TECHNICAL MEMORANDUM

235

Use of the Fujitsu VPP700 for weather forecasting at ECMWF

D. Dent, G.R. Hoffmann,
P.A.E.M. Janssen and A.J. Simmons

Research Department

May 1997

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

Use of the Fujitsu VPP700 for weather forecasting at ECMWF

D. Dent, G-R. Hoffmann, P.A.E.M. Janssen and A.J. Simmons
European Centre for Medium-Range Weather Forecasts
Shinfield Park, Reading RG2 9AX, U.K.

Abstract

On 18 September 1996 the operational numerical weather forecasting activities of the European Centre for Medium-Range Weather Forecasts (ECMWF) were transferred to run on a Fujitsu VPP700/46 computer system, after many years of use of shared-memory Cray systems. The range of current forecasting activities is described, and further information is given on the computational design and performance of the atmospheric forecast model, and on the operation of the VPP700. Developments which will utilize planned enhancements of the Fujitsu computer system at ECMWF are summarized.

1. Introduction

Numerical weather prediction has been close to the forefront in its use of the most advanced available computing facilities for almost fifty years following the pioneering forecasting experiments reported by Charney, Fjörtoft and von Neumann[1]. Until recently, the computational requirement was determined principally by the need to model the complex behaviour of the atmosphere with as fine a resolution as possible, in order to produce a realistic simulation of the evolution of weather systems within an elapsed time of at most a few hours. Use of finer resolution has been made possible both by substantial increases in computational power and by development of more efficient numerical techniques. Much improved and more comprehensive representations of the dominant physical processes have also been introduced. New types of observation have become available and better ways have been developed for processing observations to define the starting state for the atmospheric model. The resulting increases in accuracy have been dramatic both for short-range forecasts (from a few hours to a few days ahead) and for medium-range forecasts (from a few days to a week or two ahead). For example, the level of accuracy reached on average by three-day forecasts in the first medium-range prediction experiments by Miyakoda et al.[2] in 1972 was reached typically by five-day forecasts in 1979/80 during the first winter of operational forecasting at ECMWF. A measure of the improvement in ECMWF forecasts since 1980 is presented in Fig. 1. This indicates a further two-day increase in the forecast range at which a given level of accuracy is reached.

The first operational ECMWF forecast model used finite-differences and covered the globe with a regular 1.875^{o} grid. In the vertical, atmospheric variables were represented at 15 levels and soil variables at 3 levels. A ten-day forecast took between four and five hours to produce on a single-processor Cray-1 vector computer. The effective computation rate was about 40 million floating-point operations each second (40Mflops) including operational post-processing overheads, and about 50Mflops without overheads. A model using the spectral transform technique was introduced with T63 truncation in 1983, and multi-processing was introduced when the horizontal resolution was increased to T106 in 1985 on a two-processor Cray X-MP machine. The computation rate of this model exceeded 300Mflops on a four-processor X-MP in 1986, and the Gflops range was reached when an eight-

processor Cray Y-MP system was installed in 1990. The numerical formulations used operationally over this period have been reviewed in [3], and computational aspects of the multi-tasking spectral model discussed in [4].

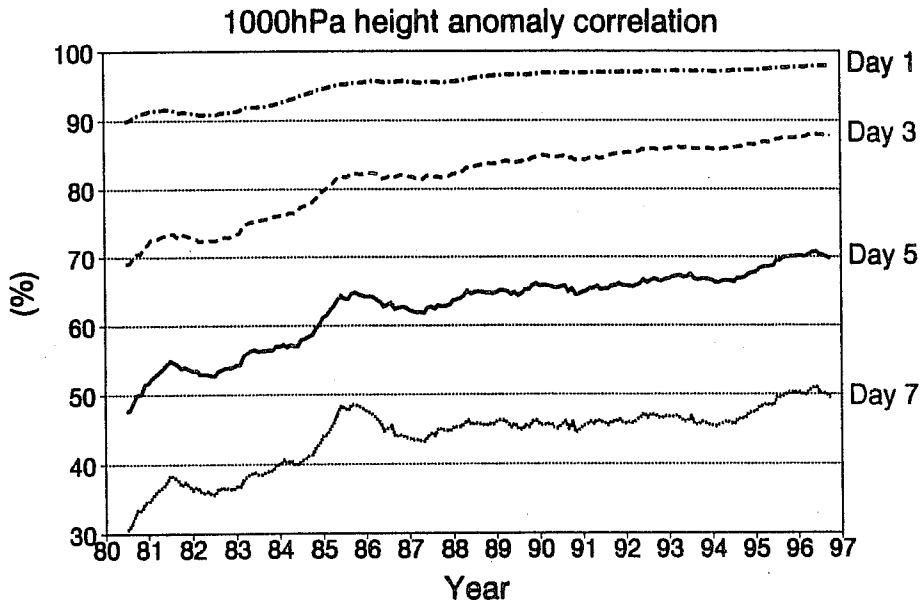

Fig. 1: Correlation between forecast and analyzed anomalies in the height of the 1000hPa pressure surface for forecast ranges of one, three, five and seven days. Plots are based on monthly averages of values computed daily from 1 January 1980 to 31 March 1997 for the extratropics. An annual running mean has been applied.

Ritchie et al.[5] have described the numerical formulation of the forecast model introduced operationally at ECMWF in 1991, and presented some computational details of its performance on the Centre's 16-processor Cray Y-MP/C90 computer. The horizontal resolution was increased to T213, vertical resolution was increased to 31 levels (L31), and a semi-Lagrangian treatment of advection was introduced with a time-step of 15 minutes. A ten-day forecast could be produced in about 75 minutes when run without overheads using 16 processors, at a computation rate approaching 6Gflops. The main operational forecast was multi-tasked over 14 processors, and with post-processing jobs and some other general work running alongside, was typically completed in around two hours.

A stage has now been reached at which refinement of the model resolution is no longer the predominant direct reason for requiring higher computing performance, at least for medium-range prediction. Better estimation of the initial state offers an important path to more accurate individual (deterministic) forecasts, although there is certainly still scope for benefit from model improvement. A very promising, but computationally demanding, method for better exploitation of observational data, particularly newer types, is four-dimensional variational data assimilation[6]. Additionally, the use of forecast information in the later medium range is hindered by variations in forecast quality which reflect both uncertainty in initial conditions and the varying degrees of predictability of different flow regimes. This has led to development of ensemble (probabilistic) prediction systems based upon multiple low-resolution integrations from perturbed initial states[7]. It was principally to satisfy the computational needs of advanced assimilation methods and ensemble prediction that ECMWF installed a 46-processor Fujitsu VPP700 system in summer 1996, providing about a five-fold increase in

computational power over its C90 system. Enhancements of the Fujitsu system will have provided a further increase by a factor of around five or more by early 1999.

2. The operational ECMWF forecasting system on the VPP700

2.1 The T213L31 forecast model

ECMWF has continued to use the T213L31 resolution for its primary forecast model since transferring its operational activities to the VPP700. A ten-day medium-range forecast is produced daily from initial conditions valid for 12UTC, and a three-day forecast is produced daily from 00UTC data to provide boundary conditions for the limited-area short-range forecast models of a number of ECMWF Members States. The same T213L31 model is used in the process of assimilating observations described in section 2.2, and a lower-resolution version is used in the Ensemble Prediction System (EPS) described in section 2.3. An account of the adaptation of this model to run on the VPP700 and of its computational performance is given in section 3.

The distribution of the 31 levels at which atmospheric variables are represented is illustrated in Fig. 2. Coordinate surfaces rise up over orographic features at low levels, and flatten with increasing height to

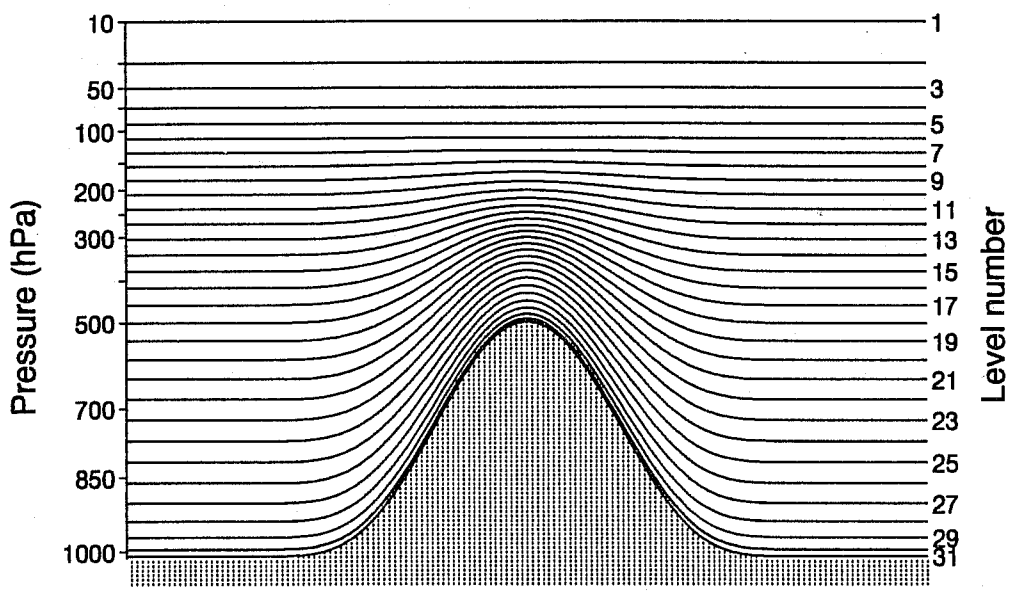

Fig. 2: Distribution of the 31 model levels, showing the dependence on a variation in surface pressure.

become surfaces of constant pressure with 20hPa spacing for the uppermost stratospheric levels. The lowest level is located at a height of about 30m and the lowest five levels lie within about the lowest kilometre of the atmosphere, where higher resolution is needed to model turbulent exchange. Vertical resolution is finer than 1km over most of the troposphere. The uppermost level is at a pressure of 10hPa, which occurs at a height of around 30km.

Two different representations are used to describe the horizontal variation of model fields. Series expansions in terms of spherical harmonics truncated at total wavenumber 213 are used to represent vorticity, divergence, virtual temperature and the natural logarithm of surface pressure, all of which are prognostic variables. The fixed model orography is also represented spectrally. Legendre and Fourier

transforms are used to evaluate fields at model grid-points, including the horizontal winds derived from vorticity and divergence and the required horizontal derivatives. From these values, the (diagnostic) vertical velocity and the horizontal pressure gradients are calculated, and basic dynamical and thermodynamical tendencies evaluated, using a semi-Lagrangian scheme for the advection process. Additional variables are represented by their values at the model grid-points. The three-dimensional fields are specific humidity, cloud water/ice content and fraction of grid volume occupied by cloud, and the semi-Lagrangian advection scheme is used for these fields also. Soil moisture and temperature are predicted for four soil layers, and snow depth is a further prognostic variable. Other fields held constant during the forecast, such as sea-surface temperature, vegetation and land-surface roughness, are also defined at the grid-points. The basic tendencies are augmented by sets of tendencies representing "parametrized" physical processes not explicitly represented in the governing "primitive" meteorological equations. These include the effects of incoming and outgoing radiation, latent heat release and other changes associated with stratiform precipitation and convection, turbulent mixing, unresolved orographically-forced wave motion, land-surface processes and wind-speed dependent ocean roughness. The variables represented at grid-points are then updated, and transformations made back to spectral space to update the fields represented by spherical-harmonic expansions. Further details and references are given in [5].

Examples of the model grid, indicating the categorization of points as land or water, are presented in Fig. 3. The points lie along lines of latitude which are located to facilitate the Legendre transform and which at T213 resolution are almost uniformly spaced with a separation of 0.56^{o} or 62km. Points are distributed in the east-west so as to maintain a similar spacing in terms of distance, although the numbers of points around latitude circles are constrained to be suitable for an efficient Fast Fourier Transform, and points are spaced slightly closer together near the poles. It can be clearly seen in Fig. 3 that points do not line up along lines of longitude (except at the Greenwich Meridian), but this does not cause any fundamental computational problem. Further discussion of this "reduced" grid is given in [8] and [9].

The horizontal resolution with which variables are represented spectrally is lower than the grid resolution. The truncation limit at total wavenumber 213 corresponds to a smallest-resolved half-wavelength of 94km, and the effective resolution is reduced a little more by application of a strongly scale-selective horizontal diffusion. The finer grid-point resolution is used in spectral models to give a largely alias-free result when the non-linear terms in the governing equations are projected back into the truncated spectral space, and is essential for computational stability in models using the traditional Eulerian advection scheme. An impression of the coarser spectral resolution can be gained from the maps of the model orography presented in Fig. 4.

The T213L31 model is run on 18 processors of the VPP700 to produce the operational ten-day forecast in an elapsed time of about one hour. The remaining processors are active at the same time carrying out product-generation tasks, computing perturbations to the analyzed initial conditions for use in the EPS and subsequently beginning to execute the EPS forecasts.

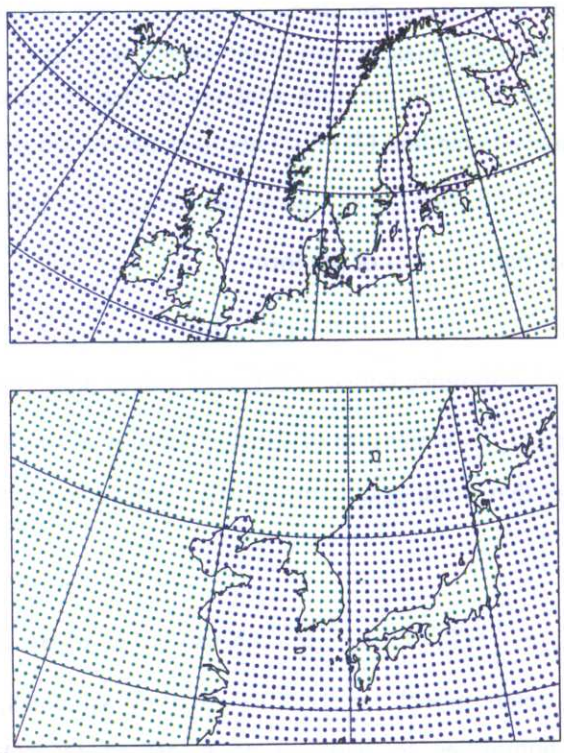

Fig. 3: Distribution of model grid-points for T213 resolution over northern Europe and eastern Asia. The blue circles indicate points treated as water or ice, and the green circles indicate points treated as land.

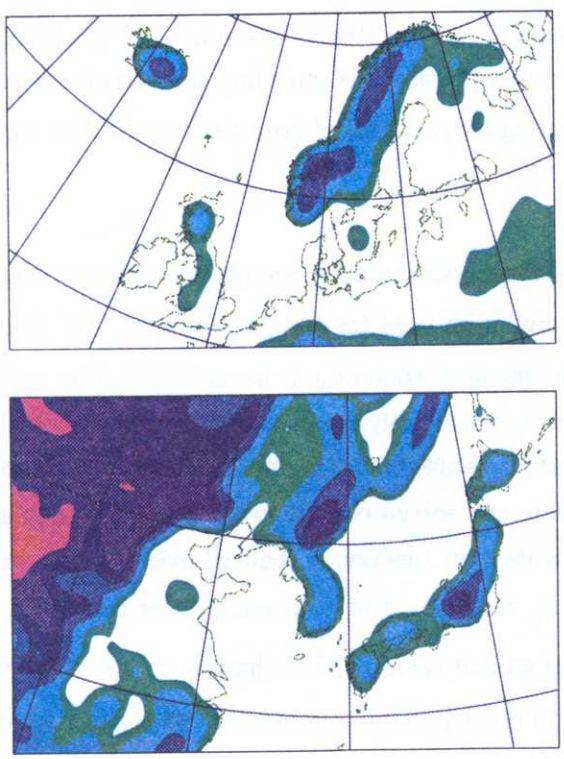

Fig. 4: Distribution of the T213 model orography for the same regions as shown in Fig. 3. Shading starts at a height of 150m, and colour changes occur with a height interval of 300m.

2.2 Variational data assimilation

The processing of observations to define the starting conditions (the initial "analysis") for the forecast model is carried out in six-hourly cycles. It uses observations made at or close to the current analysis time and a six-hour forecast from the previous analysis as "background" information. The "three-dimensional" variational analysis method (3D-Var) involves determining the model state x at time t_1 that minimizes a scalar "cost" function $J(x)$, given the background (forecast) estimate x_b of x and the set of observations o for time t_1. J is of the form:

$$J = J_o + J_b + J_c$$

where J_o is a measure of the difference between observed values and corresponding model estimates (appropriate functions of x), J_b is a measure of the deviation of x from x_b, and J_c is a safety term which penalizes any changes that would lead to unrealistic inertia-gravity oscillations in the subsequent forecast.

J_o is given generally by $J_o = \frac{1}{2}(H(x) - o)^T O^{-1} (H(x) - o)$, where O denotes the covariance of observation errors. $H(x)$ maps model variables to observed variables, for example transforming model temperatures and humidities to radiance as measured by satellite. J_b is of the form $J_b = \frac{1}{2}(x - x_b)^T B^{-1} (x - x_b)$, where B denotes the covariance of errors of the background forecast. Knowledge of B is far from complete, and the dynamical and statistical basis for the construction of J_b is an important element of variational data assimilation systems. The operational version of the ECMWF system implemented in January 1996 is described in [10]; subsequent important developments for the VPP700 have been a new pre-analysis screening and variational quality control of observations and a new method of estimating the background-error variances[11]. An improved formulation of J_b has recently been introduced.

The solution of the variational analysis problem requires knowledge of the gradient of J with respect to x at each iteration of the algorithm used for the minimization. The gradient is computed efficiently using adjoint techniques, but the dimension of x is nevertheless of the order of 6×10^6 for a model with T213L31 resolution for a combined analysis of wind, temperature and humidity. The minimization problem is reduced to a more manageable size by formulating the problem in terms of the deviation of x from its background value, and solved only at lower horizontal resolution (currently T63). Fig. 5 presents an example of how the cost function is reduced over the iterations of the minimization. The reduction comes from the J_o term; J_b, and to a much lesser extent J_c, inevitably increase as x is modified away from the background value x_b. The sharp decrease of J_o after about 30 iterations arises because of the introduction of quality control of observations at this stage of the minimization.

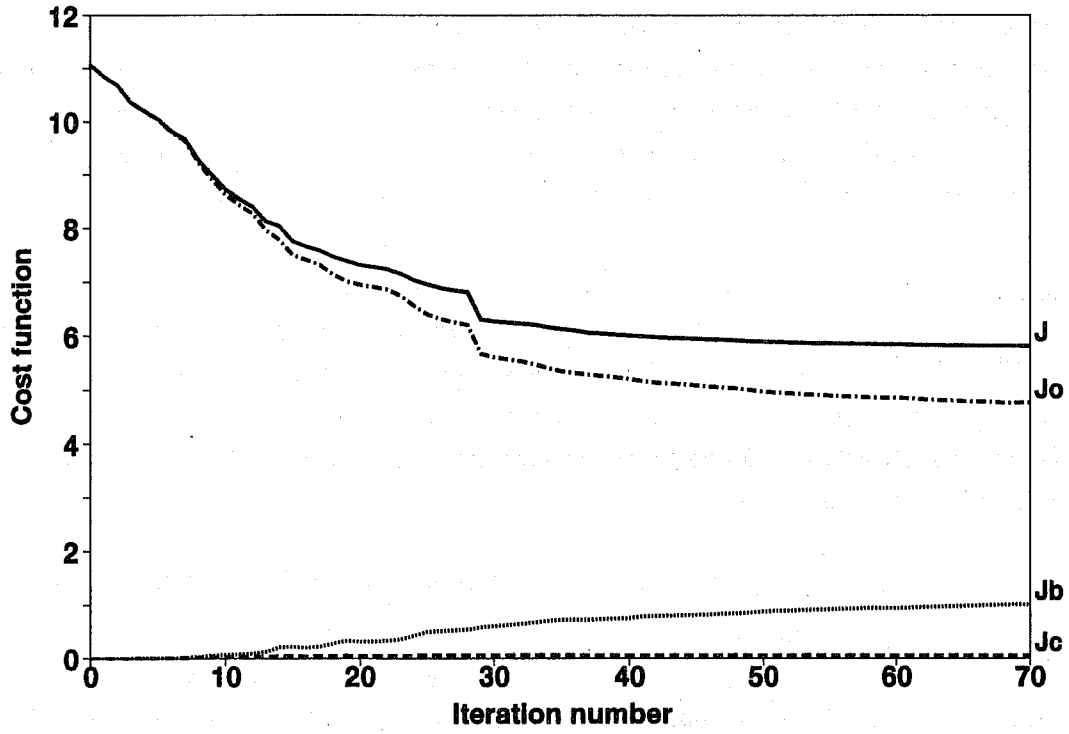

Fig. 5: Change in the cost function and its components over the iterations of a 3D-Var minimization.

Isaksen and Barros [12] have discussed the parallelization of variational data assimilation. A wide range of types of observation are used operationally, including in-situ measurements from land-stations, balloons, buoys, ships and aircraft, and several types of remotely-sensed data from satellites. Their processing poses particular problems for parallelization as they are distributed far from uniformly in space, and the amount of calculation required, for example in computing the "observation operator" $\boldsymbol{H}(\boldsymbol{x})$, varies considerably from one type to another. However, the current operational configuration of 3D-Var is not especially demanding computationally, and the main six-hourly analysis task is parallelized over six processors and completes in an elapsed time of 30-40 minutes. Other tasks (executed on either one or six processors) require a further 15 minutes or so.

2.3 Ensemble prediction

An ensemble prediction system (EPS) has been run daily at ECMWF since May 1994, following successful results in thrice-weekly trials started in December 1992[7]. The principal aims are to provide an estimate of the skill of the deterministic T213L31 forecast, to provide possible alternative evolutions of the atmospheric circulation pattern, and to provide probabilities of the occurrence of specific weather events. The system entails execution of a set of numerical forecasts using a lower-resolution model, starting from suitably perturbed initial conditions. T63 horizontal resolution and 19-level vertical resolution were used in the original implementation of the system. A control forecast was run starting from a truncated and interpolated form of the high-resolution operational analysis, and a further 32 forecasts were run from the perturbed initial conditions.

Acquisition of the VPP700, and algorithmic improvements in model efficiency, enabled a substantial improvement to be made in the configuration of the EPS in late 1996. The 31-level vertical resolution

of the main T213 forecast was adopted for the EPS forecast model, and the grid of this model was refined to be that appropriate for a conventional Eulerian T106 model, with a spacing of about 125km. The so-called "linear-grid" option was introduced, exploiting the fact that for a given grid-resolution, semi-Lagrangian models can run stably with a higher spectral resolution than can be used in Eulerian models. The spectral truncation was increased to 159, this linear-grid resolution being referred to as T_L159. In addition, the number of forecasts from perturbed initial conditions was increased to 50.

Execution of the set of forecasts is the most computationally demanding part of the EPS and is ideal for effective use of parallel processing. The memory requirement is such that each perturbed forecast can be run in just two processors of the VPP700. The control forecast and 50 forecasts from perturbed initial conditions are allocated 34 processors, and these processors can be kept well-occupied working on the forecasts in batches of 17 two-processor forecasts run in parallel. A net time of around 2.5 hours is needed to complete the forecasts.

The initial perturbations for the EPS are constructed from the "singular vectors" that grow most rapidly as deviations from the control forecast during the first two days of the forecast range[13]. The singular vectors are combined linearly to produce initial perturbations for the EPS that have a widespread geographical distribution and magnitudes consistent with estimates of analysis error. A Lanczos algorithm is used to compute the singular vectors, and it requires repeated evaluations of an operator obtained by a forward two-day integration of a model version which is linearized about the control forecast (and has simplified physical parametrizations) followed by backward integration of the adjoint of this model. Some 70 or so iterations are needed to obtain a suitable set of singular vectors, and the cost is thus equivalent to that of several hundred days of forward integration of the forecast model. At T42L31 resolution (as implemented operationally on eight processors) the calculation requires about 30% of the computation needed to execute the ten-day T213L31 forecast.

2.4 Ocean-wave forecasting

Ten-day global forecasts of ocean-wave conditions are produced daily using a wave model driven by the surface winds from the atmospheric-model forecast. The wave model is the third-generation WAM model[14]. It predicts the two dimensional wave spectrum and is based on an explicit description of the physical processes governing wave evolution, which include wind input, dissipation by white-capping and bottom friction, and nonlinear wave-wave interactions. The initial wave analysis is based on observations from the altimeter on the European Space Agency's ERS-2 satellite and on background wave information produced by the wave model using winds from the background forecast of the atmospheric data assimilation. The method of optimum interpolation is used to adjust the background state towards observed values[15].

The move to the VPP700 enabled a significant increase in the spatial resolution of the wave model. Currently, the predicted wave spectrum has 12 directions and 25 frequencies so there are 300 degrees of freedom per grid-point. The grid is an irregular latitude-longitude grid, which has a constant latitudinal increment but adjusts the size of the longitudinal increment in such a way that the distance between grid-points is almost constant, much as in the atmospheric model. The resolution is about 55km and

involves about 120,000 grid-points. Hence, the WAM model solves about 35,000,000 equations per time-step of 15 minutes.

The WAM model is coded so that the inner do loop is over the grid-points, to take full advantage of vectorisation. This is achieved by mapping the two dimensional grid into a one dimensional array, ignoring land points. On one processor the inner do-loop would have a length of 120,000, and the elapsed time of a ten-day forecast would be three hours or more. Speed-up has been achieved by distributing the grid-points over several processors and introducing message passing. Since the advection of wave energy is done by a simple first-order up-winding scheme, the overlap with neighbouring sub-domains may be precalculated and used to determine the information needed in the messages sent to the other processors. Four processors are used operationally. The elapsed time is about 50 minutes and the efficiency of speed-up is between 92% and 95%.

3. Implementation and computational performance of the ECMWF atmospheric model on the VPP700

Migration of the atmospheric model from the Cray shared-memory platforms required source code developments to remove non-standard features from the Fortran language, and to implement parallelism in a distributed memory environment. Fortran 90 features were utilized to replace Cray-specific code features, and the MPI message-passing technique was adopted. The code is now fully portable to any platform with a Fortran 90 compiler. Although the message-passing approach is rather manual and requires non-trivial effort from experienced programmers, the rewards are a portable parallel code with potentially excellent parallel performance. The alternative technique of using a compiler-specific set of directives to achieve a data parallel execution is much easier to implement but is unlikely to be portable or efficient on a range of parallel platforms.

The key to creating an efficient message-passing version lies in the identification of the inherent parallelism. Calculations in the different algorithmic steps of the spectral model involve data dependencies in different directions. For example, the Fourier transforms involve dependent zonal-wavenumber coefficients and grid-point values, but can be carried out independently for different model variables, levels and lines of latitude. Evaluation of a Legendre transform requires data for one particular zonal-wavenumber coefficient for all latitudes. In grid-point space, the physical computations involve vertical data coupling only. However, the semi-Lagrangian advection scheme requires access to data from nearby grid-columns, since for a particular grid-column, departure points at the beginning of a time-step (and corresponding field values) have to be computed for the air-parcels which arrive at that column at the end of the time-step. A degree of local communication is thereby introduced.

A transposition strategy is used to handle the problem of accessing distributed data. With this approach, the relevant data is redistributed among the processors at various stages of the algorithm so that the arithmetic computations between any two consecutive transpositions can be performed without any interprocessor communication. Such an approach is feasible because data dependencies exist only within one coordinate direction for each algorithmic component. An overwhelming practical advantage of this technique is that the message-passing logic is localized in a few routines. These are executed prior to the appropriate algorithmic stage, so that developers working with the computational routines (which constitute the vast bulk of the model code) need have no knowledge of this activity.

The distribution of work has to be done carefully to achieve good load balancing[16]. For example, in grid-space the total array of grid-columns (138346 for T213 resolution) can be evenly distributed among any number of available processors such that each receives an almost equal amount of work. For a partition of 100 processors, 54 will receive 1383 of the T213 data points, and 46 will receive 1384 data points. Each processor works on a sub-domain of the sphere whose shape (latitudinal and longitudinal extent) may be adjusted for optimal performance.

Such load balancing is "static" and is independent of the particular weather situation. A dynamic imbalance also exists due to variations in the amount of computation for particular grid-columns arising from the meteorological conditions. For example, computations for a grid-column near to the equator may have to account for substantial amounts of convective activity, while those for a polar column may have to deal with none. This dynamic imbalance is much more difficult to handle since, in principle, its reduction requires a redistribution of grid-points to processors at each model time step.

Special treatment of the reduced grid is needed to achieve good vector performance. The simplest approach of vectorising over a complete line of latitude (a row) would result in a vector length varying from 640 for equatorial rows to only 16 at the poles for the case in which each processor is allocated one or more complete rows. To prevent inefficiencies due to short vectors, a buffer is provided for each processor into which can be packed data from any of the complete or partial rows assigned to that processor. The resultant "super-row" is presented to the grid-point computational routines for processing. The buffer size is defined at run time and this allows the user to optimise vector length at the expense of the memory needed for temporary work-space. The success of this technique on the VPP700 may be judged by utilizing the profiling tool. This indicates that the application achieves a total sustained floating point computational rate of about 770 Mflops on each processor (35% of peak).

Fig. 6 quantifies message-passing and load-imbalance inefficiencies when a benchmark T213L31 version of the model utilizes 46 processors of a VPP700. It can easily be seen that these inefficiencies are small compared to the computational time. Use of a message-trace facility makes it possible to obtain detailed information about any message traffic. Most of the interprocessor traffic within the model proceeds at close to the hardware speed of more than 500 Mbytes/sec.

Performance of the T213L31 model using differing numbers of processors of a VPP700 is shown in Fig. 7. The parallel speed-up is 15.4 with 16 processors and 35.5 with 40 processors. This is based on measurements of total elapsed time, and confirms that the message-passing overheads are not excessive.

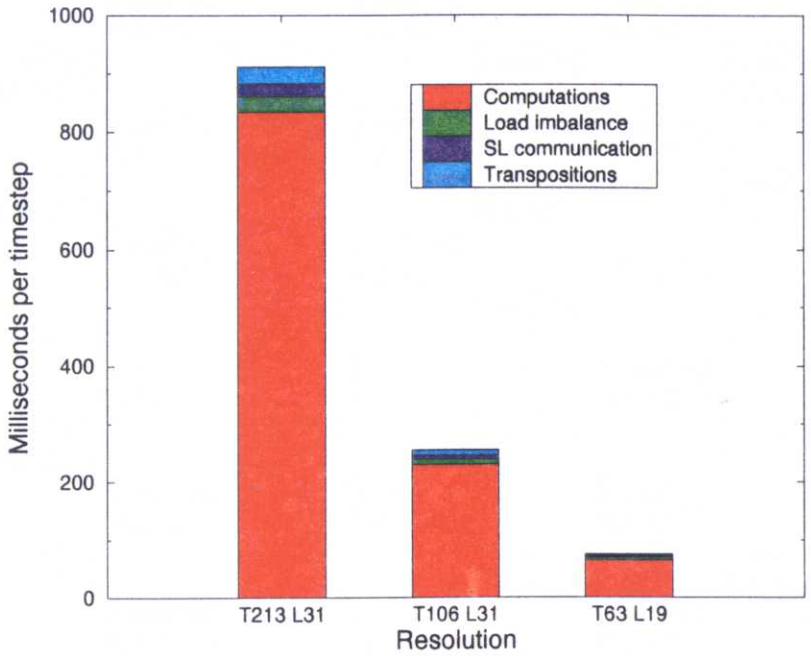

Fig. 6: Breakdown of the time spent in computation and in the message-passing associated with the transpositions and semi-Lagrangian communication, and the loss due to load imbalance, for three versions of the model using 46 VPP700 processors.

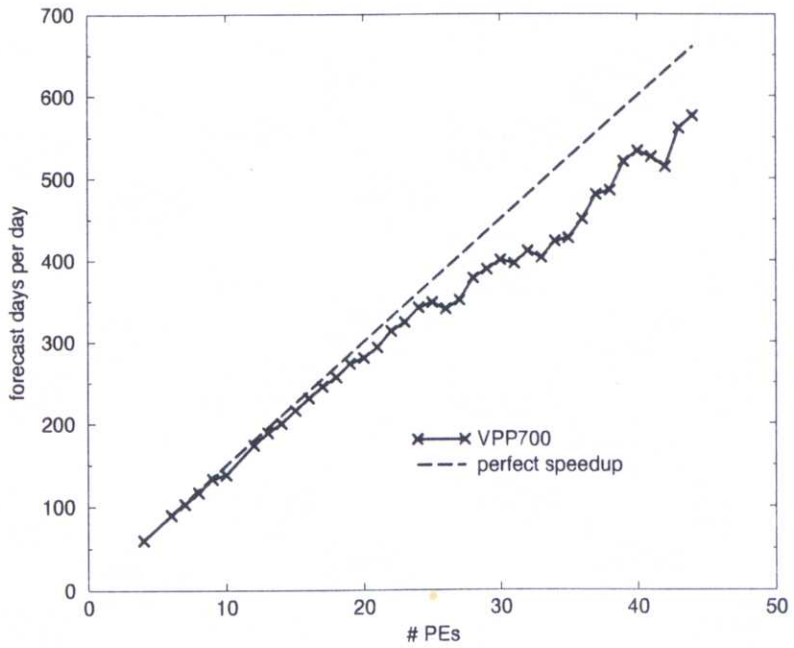

Fig. 7: Number of days that can be simulated per day of wall-clock time by the T213L31 model as a function of the number of VPP700 processors allocated to the model.

4. Operation of the VPP700

The Fujitsu VPP configuration was selected by ECMWF in August 1995 to replace its Cray C90 and T3D systems, after careful preparations and a thorough evaluation of responses to an Invitation to Tender. Details of this process are given in [17]. The VPP700/46 was delivered and installed towards the end of June 1996 and passed its Preliminary Acceptance Test on 2 July 1996. During its subsequent Performance Test it proved that it could run a benchmark version of the ECMWF model at least five times faster than the Cray C90, both for a single T213L31 forecast and for a set of T63L19 forecasts as used in the first version of the EPS. The VPP700 passed Final Acceptance on 8 August 1996 after having successfully completed a 30-day Reliability Test. It became the main operational machine of ECMWF on 18 September 1996. Its integration with the other components of the ECMWF computing environment is shown in Fig. 8.

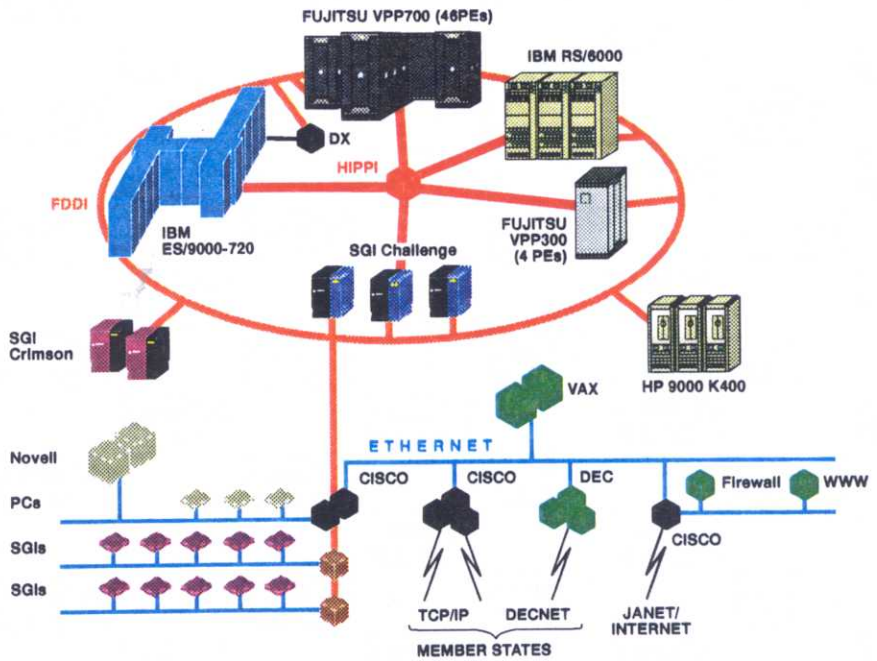

Fig. 8: Computer configuration of ECMWF in October 1996.

The VPP700 system has proven to be very stable and no major outages have occurred. The weekly mean availability of the system to users has been 164.3 hours over the period since August 1996. Of the other 3.7 hours of the week, 2.3 have been used on average for software maintenance and system sessions, and 1.2 have been lost due to software crashes. Time used for hardware maintenance or lost due to hardware crashes accounts for the very small remainder.

The growth in utilization of the Fujitsu system has been similarly encouraging. Fig. 9 shows the CPU usage of the complete system (including control and I/O processors) on a week-by-week basis since the beginning of October 1996. The VPP700 is now generally kept well supplied with streams of work based on operational and research versions of the forecasting system, and these streams occupy the processors quite effectively. Net CPU usage has increased to above the 70% level, with usage by the operating system holding more-or-less steady at well under the 10% level.

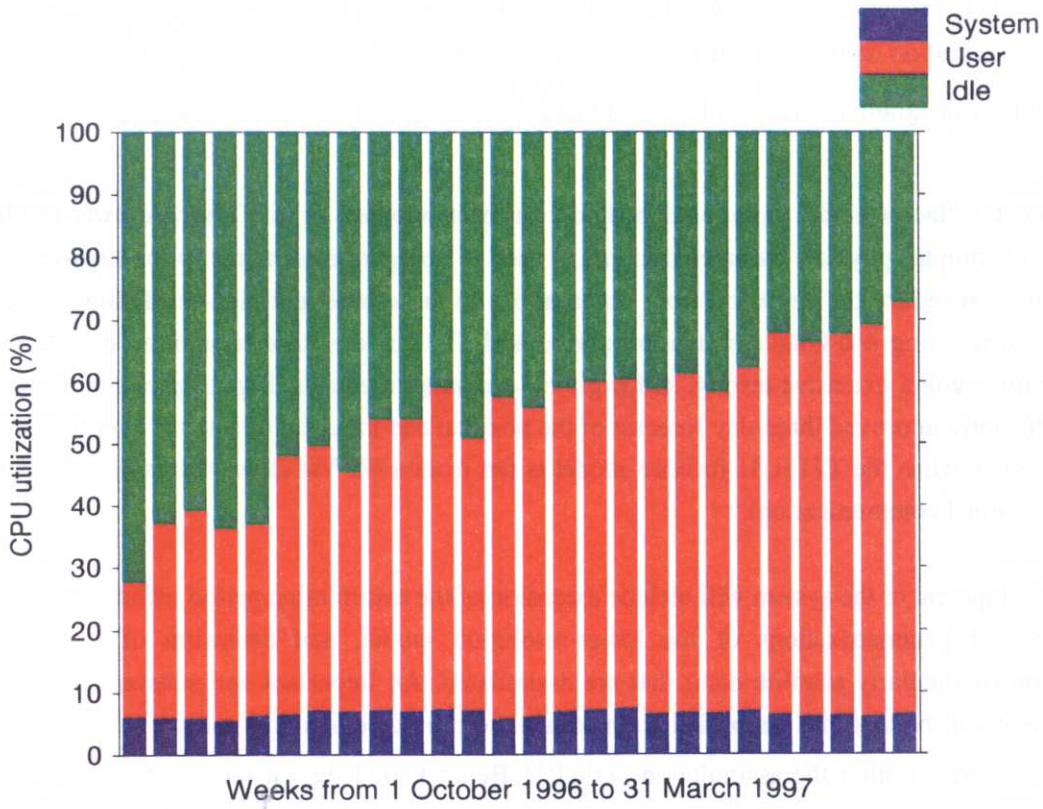

Fig. 9: Weekly CPU utilization of the VPP700/46.

5. Future developments

5.1 Data assimilation

The main current development in data assimilation is that of four-dimensional variational assimilation (4D-Var), the principal application for which the VPP700 was acquired. The formulation of 4D-Var may be set out in a way similar to that of 3D-Var. The variational problem is to determine the model state $\boldsymbol{x}$ at time t_0 that minimizes the cost function J given observations over the *time-range* $t_0 \le t \le t_1$. The observation operator $\boldsymbol{H}(\boldsymbol{x})$ now maps model variables at time t_0 to variables observed at time t, and involves a model integration from t_0 to t, followed by a transformation from model to observed variable as in 3D-Var. The background state $\boldsymbol{x}_b$ for 4D-Var is the optimal forecast valid at the end of the previous assimilation period.

Le Dimet and Talagrand[18] showed how adjoint methods can be used to calculate the required gradient of J with respect to $\boldsymbol{x}$. The J_o term is the most expensive computationally, and its gradient is computed in the following way. First, the full model is integrated forward in time from t_0 to t_1, starting from the latest estimate of $\boldsymbol{x}$, and storing the forecast state at each time-step. Then, the adjoint of the model equations linearized about the forecast states stored during the forward integration is integrated backwards from t_1 to t_0. The adjoint model includes forcing by a term which depends on the

differences between the observations and model estimates based on the stored forecast states. The backward integration starts with all fields set to zero, and the result at time t_0 gives the required components of the gradient of J_o with respect to x.

In practice, the "incremental" variational method[6] has been adopted, as in 3D-Var. A lower-resolution model (with simpler physical parametrizations) is used to compute increments to the high-resolution initial state that reduce the deviation from observations in the subsequent high-resolution forecast. A 4D-Var system using a 6-hour assimilation period and T63 for the lower-resolution calculations is currently undergoing extensive testing, and is giving encouraging results. Fig. 10 presents an example of a significantly improved three-day forecast of the position and intensity of a storm over the northern Atlantic Ocean when the T213L31 forecast model is run from a 4D-Var rather than a 3D-Var analysis (Rabier, personal communication).

Future development of the system will include extension of the assimilation period, enhancement of the resolution and parametrizations of the lower-resolution model, and extension of the types of observation (particularly satellite data) that are assimilated. An important and substantial additional development will be that of a simplified Kalman filter to provide dynamical estimates of the analysis and forecast errors within the assimilation period[11]. Better knowledge of these is vital not only for improving the error covariances used in the variational assimilation, but also for providing more realistic initial perturbations for use in the EPS.

5.2 Modelling

Model development is currently directed principally towards using the linear-grid option for the main forecast model to increase spectral resolution from T213 to T_L319. This change is expected to be accompanied by use of an improved basic dataset for the generation of the model orography and related fixed fields. Later resolution changes are expected to entail a significant increase in the vertical extent of the model, with levels up to 0.05hPa, and a finer resolution of the planetary boundary layer. Enhancement of the Fujitsu computer system in 1999 will permit a substantial increase in horizontal resolution, which may be taken to as high as T_L639, with a 31km computational grid. There will be ongoing refinements and revisions of the treatments of the existing parametrized processes, which may involve additional prognostic variables, for example separating cloud-water and cloud-ice contents and introducing turbulent and convective sub-grid-scale energies. Ozone will also be added as a prognostic variable, requiring its own parametrization of generative and destructive processes.

5.3 Ensemble prediction system

Improvements to the EPS are expected from a number of sources. Particular attention is being concentrated on the construction of more representative initial perturbations, especially through taking into account the estimated analysis error and introducing more complete physical parametrizations in the singular-vector calculations. Accounting for uncertainty in the model formulation as well as in the initial conditions is another issue for examination. Further increases in model resolution and ensemble size will be possible when the Fujitsu system is enhanced, and extension of the range of the EPS beyond ten days is likely. As a relatively new type of forecasting activity, continuing emphasis has to be

placed on exploring ways in which the wealth of data produced by the EPS can be processed to form sets of useful end-products.

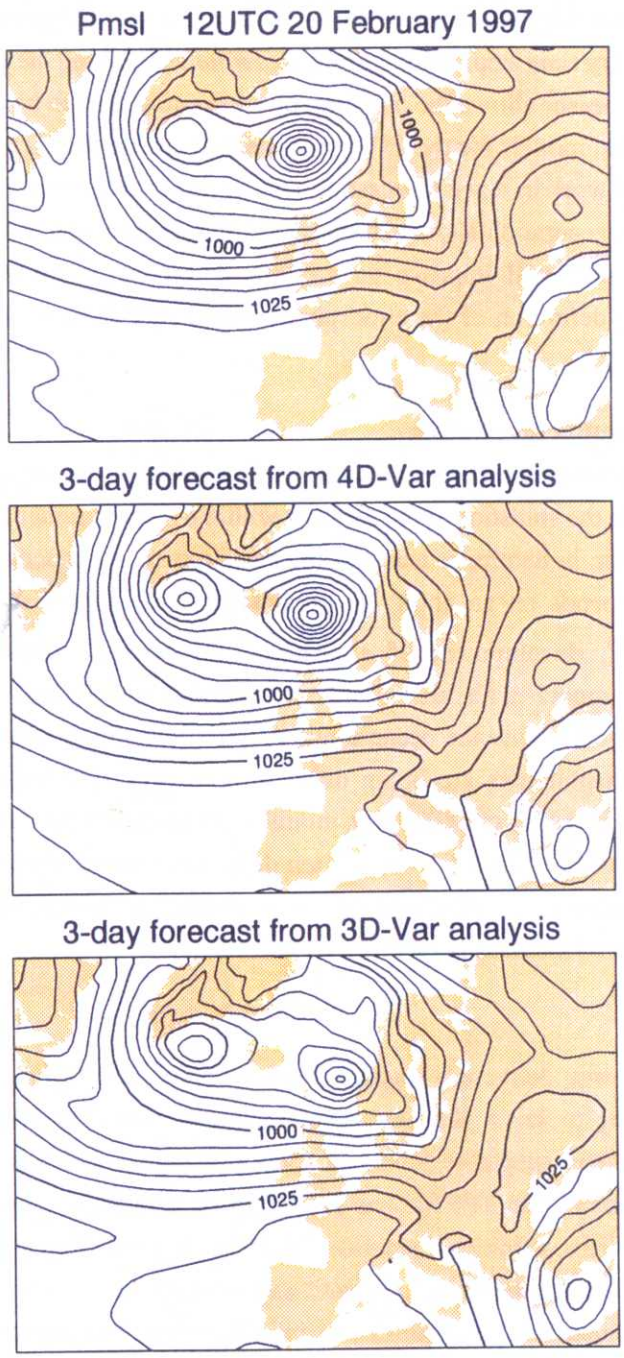

Fig. 10: Maps of mean-sea-level pressure for 12UTC 20 February 1997. The upper plot shows the 4D-Var analysis. The corresponding 3D-Var analysis is similar. Forecasts started from 4D- and 3D-Var analyses for 12UTC 17 February 1997 are shown in the lower two plots. The contour interval is 5hPa.

5.4 Wave prediction

It is planned to introduce a direct coupling of the atmospheric and ocean-wave models in the near future. At each time-step the surface winds from the atmospheric model will be used to update the wave model, which in turn will provide a new estimate of the surface roughness over sea to the atmospheric model. In developing the software for this, two different approaches to distributing the amount of work per processor have been married. This was possible because the interaction between the two models, although occurring every time step, involves only a few fields. Making the global version of these fields available on every processor allows the coupling of wind and waves to be accomplished in an efficient way. Use of the coupled models for the EPS forecasts will enable production of probabilistic forecasts of ocean waves. Other work in wave forecasting will be directed towards improving the representation of the physics of wave generation and feedback on the atmosphere, and increasing the use of wave data from satellites. Such data will improve directly the definition of the initial ocean-wave state and improve indirectly the definition of the atmospheric initial state through use of the coupled system in data assimilation.

5.5 Experimental seasonal forecasting

Work is also being undertaken on prediction on the time-scale of one or more seasons. Such forecasts will inevitably relate to the statistics of the weather within a season, rather than specific weather events, and will be probabilistic in nature, at least for the extratropics. As much of the potential for seasonal weather prediction depends on an ability to predict relatively slowly-developing anomalies in sea-surface temperature[19], development of a seasonal forecasting system requires coupling of an atmospheric forecast model with an oceanic circulation model, and implementation of a data assimilation system to determine the oceanic initial conditions. Such a system has been assembled at ECMWF, and experimental runs are being made in real time. It is expected that this activity will continue, with the system being run regularly, monitored and improved. This development may also be beneficial for the extension of the EPS beyond ten days, since some representation of the evolution of the sea-surface temperature is thought likely to be needed to realize the full potential for skilful prediction at the extended medium range.

6. Conclusion

The Fujitsu VPP700 was delivered, installed and accepted according to plan. After some initial difficulties with getting to know the new architecture, in particular its I/O behaviour, the system performance has reached acceptable levels of CPU utilization, and has been well received by the ECMWF user community. The VPP700 has proven to be a reliable system which has enabled early enhancement of ECMWF's operational activities in ensemble prediction and ocean-wave forecasting. It is providing a good throughput for the research experiments needed to test future enhancements, in particular for the development of the four-dimensional variational assimilation system which is planned for operational implementation later in 1997.

Acknowledgement

An efficient and effective range of operational forecasting activities on the Fujitsu VPP700 has been achieved through the dedicated efforts of very many members of the Research and Operations Departments of ECMWF, and of the staff of Fujitsu in Europe and Japan, to whom we express our gratitude.

References

1) Charney, J.G., Fjörtoft, R.and von Neumann, J.: Numerical integration of the barotropic vorticity equation. *Tellus*, 2, pp. 237-254, (1950).

2) Miyakoda, K., Hembree, G.D., Strickler, R.F., and Shulman, I.: Cumulative results of extended forecast experiments. I: Model performance for winter cases. *Mon. Wea. Rev.*, **100**, pp. 836-855, (1972).

3) Simmons, A.J., Burridge, D.M., Jarraud, M., Girard, C., and Wergen, W.: The ECMWF medium-range prediction models: Development of the numerical formulations and the impact of increased resolution. *Meteorol. Atmos. Phys.*, **40**, pp. 28- 60, (1989).

4) Simmons, A.J. and Dent, D.: The ECMWF multi-tasking weather prediction model. *Comp. Phys. Rep.*, **11**, pp. 165-194, (1989).

5) Ritchie, H., Temperton, C., Simmons, A., Hortal, M., Davies, T., Dent, D., and Hamrud, M.: Implementation of the semi-Lagrangian method in a high resolution version of the ECMWF forecast model. *Mon. Wea. Rev.*, **123**, pp. 489-514, (1995).

6) Courtier, P., Thépaut, J-N., and Hollingsworth, A.: A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart. J. R. Meteorol. Soc.*, **120**, pp. 1367-1387, (1994).

7) Molteni, F., Buizza, R., Palmer, T.N., and Petroliagis, T.: The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. R. Meteorol. Soc.*, **122**, pp. 73-119, (1996).

8) Hortal, M., and Simmons, A.J.: Use of reduced Gaussian grids in spectral models. *Mon. Wea. Rev.*, **119**, pp. 1057-1074, (1991).

9) Courtier, P. and Naughton, M.: A pole problem in the reduced Gaussian grid. *Quart. J. R. Meteorol. Soc.*, **120**, pp. 1389-1407, (1994).

10) Courtier, P., Andersson, E., Heckley, W., Pailleux, J., Vasiljevic, D., Hamrud, M., Hollingsworth, A., Rabier, F., and Fisher, M.: The ECMWF implementation of three dimensional variational assimilation (3D-Var). Part I: Formulation. *Submitted for publication.*

11) Fisher, M., and Courtier, P.: Estimating the covariance matrices of analysis and forecast error in variational data assimilation. *ECMWF Tech. Memo.*, No. 220, 26pp.

12) Isaksen, L., and Barros, S.R.M.: IFS 4D-variational analysis - Overview and parallel strategy. In: *Coming of Age, Proceedings of the Sixth ECMWF Workshop on the Use of Parallel Processors in Meteorology*, Eds. Hoffmann, G-R., and Kreitz, N., World Scientific, Singapore, pp. 337-351, (1995).

13) Buizza, R. and Palmer, T.N.: The singular-vector structure of the atmospheric general circulation. *J. Atmos. Sci.*, **52**, pp. 1434-1456, (1995).

14) Komen, G.J., Cavaleri, L., Donelan, M., Hasselmann, K., Hasselmann, S., and Janssen, P.A.E.M.: *Dynamics and Modelling of Ocean Waves*, Cambridge University Press, Cambridge, 532pp., (1994).

15) Lionello, P., Günther, H., and Janssen, P.A.E.M.: Assimilation of altimeter data in a global third-generation wave model. *J. Geophys. Res.*, **C97**, pp. 14453-14474, (1992).

16) Barros, S.R.M., Dent, D., Isaksen, L., and Robinson, G.: The IFS model - Overview and parallel strategies. In: *Coming of Age, Proceedings of the Sixth ECMWF Workshop on the Use of Parallel Processors in Meteorology*, Eds. Hoffmann, G-R., and Kreitz, N., World Scientific, Singapore, pp. 303-318, (1995).

17) Hoffmann, G-R.: Distributed Memory Systems For ECMWF. In: *Supercomputer 1996.* H-W Meuer (Hrsg.). München, ISBN 3-598-22413-3, pp. 69-77, (1996).

18) F-X. Le Dimet, F.-X. and Talagrand, O.: Variational algorithms for analysis and assimilation of meteorological observations. Tellus, **38A**, pp. 97-110, (1986).

19) Palmer, T.N., and Anderson, D.L.T.: The prospects for seasonal forecasting - A review. *Quart. J. R. Meteorol. Soc.*, **120**, pp. 755-793, (1994).