
- New version of analysis system		35	Sept 86	16
- Divergent structure functions		42	June 88	2
- Revised use of satellite data		39	Sept 87	4
- The variational analysis scheme - main features and some early results		62	June 93	5
- Use of TOVS satellite data at ECMWF: a new approach		61	Mar 93	3
ECMWF Preprocessing - new scheme		43	Sept 88	3
ECMWF Re-analysis (ERA) Project - plans and current status		64	Dec 93	3
Ensemble prediction		58	June 92	5
Ensemble prediction system - expert meeting		63	Sept 93	18
Ensemble prediction system: status and plans		65	Spring 94	3
ERS-1 mission		54	June 91	8
Evaporation from tropical oceans		51	Sept 90	3
Forecast model				
- Cloud cover scheme		29	Mar 85	14
- Convection - parametrisation of		43	Sept 88	6
- Increased resolution - studies of		38	June 87	10
- Initial conditions - the spin-up problem		39	Sept 87	7
- New surface/boundary layer formulation		63	Sept 93	3
- Parametrisation of gravity wave drag		35	Sept 86	10
- Revisions to physics		46	June 89	3
- Revision of the clear-sky and cloud radiative properties		61	Mar 93	3
Global forecast experiment at T213 resolution		41	Mar 88	3
Good prediction of a severe storm over southern Sweden		50	June 90	10
GORBUSH - a storm in the Mediterranean Sea		53	Mar 91	4
MARS - the ECMWF meteorological archival		32	Dec 85	15
and retrieval system	&	33	Mar 86	12

METEOROLOGY (continued)

	No	Newsletter Date	Page
Meteorological applications at ECMWF utilizing EMPRESS	64	Dec 93	11
Minimum temperature forecasts at the Regional Meteorological Service of the Emilia Romagna region (N. Italy) by the application of the Kalman filter technique	60	Dec 92	9
Monte Carlo forecast	49	Mar 90	2
Performance of the ECMWF model in tropical cyclone track forecasting over the western north Pacific during 1990-1991	58	June 92	16
Potential vorticity maps at ECMWF	50	June 90	3
Recent verification of 2m temperature and cloud cover over Europe	54	June 91	3
Skill forecasting - experimental system	40	Dec 87	7
Surface wind observations from the ERS scatterometers	66	Summer 94	3
Systematic errors - investigation of, by relaxation experiments	31	Sept 85	9
Use of reduced Gaussian grids in spectral models	52	Dec 90	3

* * * * * * * * * * * *

USEFUL NAMES AND 'PHONE NUMBERS WITHIN ECMWF

		Room*	Ext.**
DIRECTOR	- David Burridge	OB 202	2001
DEPUTY DIRECTOR and HEAD OF OPERATIONS DEPARTMENT	- Michel Jarraud	OB 010A	2003
ADVISORY: Available 9-12, 14-17 Monday to Friday			2801
Other methods of quick contact:	- Telefax (+44 1734 869450) - VMS MAIL addressed to ADVISORY - Internet mail addressed to Advisory@ecmwf.co.uk		
REGISTRATION			
Project Identifiers	- Pam Prior	OB 225	2384
User Identifiers	- Tape Librarian	CB Hall	2315
COMPUTER OPERATIONS			
Console	- Shift Leaders	CB Hall	2803
Console fax number	- +44 1734 499 840		
Reception Counter	- Tape Librarian	CB Hall	2315
Tape Requests	- Tape Librarian	CB Hall	2315
Terminal Queries	- Norman Wiggins	CB 026	2308
Telecoms Fault Reporting	- Michael O'Brien	CB 028	2306
ECMWF LIBRARY & DOCUMENTATION - DISTRIBUTION	- Els Kooij-Connally	Library	2751
LIBRARIES (ECLIB, NAG, etc.)	- John Greenaway	OB 226	2385
METEOROLOGICAL DIVISION			
Division Head	- Horst Böttger	OB 007	2060
Applications Section Head	- John Hennessy (acting)	OB 014	2400
Operations Section Head	- Bernard Strauss	OB 328	2420
Meteorological Analysts	- Andreas Lanzinger	OB 314	2425
	- Ray McGrath	OB 329	2424
	- Anders Persson	OB 315	2421
Meteorological Operations Room	-	CB Hall	2426

		Room*	Ext.**	Bleeper
COMPUTER DIVISION				
Division Head	- Geerd-R. Hoffmann	OB 009A	2050	150
Systems Software Sect.Head	- Claus Hilberg	OB 104A	2350	115
User Support Section Head	- Andrew Lea	OB 227	2380	138
User Support Staff	- Antoinette Alias	OB 224	2382	154
	- John Greenaway	OB 226	2385	155
	- Norbert Kreitz	OB 207	2381	156
	- Dominique Lucas	OB 206	2386	139
	- Pam Prior	OB 225	2384	158
Computer Operations				
Section Head	- Peter Gray	CB 023	2300	114
Security, Internal Networks and				
Workstation Section Head	- Walter Zwieflhofer	OB 140	2352	145
GRAPHICS GROUP				
Group Leader	- Jens Daabeck	CB 133	2375	159
RESEARCH DEPARTMENT				
Head of Research Department	- Anthony Hollingsworth	OB 119A	2005	
Computer Co-ordinator	- David Dent	OB 123	2702	

* CB - Computer Block
OB - Office Block

** The ECMWF telephone number is READING (01734) 499000, international +44 1734 499000, or direct dial to (01734) 499 + last three digits of individual extension number, e.g. the Director's direct number is (01734) 499001.

DEC MAIL: Contact scientific and technical staff via VMS MAIL, addressed to surname.

Internet: The ECMWF address on Internet is **ecmwf.co.uk**
Individual staff addresses are firstname.lastname, e.g. the Director's address is David.Burridge@ecmwf.co.uk