THE SPECTRAL TECHNIQUE

M. Jarraud and A.J. Simmons
European Centre for Medium Range Weather Forecasts
Reading, U.K.

1. INTRODUCTION

From 1950, and the first barotropic models used in numerical weather prediction, there has been a steady increase in the sophistication of mathematical models of the behaviour of the atmosphere, with both a progressive reduction of the approximations used in deriving the equations and a progressive refinement of the resolution of the models. Since these developments have been closely linked with increases in computing power, they have been beyond our control to a certain extent. In parallel there has also been a steady increase in sophistication and accuracy due to improvement of the numerical techniques used to discretize the continuous equations of the mathematical models. This has been the result of a major research effort carried out in many countries and recognised as a major component of the GARP. The weight of these efforts bore on both spatial and temporal discretisation techniques. We shall here concentrate on the former. Until 1972 almost all numerical models were based on finite-difference (grid-point) techniques and much of the effort was spent on them. Other techniques were regarded more as mathematical recreation than as realistic potential alternatives for operational forecast or general circulation models. However, in the last twelve years there has been a renewed interest and a rapid development of other techniques, and in particular of two of them, namely finite-element methods (which have been widely used in many other sectors of fluid dynamics) and spectral methods (which also have been used in several fields of theoretical and applied physics).

The use of spectral methods in numerical atmospheric models can be traced back to 1942 in the USSR, where Blinova made proposals for long range forecasting, using a linearised model based on expansions using spherical harmonics. In 1952 Haurwitz and Craig also proposed representing atmospheric flow patterns using similar expansions, and in 1954 Silberman solved the non-divergent barotropic vorticity equation with the help of spherical harmonics. Several theoretical studies and applications followed, bringing a number of attractive properties into light, but also unfortunately major drawbacks: the amount of computation and the size of the storage required was rapidly becoming prohibitive when resolution was increased. This was a consequence of the method (interaction coefficients) used to compute the non-linear quadratic terms. Computation of other non-linear terms (division, or more than quadratic products) was quite impracticable. Another major drawback was the inclusion of the so called "physical processes" (convection, precipitation, etc) which did not seem possible. The general feeling was thus that spectral models could be attractive and accurate tools for some limited categories of theoretical problems, mainly those using very coarse resolution (the so called "low order" models) but that they could not be competitive for routine operational numerical weather forecasts. A step forward was made in 1966 when Robert showed how the storage problem could be eliminated by decomposing the expansion functions into simpler elements. However, all other major drawbacks remained as strong as before.

The real breakthrough was the adaptation of transform methods to numerical spectral models worked out independently by Eliassen et al. (1970) and Orszag (1970): the idea is to evaluate all main quantities at the nodes of an associated grid where all non linear terms can then be computed as in a classical grid point model, thereby making possible the inclusion of physical

processes in a straightforward way. The method also considerably reduced the requirements for storage and computations, and it then became possible to envisage spectral models with substantially higher resolutions and an efficiency at least comparable with that of the most efficient grid point models of equivalent accuracy. After some preliminary attempts (Eliassen et al. 1970, Bourke, 1972, Machenhauer and Rasmussen, 1972) several groups developed more complex multi-level hemispheric or global spectral models: Machenhauer and Daley (1972,1974), Bourke (1974), Hoskins and Simmons (1975), Daley et al. (1976). This activity led to the implementation of spectral models for routine forecasts in Australia and Canada in 1976. Now spectral models are used operationally by many weather forecasting centres: in USA (NMC) since 1980, in France since 1982 and in Japan and ECMWF since 1983. Several research groups involved in general circulation studies also use spectral models since this technique has now proved most efficient and easy to implement for long global integrations.

So less than a decade after the introduction of the first primitive equation multi-level spectral model (Bourke 1974) this technique has become the most widely used numerical tool for treating the horizontal part of the equations in hemispheric or global problems. However this does not mean that they represent the ultimate step in numerical techniques for weather prediction: other techniques are being developed, or may be developed in the future, which could turn out to be more efficient methods for a comparable accuracy when the associated technical problems have been mastered.

2. GENERAL PRESENTATION

2.1 Basic description of the spectral method

This Section is partly inspired by the presentation by Machenhauer in publication No.17 of the GARP series (1979).

The equations used in numerical weather prediction can without loss of generality be written:

$$\frac{\partial F_i}{\partial t} = A_i (F_j) \quad j = 1, \ldots, J \tag{1}$$

where J is the number of prognostic variables F_i:

$$F_i = F_i(x,t) \tag{2}$$

with x representing the 3-dimensional space coordinate, and

t the time coordinate.

For all meteorological applications the F_i are supposed to be as smooth (in the mathematical sense, i.e. differentiable) as needed, at least away from the boundaries.

The A_i are operators, generally non linear and involve usually partial spatial derivatives as well as possibly spatial integrals.

When finite-difference ('grid-point') techniques are used, an ensemble of grid points (x_p, t_q) is chosen in both space and time and the continuous operators $\frac{\partial}{\partial t}$ and A_i are replaced by discrete analogues, the complexity of which depends in part on the accuracy required.

It is then possible to replace the system (1) by a new set of evolution equations at the various points x_p with the initial conditions $F_i(x_p,o)$ supposed to be known.

As an alternative the fields F_i at any time are smooth functions of x and as such can be considered as elements of some vector space H (for example the space of all continuously differentiable functions).

If f and g are two functions of H it is possible to define a scalar product:

$$(f,g) = \int_s fg^* \, dx \tag{3}$$

where g^* is the complex conjugate of g,

and a norm:

$$\|f\| = \left\{\int_s |f|^2 \, dx\right\}^{+\frac{1}{2}} \tag{4}$$

With this scalar product and this norm H becomes a so called "Hilbert" space.

Let us now assume that we know an "Hilbertian basis" of H, that is an ensemble of elements $\{e_m(x)\}$ of H, orthonormal (ie. $(e_m,e_n) = 0$ for $n \neq m$ and $(e_m,e_m)=1$) which is such that any element F of H can be expressed as

$$F = \sum_{m \in M} F^m e_m \tag{5}$$

It is not a basis since in general M is not finite.

The F^m are the orthogonal projection of F on the sub space generated by the e_m. In other words:

$$F^m = (F,e_m) = \int_s F \, e_m^* \, dx \tag{6}$$

In order to avoid unnecessary complications, we shall now present the basic principle of the method for a single variable

$$F = F(x,t).$$

The system (1) reduces to the following equation:

$$\frac{\partial F}{\partial t} = A(F) \tag{7}$$

with the initial condition $F(x,o) = f(x)$ (8)

Since we cannot cope with an infinite number of components we have to project F on a finite dimensional sub space $\bar{H}$ of H (a truncation procedure).

F is thus approximated by

$$\bar{F}(x,t) = \sum_{m=1}^{M} F^m(t)\, e_m(x) \tag{9}$$

It is worth noting that this approximation corresponds to a least-square minimisation:

Let us compute $\tilde{F}^m$ such that $\tilde{F}(x) = \sum_{m=1}^{M} \tilde{F}^m e_m(x)$ (10

minimises $J(\tilde{F}) = \int_s (\tilde{F}-F)^2\, dx$

A necessary and sufficient condition is

$$\frac{\partial J}{\partial \tilde{F}^m} = 0 \text{ for } m=1,\ldots,M \tag{11}$$

$$\Leftrightarrow \int_s (F-\tilde{F}) \frac{\partial \tilde{F}^*}{\partial \tilde{F}_m}\, dx = 0 \quad \text{for } m=1, \ldots M \tag{12}$$

and since $\frac{\partial \tilde{F}}{\partial \tilde{F}^m} = e_m(x)$ (13)

as a consequence of (10), this condition is equivalent to

$$\int_s (F-\tilde{F})\, e_m^*(x)\, dx = 0 \qquad (14)$$

Since the e_m are orthonormal

$$(14) \Leftrightarrow \tilde{F}^m = \int_s F\, e_m^*(x)\, dx \qquad (15)$$

which is precisely the definition of F^m (cf (6))

Thus

$$\underline{\underline{\tilde{F} = \bar{F}}} \qquad \text{q.e.d}$$

The next step is to solve equation (7) for $\bar{F}$

$$\frac{\partial \bar{F}}{\partial t} \stackrel{?}{=} A(\bar{F}) \qquad (16)$$

This would correspond to a perfectly valid and possible evolution problem with the initial condition

$$F(x,o) = \bar{F}(x,o) \qquad (17)$$

Unfortunately this is almost always impossible since $\bar{F}$ belongs to $\bar{H}$ (the truncated subspace) and in general

$\underline{A(\bar{F}) \text{ does not}}$

So we are going to replace equation (16) by

$$\frac{\partial \bar{F}}{\partial t} = \overline{A(\bar{F})} \qquad (18)$$

that is we truncate $A(\bar{F})$ so that it belongs to $\bar{H}$ too. (18) is called the $\underline{\text{truncated equation}}$

From (6) it is easy to see that (18) is equivalent to

$$\boxed{\frac{dF^m}{dt} = \left\{A(\bar{F})\right\}^m = \int_s A(\bar{F})\, e_m^*(x)\, dx} \qquad (19)$$

It is now clear that the method is a Galerkin method with the e_m used as test functions. We have thus obtained a set of simple ordinary time differential equations, the right hand sides of which can be evaluated by some quadrature procedures.

As already mentioned, since $A(\bar{F})$ does not belong to $\bar{H}$

$$R(\bar{F}) = \frac{\partial \bar{F}}{\partial t} - A(\bar{F}) \text{ is different from 0.}$$

The equation solved, (18), is such that

$$R(\bar{F}) = \overline{A(\bar{F})} - A(\bar{F}) \tag{20}$$

which is a least squares minimisation of the general residual $R(\bar{F})$ as already shown.

Another trivial consequence of (20) is that, since the e_m are orthogonal, $R(\bar{F})$ belongs to the complement of $\bar{H}$ (i.e. $H - \bar{H}$) and therefore is orthogonal to all elements of $\bar{H}$ and in particular to $\bar{F}$.

In other words:

$$\int_S R(\bar{F})\, \bar{F}\, dx = 0 \tag{21}$$

This property will prove very important for some applications.

Since the set $\{e_m\}$ form an Hilbertian basis, it can be shown that for any element F of H we have the Bessel-Parseval equality:

$$\text{i.e. } \sum_{m \in M} \left| F^m \right|^2 = \| F \|^2 \tag{22}$$

In particular, since $\bar{F}$ belongs to H

$$\sum_{m=1}^{M} \left| F^m \right|^2 = \| \bar{F} \|^2 \tag{23}$$

which means that when $\bar{H}$ converges to H, $\bar{F}$ converges uniformly to F and $R(\bar{F})$

converges uniformly to zero. So if the initial state F(x,o) can be represented by any finite expansion, the truncated solution converges to the exact solution when $\bar{H}$ converges to H.

In conclusion let us mention that the F^m form the "spectrum" of F, hence the name "spectral" method.

2.2 Choice of the expansion functions and basic properties

In practice, since the atmosphere is highly anisotropic, it is often convenient to use different techniques for the horizontal and vertical parts of the equations. Most spectral models use a truncated expansion only in the horizontal, the vertical being treated with classical finite difference techniques. A few however are combined with a finite-element treatment in the vertical.

In this Section we shall therefore restrict ourselves to the application of the spectral technique to the horizontal part of the equations.

Since we are dealing with global or hemispheric problems on the earth, it is desirable to choose expansion functions suited to spherical geometry. It is also further desirable to choose functions simplifying the operator A in equation (7):

$$\frac{\partial F}{\partial t} = A(F).$$

As mentioned earlier, A includes space differential operators (horizontal and vertical derivatives, horizontal Laplacian), non linear terms and possibly some vertical integrals.

Such a simplification is achieved if the expansion functions are eigenfunctions of some sub-operators of A.

This is the case with spherical harmonics

$$Y_n^m(\lambda,\theta) = e^{im\lambda} \; P_n^m(\sin\theta) \tag{24}$$

where λ is longitude and θ is latitude. m is the zonal wavenumber and n is often called the 2-dimensional index or total wavenumber. $P_n^m(\sin\theta)$ is the so called "associated Legendre" function of the first kind, of order m and degree n, a solution of the Legendre equation (see Appendix). There are several analytical forms for the P_n^m. A classical one is given by the Rodrigues formula:

if $\mu = \sin\theta$,

$$P_n^m(\mu) = \sqrt{(2n+1)\,\frac{(n-m)!}{(n+m)!}}\;\frac{(1-\mu^2)^{\frac{|m|}{2}}}{2^n n!}\;x\;\frac{d^{n+|m|}(1-\mu^2)^n}{d\,\mu^{n+|m|}} \tag{25}$$

The Y_n^m are eigenfunctions in particular of the 2-dimensional Laplacian on the sphere:

$$\nabla^2 Y_n^m = \frac{-n(n+1)}{r^2} Y_n^m \tag{26}$$

r = radius of the sphere

This is an obvious consequence of equation (A5) in the Appendix.

Note for the sake of rigour that this holds only in pure spherical geometry (λ,θ,r). If another vertical coordinate (σ for example) is used, it is easy to see that $\nabla^2_\sigma Y_n^m \neq \frac{-n(n+1)}{r^2} Y_n^m$ since r is not constant on a 'horizontal' (σ = constant) surface. It remains however a very good approximation to make, and an approximation consistent with others made in the derivation of the 'primitive' equations usually used in numerical weather prediction and general circulation studies.

The spherical harmonics are also eigenfunctions of the zonal differentiation operator.

$$\frac{\partial Y_n^m}{\partial \lambda} = im\, Y_n^m \tag{27}$$

This is a direct consequence of (24).

Amongst the other important properties of the Legendre functions and spherical harmonics, let us mention

$$\frac{1}{2}\int_{-1}^{1} P_n^m P_{n'}^m \, d\mu = 0 \text{ if } n \neq n' \tag{28}$$

$$= 1 \text{ if } n = n'$$

This is a consequence of the fact that the P_n^m are solutions of the Legendre equation (equation (A.10) of the Appendix) and of the particular normalisation chosen in definition (25). (24) and (28) imply

$$\frac{1}{4\pi}\int_{-1}^{1}\int_{0}^{2\pi} Y_n^m Y_{n'}^{m'*} \, d\lambda \, d\mu = 0 \text{ if } (n,m) \neq (n',m') \tag{29}$$

where Y_n^{m*} represents the complex conjugate of Y_n^m .

Thus in the space of continuous complex functions on the sphere H(S) with the scalar product

$$(f,g) = \frac{1}{4\pi}\int_{1}^{1}\int_{0}^{2\pi} f\, g^* \, d\lambda d\mu \tag{30}$$

the $Y_n^m(\lambda,\theta)$ form an orthogonal set. With this scalar product H(S) is a Hilbert space and it is possible to show that the Y_n^m form an Hilbertian basis of H (see Hobson 1931). Thus every continuous function $F(\lambda,\theta)$ on the sphere can be expressed as

$$F(\lambda,\theta) = \sum_m \sum_n F_n^m Y_n^m(\lambda,\theta) \tag{31}$$

If $F = F(\lambda,\theta,\eta,t)$ where η is some vertical coordinate, then

$$F = \sum_m \sum_n F_n^m(\eta,t)\, Y_n^m(\lambda,\theta) \tag{32}$$

As a consequence of (6) the F_n^m are computed as

$$F_n^m(\eta,t) = \frac{1}{4\pi}\int_{-1}^{1}\int_{0}^{2\pi} F(\lambda,\mu,\eta,t)\, Y_n^{m*}(\lambda,\mu)\, d\lambda\, d\mu \tag{33}$$

The F_n^m are called spectral components.

Note that (33) can be rewritten:

$$F_n^m = \frac{1}{2\pi}\int_0^{2\pi}\left\{\frac{1}{2}\int_{-1}^{1} F\, P_n^m\, d\mu\right\} e^{-im\lambda}\, d\lambda \tag{34}$$

that is as the convolution of a classical Fourier transform

$$\frac{1}{2\pi}\int_0^{2\pi}(\,)\, e^{-im\lambda} d\lambda$$

by a so called Legendre transform

$$\frac{1}{2}\int_{-1}^{1}(\,)\, P_n^m(\mu) d\mu \ .$$

(32) can be separated in a similar way into inverse Fourier and Legendre transforms.

A consequence of the definition of the associated Legendre functions (25) is:

$$P_n^m(\mu) = P_n^{-m}(\mu) \tag{35}$$

and

$$P_n^m(\mu) \equiv 0 \text{ if } |m|>n \tag{36}$$

$$(35) \Rightarrow Y_n^{-m} = Y_n^{*m} \tag{37}$$

and (36) implies that (32) can be written

$$F = \sum_{m=-\infty}^{+\infty}\ \sum_{n=|m|}^{\infty} F_n^m\, Y_n^m \tag{38}$$

Another useful property is

$$P_n^m(-\mu) = (-1)^{n+|m|} P_n^m(\mu) \tag{39}$$

which corresponds to either symmetry or antisymmetry with respect to the equator.

Many recurrence formulae link the various P_n^m. For us, their main usage nowadays is to compute the values of the P_n^m used in the model.

As an example, the one used at ECMWF is

$$P_n^m(\mu) = c_n^m P_{n-2}^{m-2}(\mu) - d_n^m \mu P_{n-1}^{m-2}(\mu) + e_n^m P_{n-1}^m(\mu) \tag{40}$$

with

$$c_n^m = \left(\frac{2n+1}{2n-3}\,\frac{m+n-1}{m+n}\,\frac{m+n-3}{m+n-2}\right)^{\frac{1}{2}}$$

$$d_n^m = \left(\frac{2n+1}{2n-1}\,\frac{m+n-1}{m+n}\,\frac{n-m+1}{m+n-2}\right)^{\frac{1}{2}}$$

$$e_n^m = \left(\frac{2n+1}{2n-1}\,\frac{n-m}{n+m}\right)^{\frac{1}{2}} \quad , \text{ for } m > 0.$$

It has the advantage of being mathematically stable when n and m increase (Belousov, 1962).

Another recurrence formula is used to compute the meridional derivatives (unlike the case for the zonal derivatives, the spherical harmonics are not eigenfunctions of $\frac{\partial}{\partial\mu}$).

$$(\mu^2-1)\,\frac{dP_n^m}{d\mu} = n\,\varepsilon_{n+1}^m P_{n+1}^m - (n+1)\,\varepsilon_n^m P_{n-1}^m \tag{41}$$

with $\varepsilon_n^m = \left(\frac{n^2-m^2}{4n^2-1}\right)^{\frac{1}{2}}$

2.3 The truncation problem

As already mentioned in Sect.2.1, we have to restrict ourselves to a finite number of components. This is called the truncation procedure.

At this point it is important to note that as m and n increase they correspond to features with decreasing horizontal scales. So when using spherical harmonic expansions (Y_n^m) we have a direct control on the scales we wish to neglect, very similar to the one in grid point methods when selecting the size of the grid intervals. We are thus led to the following choice for the truncated field

$$\bar{F} = \sum_{m=-M}^{M} \sum_{n=|m|}^{N(m)} F_n^m \; Y_n^m \tag{42}$$

The fact that m runs from -M to M ensures that $\bar{F}$ is real since (37) and (33) imply

$$F_n^{-m} = F_n^{m*}, \text{and thus that } F_n^{-m} Y_n^{-m} + F_n^m Y_n^m \text{ is real.}$$

A problem is then to know how quickly $\bar{F}$ converges to F.

Let us assume that F is infinitely differentiable. This means that $G = \nabla^{2p} F$ can be expanded in terms of spherical harmonics for any integer p:

$$G = \sum_n \sum_m G_n^m \; Y_n^m \tag{43}$$

with

$$G_n^m = \frac{1}{4\pi} \int_0^{2\pi} \int_{-1}^{1} G \; Y_n^{m*} \, d\mu \, d\lambda \tag{44}$$

Green's theorem states that

$$\frac{1}{4\pi} \int_0^{2\pi} \int_{-1}^{1} g \nabla^2 f \, d\mu \, d\lambda = \frac{1}{4\pi} \int_0^{2\pi} \int_{-1}^{1} f \, \nabla^2 g \, d\mu \, d\lambda \tag{45}$$

Thus repeated use of this theorem shows that

$$G_n^m = \left(\frac{-n(n+1)}{r^2} \right)^p F_n^m \tag{46}$$

The Bessel-Parseval equality applied to G yields

$$\|G\|^2 = \sum_{n,m} \left|G_n^m\right|^2 \tag{47}$$

which implies that $\left|G_n^m\right| \to 0$ when $m,n \to \infty$. Therefore (46) implies

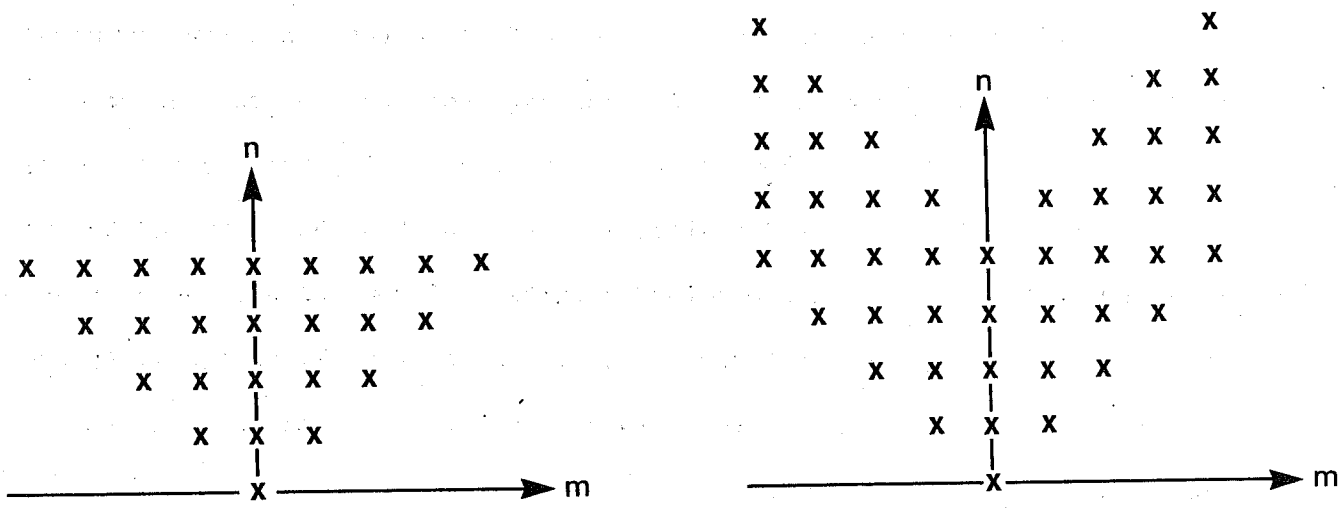
$$F_n^m \, n^{2p} \to 0 \text{ when } n \to \infty \text{ for any } p, \tag{48}$$

which shows that for a smooth field, $\bar{F}$ converges uniformly to F faster than any finite power of $\frac{1}{n}$ where n tends to infinity (Orszag, 1974). This extremely rapid convergence is another important argument in favour of spherical harmonic expansions since it explains why they can achieve an accuracy comparable with grid point methods with substantially fewer degrees of freedom.

If F is not as smooth (as, for example, in the case of the Earth's orography), the convergence can be slower.

The main question left is therefore to chose N(m). The two most common truncations used in numerical prediction models are the so called triangular and rhomboidal ones.

These can be best represented in the (m,n) plane as shown below in Fig.1.



Triangular truncation Fig. 1 Rhomboidal truncation

Early spectral models employed almost exclusively rhomboidal truncations, but the most recently implemented operational spectral models (France, Japan and ECMWF) use triangular truncations.

The reason for choosing a truncation should be to make the best usage of a given number of degrees of freedom (i.e. of a given number of retained spectral components) which means either getting the most accurate results for a given cost or the most economical result for a given accuracy. All this can of course depend on the nature of the problem studied and/or on the computer facilities available.

This being said, the arguments in favour of a rhomboidal truncation are mainly historical. On the technical side, rhomboidal truncations can lead to easier initial programming (the spectral components can be stored in rectangular arrays, i.e. matrices) although with the establishment of the technique and its operational implementation this no longer becomes a valid criterion for choice. It should also be mentioned that several spectral models derived from the same initial rhomboidal code.

A strong reason was put forward by Ellsaesser (1966): he showed that for very coarse resolutions (i.e. keeping only up to five to ten zonal wave numbers) rhomboidal truncations could maximize the variance retained for the rotational kinetic energy at 500 mb. However Baer (1972) analysing the kinetic energy for January and February 1969 at five levels (700, 500, 300, 250 and 200 mb) over the northern hemisphere obtained results in favour of a triangular shaped truncation for moderate resolutions (Fig.2) More recently Savijärvi (personal communication) has obtained results supporting even more strongly a similar conclusion.

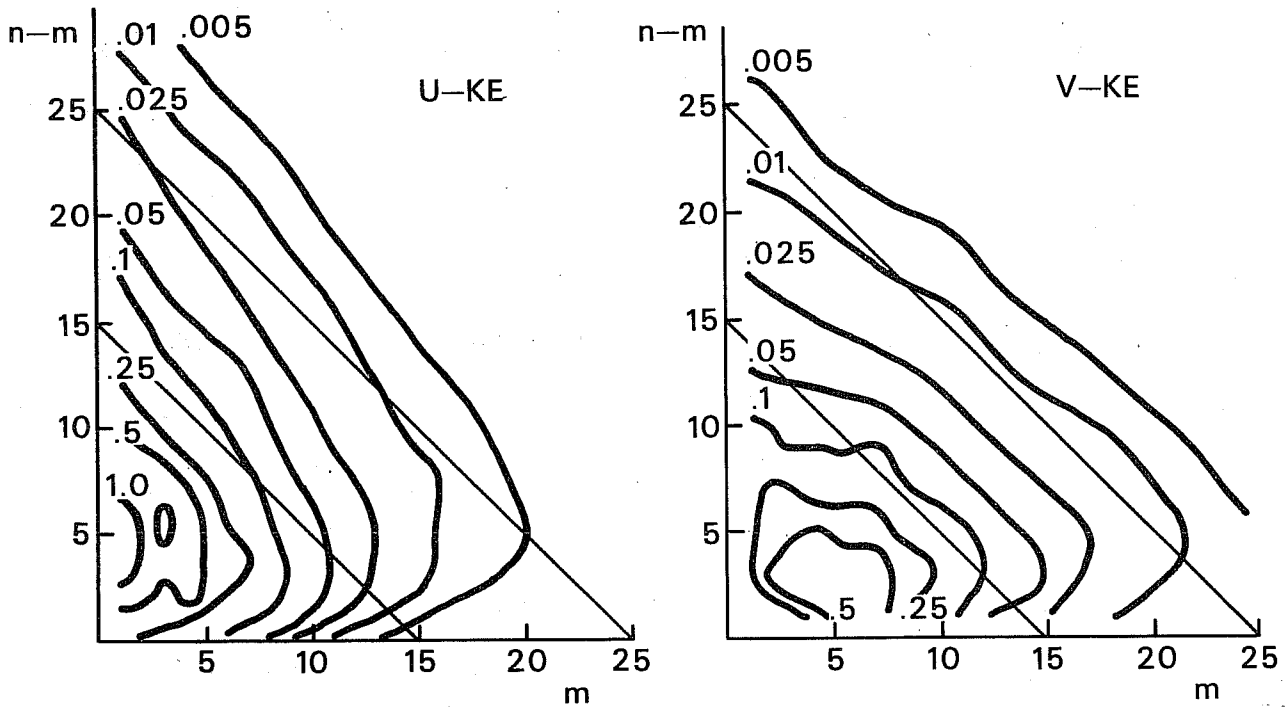


Fig. 2 Percentage of total KE in each spectral component (after Baer, 1972).

A theoretical advantage of a triangular truncation is that it is isotropic. In other words, if $\bar{F}_1$ is an expansion of F triangularly truncated at wavenumber M in one spherical coordinate system (λ_1,θ_1) and $\bar{F}_2$ in another one (λ_2, θ_2) then $\bar{F}_1 = \bar{F}_2$.

This results from an interesting property of the spherical harmonics (see for example Courant and Hilbert 1953):

$$Y_n^m(\lambda_1,\theta_1) = \sum_{m'=-n}^{n} C_n^{m,m'} Y_n^{m'}(\lambda_2,\theta_2) \qquad (49)$$

The resolution in a triangularly truncated system is therefore uniform on the sphere. In other words, whatever the position of the pole the truncated field is the same.

Rhomboidal truncations on the other hand are only invariant under rotations around the earth's axis and correspond to a uniform resolution only in the east-west direction.

Incidentally, (49) also explains why n is called the two-dimensional index.

n defines a scale which is invariant by any rotation of the (λ,θ) coordinate system and is thus truly two dimensional (m on the other hand only defines an east-west scale at a <u>given</u> latitude).

Therefore when using a triangular truncation, all components corresponding to a given spatial scale are either kept or discarded in a consistent way. This is also consistent with the isotropic horizontal diffusion (∇^{2p} form) used in many numerical prediction models.

The anisotropy of the rhomboidal truncation has been mentioned by some authors as an advantage, claiming that for a given number of degrees of freedom, a rhomboidal truncation provides an increase north south resolution, at the expense of the east west one. This would of course be an advantage in resolving the strong north south gradients near the jet for example. This is however not true as we can easily show.

Let us assume that we have a perfectly zonal flow (see Fig. 3).

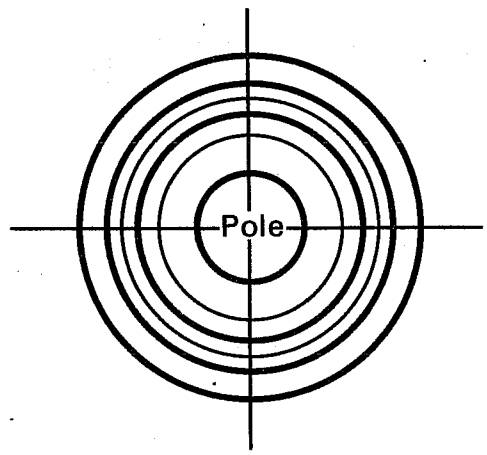


Fig. 3

It can be analysed in terms of $Y_n^o(\mu)$ only. Since it does not depend on λ, m has to be zero. Note that definition (25) implies

$$Y_n^o(\mu) = P_m^o(\mu) = \frac{\sqrt{2n+1}}{2^n\, n!} \frac{d^n(1-\mu^2)^n}{d\,\mu^n} \tag{50}$$

The $P_n^o(\mu)$ are called Legendre polynomials and the $Y_n^o(\mu)$ are called zonal harmonics.

If we now compare, as in Fig.4 a rhomboidal and a triangular truncation with the same number of degrees of freedom (i.e. of spectral components retained) one can distinguish four different parts.

There is a part A common to both truncations. There are two parts (B and D) kept by the triangular truncation and not by the rhomboidal and one part (C) for which the reverse is true.

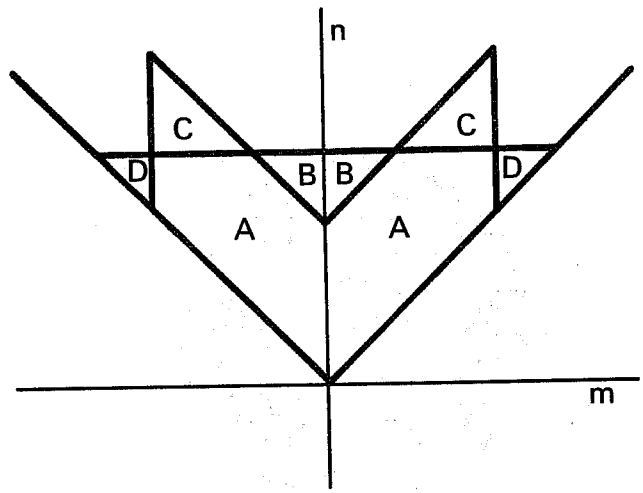


Fig. 4

It is clear that the triangular truncation will better represent the north south gradient of the type shown in Fig.3 since it keeps more zonal harmonics (m=0) than the rhomboidal one. In the real atmosphere, the flow is never purely zonal, but it has a strong zonal component together with a substantial amplitude in the planetary waves (m=1-3), and the north south gradients associated with them will be more accurately represented by the triangular truncation due to part B. This may prove more important in the winter stratosphere when the jet is strong and there is a large amplitude in the

low zonal wave numbers. The impact of part D is easy to explain: it reduces the east west resolution for the rhomboidal truncation at low latitudes. (At high latitudes the $P_n^m(\mu)$ for high m have extremely small values). The importance of part C is much more difficult to evaluate, precisely because of its anisotropic nature. From what has been said earlier, there is no doubt that it introduces smaller scales than in the corresponding triangular truncation. An important point is that at high latitudes (μ tending to 1) we have the following inequality.

$$\left|P_{n_1}^m(\mu)\right| < \left|P_{n_2}^m(\mu)\right| \text{ if } n_1 < n_2.$$

Thus for a given zonal wavenumber m, additional components with larger two dimensional index n will contribute to increase resolution at high latitudes, but not uniformly, and paradoxically not only in the meridional direction. Therefore part C brings increased resolution at high latitudes in the medium part of the spectrum.

From a computational point of view the triangular truncation is slightly more efficient than the rhomboidal one for the same number of degrees of freedom. Firstly, since shorter waves are represented in the rhomboidal truncation, a slightly smaller time step has to be used. Secondly the associated grid (Gaussian grid, - see Sect.3) where all the non linear terms (including the physics) are computed is larger by about 20%. These two points lead to an overall increase in cost of about 25%.

Practical comparisons on the other hand have been carried out by a few groups. Simmons and Hoskins in 1975 compared the two truncations for a growing baroclinic wave. No significant difference was found with the higher versions of these truncations (triangular T42 or the equivalent rhomboidal one). For

the lower resolution (triangular T21 versus rhomboidal R16) the rhomboidal showed some slight advantages. It should however be noted that there was a maximum in the amplitude of the vorticity field for zonal wavenumber 8, and strong non linear generation of wavenumber 16. There was thus significant activity precisely in the spectral range where this particular rhomboidal truncation can be expected a priori to improve according to our discussion above. Therefore the problem appears biased towards the rhomboidal low resolution truncation.

Daley and Bourassa (1978) performed a number of short term (36 h) comparisons between triangular T29 and rhomboidal R20 multilevel spectral models. The models proved extremely similar for the mid and high latitudes. Similar conclusions were obtained from another comparison with higher resolutions (triangular T40 versus rhomboidal R29).

Some previously unpublished experiments have also been carried out at ECMWF comparing a triangular T40 with a rhomboidal R28 truncation in a series of six 10 day forecasts from initial conditions chosen from dates within February, 1976.

The results did not contradict those of Daley and Bourassa. However, at a range beyond two days they showed a distinct advantage for the triangular truncation. This advantage increased with height (see Fig.5) to become more substantial in the higher troposphere, which is consistent with the previous discussion.

The triangular truncation was subsequently selected as the basic truncation for the ECMWF spectral model on the basis of these arguments and results.

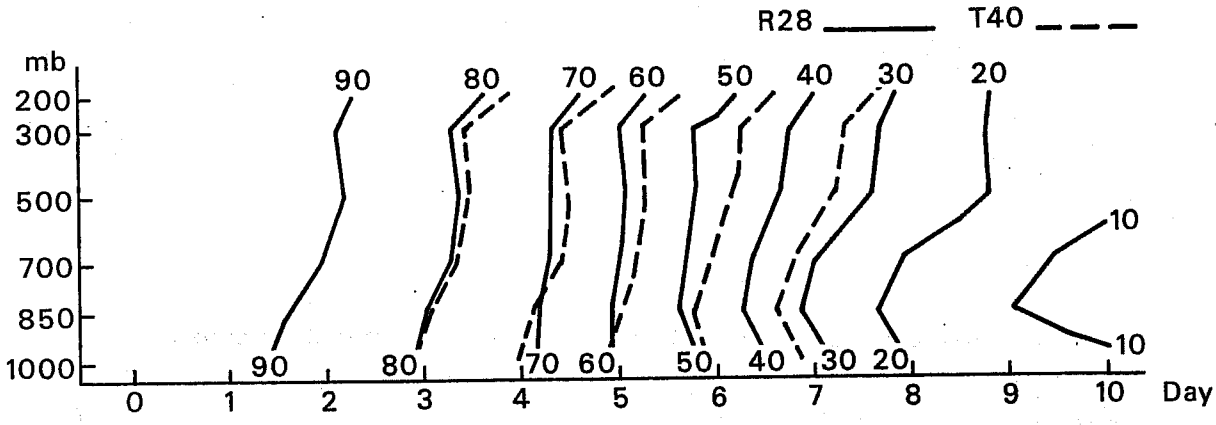


Fig. 5 Anomaly correlation (82.5-20°N) as a function of height.

3. IMPLEMENTATION IN ATMOSPHERIC MODELS

3.1 The barotropic vorticity equation

3.1.1 Governing equations

We first illustrate the way the spectral method is implemented in atmospheric models with spherical geometry by considering the simplest case; that of the barotropic vorticity equation. The governing prognostic equation is

$$\frac{\partial \xi}{\partial t} = - \frac{U}{a(1-\mu^2)} \frac{\partial \zeta}{\partial \lambda} - \frac{V}{a} \frac{\partial \zeta}{\partial \mu} \qquad (51)$$

where ξ is the relative vorticity, ζ the absolute vorticity, U and V the eastward and northward velocities multiplied by the cosine of latitude, and a the radius of the earth. As before, t is time, λ is longitude and μ is the sine of latitude. The relative vorticity is defined in terms of U and V by

$$\xi = \frac{1}{a} \left\{ \frac{1}{1-\mu^2} \frac{\partial V}{\partial \lambda} - \frac{\partial U}{\partial \mu} \right\} \qquad (52)$$

and the absolute vorticity is given by

$$\zeta = 2\Omega\mu + \xi \tag{53}$$

where Ω is the earth's rotation rate. Velocities are non-divergent enabling us to introduce a stream function ψ (to within an additive constant) such that

$$U = -\frac{1}{a}(1-\mu^2)\frac{\partial\psi}{\partial\mu} \tag{54}$$

$$V = \frac{1}{a}\frac{\partial\psi}{\partial\lambda} \tag{55}$$

whence

$$\xi = \nabla^2\psi \tag{56}$$

The non-divergence of winds also allows (51) to be rewritten thus:

$$\frac{\partial\xi}{\partial t} = -\frac{1}{a(1-\mu^2)}\frac{\partial}{\partial\lambda}(U\zeta) - \frac{1}{a}\frac{\partial}{\partial\mu}(V\zeta) \tag{57}$$

3.1.2 Spectral representation

Choosing spherical harmonic expansion functions as discussed in Sect.2.2, we represent ξ by a truncated series:

$$\xi = \sum_{m=-M}^{M}\sum_{n=|m|}^{N(m)} \xi_n^m P_n^m(\mu)\, e^{im\lambda} \tag{58}$$

Since the spherical harmonics are eigenfunctions of the operator ∇^2, we may use (26) and (56) to write

$$\psi = -a^2 \sum_{m=-M}^{M}\sum_{\substack{n=|m| \\ n\neq o}}^{N(m)} \frac{\xi_n^m}{n(n+1)} P_n^m(\mu)\, e^{im\lambda} \tag{59}$$

where the restriction $n\neq o$ corresponds to the choice $\psi_o^o = 0$.

From (59), using (54) and (55),

$$U = a \sum_{m=-M}^{M}\sum_{\substack{n=|m| \\ n\neq o}}^{N(m)} \frac{\xi_n^m}{n(n+1)} \left\{(1-\mu^2)\frac{dP_n^m}{d\mu}\right\} e^{im\lambda} \tag{60}$$

and

$$V = -a \sum_{m=-M}^{M}\sum_{\substack{n=|m| \\ n\neq o}}^{N(m)} \frac{im\xi_n^m}{n(n+1)} P_n^m(\mu)\, e^{im\lambda} \tag{61}$$

The truncated spectral form of equation (57) follows from the general case represented by (19):

$$\frac{\partial \xi_n^m}{\partial t} = -\frac{1}{4\pi a} \int_{-1}^{1} \int_{0}^{2\pi} \left\{ \frac{1}{1-\mu^2} \frac{\partial}{\partial \lambda} (U\zeta) + \frac{\partial}{\partial \mu} (V\zeta) \right\} P_n^m(\mu)\, e^{-im\lambda}\, d\lambda d\mu \qquad (62)$$

and using integration by parts, this equation may be written :

$$\frac{\partial \xi_n^m}{\partial t} = -\frac{1}{4\pi a} \int_{-1}^{1} \int_{0}^{2\pi} \left\{ \frac{im(U\zeta)P_n^m(\mu)}{1-\mu^2} - (V\zeta) \frac{dP_n^m}{d\mu} \right\} e^{-im\lambda}\, d\lambda d\mu \qquad (63)$$

From (53), (58), (60) and (61), the integral in (63) is known in terms of the ξ_n^m and the known analytical functions P_n^m, $dP_n^m/d\mu$ and $e^{im\lambda}$. Thus given an initial distribution of vorticity, its subsequent evolution may be computed using an appropriate time discretization once a computationally feasible technique for evaluating the integral can be found. As discussed in the introduction, the discovery of efficient "transform" techniques for evaluating such integrals has been central to the development of spectral models as practical meteorological tools.

3.1.3 Interaction coefficients

Perhaps the most obvious approach to the evaluation of integrals of form (63), and the approach adopted in the early spectral models referred to previously, is to insert directly the spectral expansions into the integrand. Consider the simple product

$$C = AB \qquad (64)$$

where A and B are represented by truncated spectral expansions:

$$A = \sum_{m=-M}^{M} \sum_{N|m|}^{N(m)} A_n^m\, Y_n^m \qquad (65)$$

$$B = \sum_{m=-M}^{M} \sum_{n=|m|}^{N(m)} B_n^m\, Y_n^m \qquad (66)$$

The spectral coefficients of C are given by

$$C_n^m = \frac{1}{4\pi} \int_{-1}^{1} \int_{o}^{2\pi} AB \, (Y_n^m)^* \, d\lambda \, d\mu \qquad (67)$$

and directly substituting the expansions (65) and (66) into (67) gives

$$C_n^m = \sum_{p=-M}^{M} \sum_{q=|p|}^{N(p)} \sum_{r=-M}^{M} \sum_{s=|r|}^{M} I_{qsn}^{prm} \, A_q^p \, B_s^r \qquad (68)$$

where the I_{qsn}^{prm} are known as "interaction coefficients", and are given by

$$I_{qsn}^{prm} = \frac{1}{4\pi} \int_{-1}^{1} \int_{o}^{2\pi} Y_q^p \, Y_s^r \, (Y_n^m)^* \, d\lambda \, d\mu \qquad (69)$$

Such interaction coefficients may be pre-computed and stored, and repeatedly used during the integration in time of a model such as governed by Eqn.(63). This approach, however, is quite impractical for realistic weather prediction or climate models. Even allowing for the fact that many of the interaction coefficients are zero, the computation required per timestep increases markedly as resolution is increased, with $0(N^5)$ operations/timestep for large N, with M~N (Orszag, 1970). Computer storage requirements also become prohibitive as resolution increases, and there is no feasible way that highly non-linear parameterisations can be included. It was thus not until an alternative approach was developed by Eliasen et al. (1970) and Orszag (1970) that spectral models became competitive with the long-established finite-difference models.

3.1.4 The spectral transform method

The alternative approach is based on evaluation of grid point values of basic and derived variables (or "transforms" from spectral to grid-point space) and use of numerical quadrature to evaluate integrals such as (63). Considering again the simple product C = AB, grid-point values of A (and B) are computed in a two-stage process.

(i) Inverse Legendre Transformation

$$A_m(\mu_j) = \sum_{n=|m|}^{N(m)} A_n^m P_n^m(\mu_j) \tag{70}$$

(ii) Inverse Fourier Transformation

$$A(\lambda_i, \mu_j) = \sum_{m=-M}^{M} A_m(\mu_j)\, e^{im\lambda_i} \tag{71}$$

With values $B(\lambda_i, \mu_j)$ similarly calculated, grid-point values of $C(\lambda_i, \mu_j)$ are computed by simple multiplication.

These values of C are then used in quadrature formulae to evaluate longitudinal and latitudinal integrals giving the transformation from grid-point back to spectral space. In continuous form, these transformations are:

(i) Fourier Transformation

$$C_m(\mu) = \frac{1}{2\pi} \int_0^{2\pi} C\, e^{-im\lambda}\, d\lambda \tag{72}$$

(ii) Legendre Transformation

$$C_n^m = \frac{1}{2} \int_{-1}^{1} C_m(\mu)\, P_n^m(\mu)\, d\mu \tag{73}$$

The methods used to evaluate the integrals (72) and (73) are generally such as to give exact (or non-aliased) results for the contribution of quadratic non linear terms to the computation of those spectral coefficients retained in the truncation. Considering first the longitudinal integral, the integrand in (72) is the product of three trigonometric functions, and includes zonal wavenumbers up to 3M. Following Machenhauer and Rasmussen (1972), the integral is evaluated using a regular grid of points in the east/west, with

$$\lambda_i = \frac{2\pi i}{I} \tag{74}$$

and

$$C_m(\mu_j) = \frac{1}{I} \sum_{i=1}^{I} C(\lambda_i, \mu_j) e^{-im\lambda_i} \qquad (75)$$

The result is exact for

$$I \geq 3M+1 \qquad (76)$$

In practice, I is chosen to allow efficient "fast Fourier transform" techniques (Cooley and Tukey, 1965, Temperton, 1983) to be used to compute the sums (71) and (75), and truncation are usually such that (3M+1) is equal or just less than one such I, for example an integer of the form $2^p\ 3^q\ 5^r$ for integers p, q and r. Thus spectral models have typically been run, or are anticipated to run, with zonal truncations given by M = 15, 21, 31, 42, 63, 84, 95, etc.

Considering now the latitudinal integral, we note that

(i) the associated Legendre functions $P_n^m(\mu)$, as given by (25), are of the form

$$(1 - \mu^2)^{|m|/2} \times (\text{Polynomial in } \mu \text{ of degree } (n-|m|))$$

(ii) the longitudinal integral of a product of three spherical harmonics vanishes unless the three zonal wavenumbers, m_1, m_2 and m_3, of the harmonics satisfy

$$m_1 + m_2 + m_3 = 0$$

It follows from these properties that the integrand $c_m(\mu)\ P_n^m(\mu)$ in (73) is a polynomial in μ, of degree at most

3M for triangular truncation (N=M) ,

2M+3K for rhomboidal truncation (N=m+K),

Following Eliassen et al (1970), the integral (73) is evaluated by Gaussian quadrature:

$$C_n^m = \tfrac{1}{2} \sum_{j=1}^{J} g_j\ C_m(\mu_j)\ P_n^m(\mu_j) \qquad (77)$$

where the μ_j are the zeros of $P_J^o(\mu)$, that is

$$P_J^o(\mu_j) = 0 \qquad , \quad j = 1, 2 \ldots J, \qquad (78)$$

and

$$g_j = \frac{2(1 - \mu_j^2)}{(J\ P^0_{J-1}(\mu_j))^2} \qquad (79)$$

This Gaussian quadrature is exact for an integrand which is a polynomial in μ of degree $\leq 2J-1$, and in view of the particular form of $C_m(\mu)P^m_n(\mu)$, J must satisfy

$$\left.\begin{aligned} &J \geq \frac{3M+1}{2} \text{ for triangular truncation,} \\ \text{and} \\ &J \geq \frac{2M+3K+1}{2} \text{ for rhomboidal truncation.} \end{aligned}\right\} \qquad (80)$$

These conditions apply to the simple product (64). Returning to the particular form of the barotropic vorticity equation, it may be shown from (63) that the minimum number of "Gaussian" latitudes becomes 3M/2 or (2M+3K)/2 in this case, although for the primitive equations considered in Section 3.2, it is conditions (80) that apply for non-aliased treatment of all quadratic terms.

3.1.5 The computational sequence

The sequence of calculations for a barotropic vorticity equation model using the spectral transform technique may now be summarized.

At a particular forecast time, t, we start from the spectral coefficients of relative vorticity $\xi^m_n(t)$, and depending on the time-scheme we may also have available values at previous times, particularly $\xi^m_n(t-\Delta t)$ say. The bulk of the model computation involves a sequence of calculations which is repeated for each Gaussian latitude $\sin^{-1}\mu_j$. These calculations comprise

(i) Inverse Legendre Transforms

Compute

$$\xi_m(\mu_j) = \sum_{n=|m|}^{N(m)} \xi^m_n\, P^m_n(\mu_j)$$

and also, using (60) and (61),

$$U_m(\mu_j) = a \sum_{\substack{n=|m| \\ n\neq 0}}^{N(m)} \frac{\xi_n^m}{n(n+1)} (1-\mu_j^2) \frac{dP_n^m(\mu_j)}{d\mu}$$

and

$$V_m(\mu_j) = -ima \sum_{\substack{n=|m| \\ n\neq 0}}^{N(m)} \frac{\xi_n^m}{n(n+1)} P_n^m(\mu_j)$$

(ii) Inverse Fourier Transforms

Compute grid point values $\xi(\lambda_i, \mu_j)$, $U(\lambda_i, \mu_j)$, $V(\lambda_i, \mu_j)$.

(iii) Grid-point calculations

Compute $\zeta(\lambda_i, \mu_j)$ from $\xi(\lambda_i, \mu_j)$ using (53), and form the products $U\zeta$ and $V\zeta$.

(iv) Fourier Transforms

Compute $(U\zeta)_m$ and $(V\zeta)_m$ as in Eqn.(75).

(v) Legendre Transforms

Compute the contribution from the current latitude, μ_j, to $\partial\xi_n^m/\partial t$. From (63) and (77) this contribution is

$$-\frac{g_j}{2a}\left\{(U\zeta)_m \frac{imP_n^m(\mu_j)}{1-\mu_j^2} - (V\zeta)_m \frac{dP_n^m}{d\mu}(\mu_j)\right\}$$

Contributions are accumulated latitude by latitude.

Once these computations are complete, $\partial\xi_n^m/\partial t$ is known, and new values of ξ_n^m may be computed, for example using the simple centred scheme

$$\xi_n^m(t+\Delta t) = \xi_n^m(t-\Delta t) + 2\Delta t \frac{\partial \xi_n^m}{\partial t} \qquad (81)$$

The sequence of calculations may then be repeated.

3.1.6 Additional comments

(i) The computational grid

The grid on which non-linear calculations are performed is regular in the east/west direction, but formally irregular in the north/south. In practice, however, Gaussian latitudes (determined by (78)) are almost regular in θ, particularly as resolution increases. For example, with J=96 (as in the operational model at ECMWF, which uses triangular truncation at wavenumber 63 (T63)), the Gaussian latitudes all lie within 1.4 km of a regular grid starting at 88.585°N with a grid interval of 1.8649°.

(ii) Computational cost

The transform method requires $0(N^3)$ operations/timestep for large N, with $M \sim N$ (Eliassen et al. 1970, Orszag, 1970). It thus has a substantial advantage over the interaction coefficient method in the asymptotic limit, and this advantage is realised at quite low resolutions. For example, Bourke (1972) showed that for a shallow-water equation model rhomboidally truncated with $M = K = 15$, the transform method was faster by a factor of 10.

The $0(N^3)$ asymptotic operation count arises because of the computation of the Legendre transforms; the fast Fourier transforms yield an $0(N^2 \log N)$ behaviour, and the non-linear grid point calculations involve $0(N^2)$ operations. In practice, however, models are far from the asymptotic limit. In the multi-level T63 model used at ECMWF, the Legendre transforms account for about 20% of the computational cost, and the Fourier transforms about 5%. Extrapolating leads to figures of 33% for the Legendre transforms at a resolution of T126 and 50% for T252, while Fourier transforms remain at close to 5% of the net cost, assuming no corresponding increase in the complexity of the physical parameterisations. Thus although the cost of the Legendre

transforms is not prohibitive for current or immediate future resolutions, it is by no means negligible, and the search for more efficient transform techniques continues (Orszag, 1979).

(iii) Optimisation

A property of the Legendre functions which may be derived directly from the definition (25) is

$$P_n^m(\mu) \sim (1-\mu^2)^{|m|/2} \quad \text{as } \mu \to \pm 1.$$

Thus for large m the functions become vanishingly small as the poles are approached, and the contributions to spectral tendencies from polar regions become less than unavoidable round-off error for sufficiently large zonal wavenumbers. This means that in practice 3M+1 longitudinal points may not be needed at all latitudes, and that a decreasing number of points may be used as the poles are approached without significant loss of accuracy (Machenhauer, 1979). This has been confirmed in some preliminary experiments at ECMWF, but exhaustive testing has yet to be carried out.

It is unclear to what extent there is further scope for optimisation by use of grids which are generally coarser than required for alias-free calculation of quadratic terms. Unsatisfactory results have been reported by Bourke et al. (1977) and Machenhauer (1979) using quite low resolutions, although successful integration of a higher-resolution bi-Fourier model with a very selective horizontal diffusion has been reported by Sadourny (personal communication).

3.2 The primitive equations

3.2.1 Basic variables and spectral expansions

We illustrate the application of the spectral transform method to multi-level

primitive-equation models for the choice of prognostic variables made for the operational ECMWF model. These variables are:

relative vorticity - ξ

divergence - D

temperature - T

specific humidity - q

natural logarithm of surface pressure - $\ell n(p_s)$

Each is expanded in truncated series of spherical harmonic functions. For example,

$$\xi(\lambda,\mu,\eta,t) = \sum_{m=-M}^{M} \sum_{m=|m|}^{N(m)} \xi_n^m(\eta,t)\, P_n^m(\mu)\, e^{im\lambda} \tag{82}$$

Here η is the vertical coordinate, which we shall take to be a general pressure- based terrain-following coordinate. It must be a monotonic function of pressure (p) and depend also on surface pressure:

$$\eta = \eta(p\ ,\ p_s) \tag{83}$$

where

$$\eta(0,p_s) = 0 \text{ and } \eta(p_s,p_s) = 1 \tag{84}$$

We shall not be concerned here with the discretization of $\xi_n^m(\eta,t)$ and other spectral coefficients in the vertical, but shall simply assume that all relevant integrals and derivatives with respect to η are calculable using grid-point values of the basic variables.

To compute wind components we introduce the stream function, ψ and velocity potential, χ. Then

$$U = \frac{1}{a}\left\{-(1-\mu^2)\frac{\partial\psi}{\partial\mu} + \frac{\partial\chi}{\partial\lambda}\right\} \tag{85}$$

and

$$V = \frac{1}{a}\left\{\frac{\partial\psi}{\partial\lambda} + (1-\mu^2)\frac{\partial\chi}{\partial\mu}\right\} \tag{86}$$

with

$$\xi = \nabla^2\psi \tag{87}$$

and $$D = \nabla^2\chi \tag{88}$$

Truncated expansions for ψ and χ are thus given by

$$\psi = -a^2 \sum_{m=-M}^{M} \sum_{\substack{n=|m| \\ n\neq 0}}^{N(m)} \frac{\xi_n^m}{n(n+1)} P_n^m(\mu) e^{im\lambda} \tag{89}$$

and $$\chi = -a^2 \sum_{m=-M}^{M} \sum_{\substack{n=|m| \\ n\neq 0}}^{N(m)} \frac{D_n^m}{n(n+1)} P_n^m(\mu) e^{im\lambda} \tag{90}$$

(85), (86), (89) and (90) thus give truncated expansions for U and V:

$$U = a \sum_{m=-M}^{M} \sum_{\substack{n=|m| \\ n\neq 0}}^{N(m)} \frac{1}{n(n+1)} \left\{ \xi_n^m (1-\mu^2) \frac{dP_n^m}{d\mu} - im D_n^m P_n^m \right\} e^{im\lambda} \tag{91}$$

and $$V = -a \sum_{m=-M}^{M} \sum_{\substack{n=|m| \\ n\neq 0}}^{N(m)} \frac{1}{n(n+1)} \left\{ im\xi_n^m P_n^m + D_n^m (1-\mu^2) \frac{dP_n^m}{d\mu} \right\} e^{im\lambda} \tag{92}$$

3.2.2 Governing equations

The governing primitive equations for U, V, T and q may be written in the form

$$\frac{\partial U}{\partial t} - \zeta V + \dot{\eta} \frac{\partial U}{\partial \eta} + \frac{R_d T_v}{a} \frac{\partial \ln p}{\partial \lambda} + \frac{1}{a} \frac{\partial}{\partial \lambda} (\phi + E) = P_U \tag{93}$$

$$\frac{\partial V}{\partial t} + \zeta U + \dot{\eta} \frac{\partial V}{\partial \eta} + \frac{R_d T_v}{a} (1-\mu^2) \frac{\partial \ln p}{\partial \mu} + \frac{(1-\mu^2)}{a} \frac{\partial}{\partial \mu} (\phi + E) = P_V \tag{94}$$

$$\frac{\partial T}{\partial t} + \frac{U}{a(1-\mu^2)} \frac{\partial T}{\partial \lambda} + \frac{V}{a} \frac{\partial T}{\partial \mu} + \dot{\eta} \frac{\partial T}{\partial \eta} - \frac{\kappa T_v \omega}{(1+(\delta-1)q)p} = P_T \tag{95}$$

and

$$\frac{\partial q}{\partial t} + \frac{U}{a(1-\mu^2)} \frac{\partial q}{\partial \lambda} + \frac{V}{a} \frac{\partial q}{\partial \mu} + \dot{\eta} \frac{\partial q}{\partial \eta} = P_q \tag{96}$$

The continuity equation is

$$\frac{\partial}{\partial \eta}\left(\frac{\partial p}{\partial t}\right) + \tilde{D}\frac{\partial p}{\partial \eta} + \frac{\partial}{\partial \eta}\left(\dot{\eta}\frac{\partial p}{\partial \eta}\right) = 0 \tag{97}$$

where $$\tilde{D} = D + \frac{U}{a(1-\mu^2)}\frac{\partial}{\partial \lambda}\left(\ln \frac{\partial p}{\partial \eta}\right) + \frac{V}{a}\frac{\partial}{\partial \mu}\left(\ln \frac{\partial p}{\partial \eta}\right) \tag{98}$$

The hydrostatic equation gives

$$\phi = \phi_s + \int_{\eta}^{1} \frac{R_d T_v}{p}\frac{\partial p}{\partial \eta}\, d\eta \tag{99}$$

and the quantity E in (93) and (94) is given by

$$E = \frac{1}{2}(U^2 + V^2)/(1 - \mu^2) \tag{100}$$

In these equations R_d is the gas constant for dry air and $\kappa = R_d/C_{pd}$, with C_{pd} the specific heat of dry air at constant pressure. P_U, P_V, P_T and P_q represent tendencies resulting from the parameterized processes, which are not specified here. T_v is the virtual temperture, given by

$$T_v = T\left(1 + \left(\frac{1}{\varepsilon} - 1\right) q\right) \tag{101}$$

where ε is the ratio of the gas constants of dry air and water vapour, while δ in (95) is the ratio of the specific heats at constant pressure of water vapour and dry air.

The pressure-coordinate vertical velocity, ω, is given by

$$\omega = -\int_{o}^{\eta} \tilde{D}\frac{\partial p}{\partial \eta}\, d\eta + \frac{U}{a(1-\mu^2)}\frac{\partial p}{\partial \lambda} + \frac{V}{a}\frac{\partial p}{\partial \mu} \tag{102}$$

and explicit expressions for the rate of change of surface pressure and for $\dot{\eta}$ are obtained by integrating (97) using the boundary conditions $\dot{\eta}=0$ at $\eta=0$ and $\eta=1$:

$$\frac{\partial p_s}{\partial t} = -\int_{o}^{1} \tilde{D}\frac{\partial p}{\partial \eta}\, d\eta \tag{103}$$

and $$\dot{\eta}\frac{\partial p}{\partial \eta} = -\frac{\partial p}{\partial t} - \int_{0}^{\eta} \tilde{D}\frac{\partial p}{\partial \eta}\, d\eta \tag{104}$$

(103) may also be written

$$\frac{\partial \ln p_s}{\partial t} = -\frac{1}{p_s}\int_0^1 \tilde{D}\,\frac{\partial p}{\partial \eta}\,d\eta \tag{105}$$

In the special case of sigma coordinates,

$$\eta \equiv \sigma = p/p_s \tag{106}$$

and

$$\frac{\partial p}{\partial \eta} = p_s \tag{107}$$

(98), (99), (102), (104) and (105) then reduce to the more familiar forms

$$\tilde{D} = D + \frac{U}{a(1-\mu^2)}\frac{\partial}{\partial \lambda}\ln p_s + \frac{V}{a}\frac{\partial}{\partial \mu}\ln p_s \tag{108}$$

$$\phi = \phi_s + \int_\sigma^1 \frac{R_d T_v}{\sigma}\,d\sigma \tag{109}$$

$$\frac{\omega}{p} = -\frac{1}{\sigma}\int_0^\sigma \tilde{D}\,d\sigma + (\tilde{D} - D) \tag{110}$$

$$\dot{\sigma} = -\sigma\frac{\partial \ln p_s}{\partial t} - \int_0^\sigma \tilde{D}\,d\sigma \tag{111}$$

and

$$\frac{\partial \ln p_s}{\partial t} = -\int_0^1 \tilde{D}\,d\sigma \tag{112}$$

3.2.3 Spectral vorticity and divergence equations

The use of equations in vorticity and divergence form stems from the work of Bourke (1972). The unapproximated equations in physical space are derived from the horizontal momentum equations (93) and (94):

$$\frac{\partial \xi}{\partial t} = \frac{1}{a(1-\mu^2)}\frac{\partial F_v}{\partial \lambda} - \frac{1}{a}\frac{\partial F_u}{\partial \mu} \tag{113}$$

$$\frac{\partial D}{\partial t} = \frac{1}{a(1-\mu^2)}\frac{\partial F_u}{\partial \lambda} + \frac{1}{a}\frac{\partial F_v}{\partial \mu} - \nabla^2 G \tag{114}$$

where

$$F_u = \zeta V - \dot{\eta} \frac{\partial U}{\partial \eta} - \frac{R_d T_v}{a} \frac{\partial \ln p}{\partial \lambda} + P_U \tag{115}$$

$$F_v = -\zeta U - \dot{\eta} \frac{\partial V}{\partial \eta} - \frac{R_d T_v}{a} (1 - \mu^2) \frac{\partial \ln p}{\partial \mu} + P_V \tag{116}$$

and

$$G = \phi + E \tag{117}$$

The truncated spectral forms are then given, as in (19) and (62), by

$$\frac{\partial \xi_n^m}{\partial t} = \frac{1}{4\pi a} \int_{-1}^{1} \int_{0}^{2\pi} \left(\frac{1}{1-\mu^2} \frac{\partial F_v}{\partial \lambda} - \frac{\partial F_u}{\partial \mu} \right) P_n^m(\mu) e^{-im\lambda} \, d\lambda \, d\mu \tag{118}$$

and

$$\frac{\partial D_n^m}{\partial t} = \frac{1}{4\pi a} \int_{-1}^{1} \int_{0}^{2\pi} \left(\frac{1}{1-\mu^2} \frac{\partial F_u}{\partial \lambda} + \frac{\partial F_v}{\partial \mu} \right) P_n^m(\mu) e^{-im\lambda} \, d\lambda \, d\mu$$

$$- \frac{1}{4\pi a} \int_{-1}^{1} \int_{0}^{2\pi} (\nabla^2 G) \, P_n^m(\mu) e^{-im\lambda} \, d\lambda \, d\mu \tag{119}$$

and integrating by parts as for the barotropic vorticity equation we obtain

$$\frac{\partial \xi_n^m}{\partial t} = \frac{1}{4\pi a} \int_{-1}^{1} \int_{0}^{2\pi} \left\{ \frac{im}{1-\mu^2} F_v P_n^m + F_u \frac{dP_n^m}{d\mu} \right\} e^{-im\lambda} \, d\lambda \, d\mu \tag{120}$$

and

$$\frac{\partial D_n^m}{\partial t} = \frac{1}{4\pi a} \int_{-1}^{1} \int_{0}^{2\pi} \left\{ \frac{im}{1-\mu^2} F_u P_n^m - F_v \frac{dP_n^m}{d\mu} \right\} e^{-im\lambda} \, d\lambda \, d\mu$$

$$+ \frac{n(n+1)}{4\pi a^2} \int_{-1}^{1} \int_{0}^{2\pi} G \, P_n^m \, e^{-im\lambda} \, d\lambda \, d\mu$$

Here we have again used the property that the spherical harmonics are eigenfunctions of the operator ∇^2.

The computation of tendencies proceeds in the manner discussed in detail for the barotropic vorticity equation. From the spectral representation of basic and derived variables, values of ξ, D, U, V, T, q, $\ln p_s$, $\partial(\ell n p_s)/\partial\lambda$ and $\partial(\ell n p_s)/\partial\mu$ are evaluated at the points of a "Gaussian" grid. Corresponding grid-point values of F_u, F_v and G are computed, using for example finite differences in the vertical, and the integrals (120) and (121) evaluated by numerical quadrature. The choice of Gaussian grid will be discussed further below.

3.2.4 The surface-pressure equation

From (105) we obtain

$$\frac{\partial}{\partial t}(\ln p_s)_n^m = \frac{1}{4\pi}\int_{-1}^{1}\int_{0}^{2\pi} F_p\, P_n^m(\mu)\, e^{-im\lambda}\, d\lambda\, d\mu \qquad (122)$$

where

$$F_p = -\frac{1}{p_s}\int_o^1 \tilde{D}\,\frac{\partial p}{\partial \eta}\, d\eta \qquad (123)$$

Grid point fields required for the computation of tendencies of vorticity and divergence are also sufficient for the calculation of F_p. Tendencies of the spectral coefficients of $\ln p_s$, as given by (122), are then evaluated by the same quadrature as for other tendencies.

3.2.5 The temperature and humidity equations

Although the treatment of the temperature and humidity equations is generally similar to that of the other prognostic equations, two different approaches have been adopted for the actual implementation of the technique. For adiabatic models they yield identical results, apart from round-off error, but the choice of approach has implications for the way the parameterizations of convection and large-scale precipitation are included in a model.

To illustrate the two approaches we write equations (95) and (96) in the forms

$$\frac{\partial T}{\partial t} = - \frac{U}{a(1-\mu^2)} \frac{\partial T}{\partial \lambda} - \frac{V}{a} \frac{\partial T}{\partial \mu} + R_T \tag{124}$$

and

$$\frac{\partial q}{\partial t} = - \frac{U}{a(1-\mu^2)} \frac{\partial q}{\partial \lambda} - \frac{V}{a} \frac{\partial q}{\partial \mu} + R_q \tag{125}$$

where values of R_T and R_q can be computed directly on the Gaussian grid using those grid-point fields used for the calculation of tendencies of vorticity and divergence.

In the first, and more traditional approach, (124) and (125) are written

$$\frac{\partial T}{\partial t} = - \frac{1}{a(1-\mu^2)} \frac{\partial}{\partial \lambda} (UT) - \frac{1}{a} \frac{\partial}{\partial \mu} (VT) + (DT + R_T) \tag{126}$$

and

$$\frac{\partial q}{\partial t} = - \frac{1}{a(1-\mu^2)} \frac{\partial}{\partial \lambda} (Uq) - \frac{1}{a} \frac{\partial}{\partial \mu} (Vq) + (Dq + R_q) \tag{127}$$

Values of UT, VT, DT, Uq, Vq and Dq, as well as R_T and R_q are evaluated on the Gaussian grid, and integration by parts is used to write the spectral equations in a form suitable for the application of the quadrature formulae:

$$\frac{\partial T_n^m}{\partial t} = - \frac{1}{4\pi a} \int_{-1}^{1} \int_{0}^{2\pi} \left\{ \frac{im}{1-\mu^2} (UT) P_n^m - (VT) \frac{dP_n^m}{d\mu} \right\} e^{-im\lambda} \, d\lambda \, d\mu$$

$$+ \frac{1}{4\pi a} \int_{-1}^{1} \int_{0}^{2\pi} (DT + R_T) P_n^m e^{-im\lambda} \, d\lambda \, d\mu \tag{128}$$

and

$$\frac{\partial q_n^m}{\partial t} = - \frac{1}{4\pi a} \int_{-1}^{1} \int_{0}^{2\pi} \left\{ \frac{im}{1-\mu^2} (Uq) P_n^m - (Vq) \frac{dP_n^m}{d\mu} \right\} e^{-im\lambda} \, d\lambda \, d\mu$$

$$+ \frac{1}{4\pi a} \int_{-1}^{1} \int_{0}^{2\pi} (Dq + R_q) P_n^m e^{-im\lambda} \, d\lambda \, d\mu \tag{129}$$

In this approach, tendencies of T and q are not directly available on the Gaussian grid, and parameterizations involving condensation have to be performed after transformations to and from spectral space, for example at the beginning of the next timestep as in the original ECMWF spectral model (Baede et al. 1979).

In the second approach, the additional fields

$$\frac{\partial T}{\partial \lambda}\ ,\ \frac{\partial T}{\partial \mu}\ ,\ \frac{\partial q}{\partial \lambda} \text{ and } \frac{\partial q}{\partial \mu}$$

are evaluated at grid-points from the spectral expansions of T and q. This allows computation of the total grid-point tendencies as given by the right-hand sides of (124) and (125), and spectral tendencies are computed by direct Fourier and Legendre transforms of these tendencies. This is the approach adopted in the operational ECMWF spectral model, as discussed further in a subsequent contribution to these proceedings.

3.2.6 The computational grid

It is common practice to apply the spectral transform method to multi-level primitive-equation models using the lowest resolution Gaussian grid which according to conditions (76) and (80) guarantees alias-free quadratic products. For the dry, adiabatic primitive equations using sigma coordinates, the highest non-linearity occurs in the form of triple products, and aliasing can be avoided by choosing a sufficiently large grid. More generally, some aliasing is inevitable, but there is much experience which points to insignificantly different results when grids finer than the minimum determined by (76) and (80) are used (e.g. Bourke, 1974; Hoskins and Simmons, 1975; Bourke et al. 1977). Some instances when aliasing becomes evident will be mentioned in our paper on the operational ECMWF spectral model.

3.3 Conservation properties

3.3.1 Barotropic vorticity equation

The following integral quantities are conserved by motion governed by the barotropic vorticity equation:

Angular momentum - $$M = \frac{1}{4\pi} \int_{-1}^{1} \int_{o}^{2\pi} a\, U \, d\lambda d\mu \qquad (130)$$

Kinetic energy - $$K = \frac{1}{8\pi} \int_{-1}^{1} \int_{o}^{2\pi} (u^2 + v^2) \, d\lambda d\mu \qquad (131)$$

Enstrophy - $$E = \frac{1}{8\pi} \int_{-1}^{1} \int_{o}^{2\pi} \xi^2 \, d\lambda d\mu \qquad (132)$$

This may be demonstrated by first rewriting expressions (130) and (131) using integration by parts. This gives

$$M = \frac{a^2}{4\pi} \int_{-1}^{1} \int_{o}^{2\pi} \mu\xi \, d\lambda d\mu \qquad (133)$$

$$K = -\frac{1}{8\pi} \int_{-1}^{1} \int_{o}^{2\pi} \psi\xi \, d\lambda d\mu \qquad (134)$$

from which

$$\frac{\partial M}{\partial t} = \frac{a^2}{4\pi} \int_{-1}^{1} \int_{o}^{2\pi} \mu \frac{\partial \xi}{\partial t} \, d\lambda d\mu = 0 \qquad (135)$$

$$\frac{\partial K}{\partial t} = -\frac{1}{4\pi} \int_{-1}^{1} \int_{o}^{2\pi} \psi \frac{\partial \xi}{\partial t} \, d\lambda d\mu = 0 \qquad (136)$$

$$\frac{\partial E}{\partial t} = \frac{1}{4\pi} \int_{-1}^{1} \int_{o}^{2\pi} \xi \frac{\partial \xi}{\partial t} \, d\lambda d\mu = 0 \qquad (137)$$

Here the equality to zero in each of (135), (136) and (137) is derived by substituting Eqn. (57) for $\partial\xi/\partial t$ into these expressions, and integrating by parts using the basic relationships (52) to (56).

Corresponding conservation relations hold for a spectral barotropic model. To see this, it should be noted that the relationships (135) to (137) will hold for the special case in which ξ (and thus ψ) is given by a truncated spectral expansion of form (58), provided $\partial\xi/\partial t$ is the exact (or untruncated) tendency of ξ. However, the basic spectral technique is such that

$$\frac{\partial \xi}{\partial t} = \overline{\frac{\partial \xi}{\partial t}} + R(\xi) \qquad (138)$$

where $\overline{\frac{\partial \xi}{\partial t}}$ represents the truncated tendency computed by the model, and the residual $R(\xi)$ is <u>orthogonal</u> to the truncated ξ. Since ψ has the same truncation as ξ and μ is proportional to P_1^0, it follows that $R(\xi)$ is also orthogonal to ψ and μ. Thus the relations (135) to (137) hold for a <u>truncated</u> ξ when the <u>truncated</u> tendency $\partial\xi/\partial t$ is used. It follows that apart from time-truncation and round-off error, these conservation relations hold for the spectral model.

Satisfying these conservation relations contributes to the computational stability of the spectral technique, and at first sight appears to be physically realistic. However, if in reality energy or enstrophy is being transferred from resolved to unresolved scales, this process will not be represented in a non-dissipative model, and can lead to an accumulation of energy, and particularly enstrophy, in the shorter retained scales, a process referred to as spectral blocking (Puri and Bourke, 1974). This is most marked for relatively low resolution models, for which it may be alleviated by an appropriate choice of scale-selective horizontal diffusion. The benefits of such a diffusion are illustrated quite distinctly in some idealised barotropic integrations using T21 and T42 resolution reported by Simmons et al. (1983, Figs.22-24). Some further discussion concerning the incorporation of horizontal diffusion in spectral models is given in Sect.3.5, and the subject is treated more fully by Sadourny elsewhere in these proceedings.

3.3.2 The primitive equations

In the usual formulation of primitive-equation spectral models none of the principal atmospheric conservation relations are maintained. This is in part due to the presence of triple products in some quantities, for example the kinetic energy in shallow-water equation and sigma-coordinate models, and in part due to the use of $\ln p_s$ rather than p_s as the basic prognostic variable in multi-level models with terrain-following coordinates, as discussed further in our later contribution to these proceedings. However, experience with both shallow-water equation and multi-level models is that conservation properties in adiabatic simulations are generally good (e.g. Eliasen et al. 1970; Bourke, 1972, 1974; Hoskins and Simmons, 1975).

3.4 Semi-implicit time schemes

We conclude our discussion of the implementation of the spectral technique by indicating in this and the following section how the approach facilitates adoption of a semi-implicit time scheme and linear forms of horizontal diffusion. In both cases, the property that the spherical harmonics are eigenfunctions of the two-dimensional Laplacian is basic to the application.

We consider first the semi-implicit treatment of gravity-wave terms, following Robert et al. (1972), and more specifically the first applications to multi-level spectral models by Bourke (1974) and Hoskins and Simmons (1975). The scheme can be written in the following form for the divergence, temperature and surface pressure equations

$$\delta_t D = \mathscr{D} - \nabla^2 \{\gamma \Delta_{tt} T + \varepsilon \Delta_{tt} \ln p_s\} \tag{139}$$

$$\delta_t T = \mathscr{T} - \tau \Delta_{tt} D \tag{140}$$

$$\delta_t \ln p_s = \mathscr{P} - \pi \Delta_{tt} D \tag{141}$$

where the finite difference operators δ_t and Δ_{tt} are defined by

$$\delta_t X = \frac{1}{2\Delta t}\left\{X(t+\Delta t) - X(t-\Delta t)\right\} \tag{142}$$

$$\Delta_{tt} X = \frac{1}{2}\left\{X(t+\Delta t) + X(t-\Delta t)\right\} - X(t) \tag{143}$$

In (139) to (141) $\mathscr{D}$, $\mathscr{T}$ and $\mathscr{P}$ represent the total tendency of D, T and $\ln p_s$ calculated according to Eqns. (114), (95) and (195) using values at time t for the unparameterized terms. We assume a vertical discretization so that D, T, $\mathscr{D}$ and $\mathscr{T}$ are horizontally-varying column vectors with each element representing, for example, one level in the vertical. γ and τ are constant matrices, and ε and π constant (column and row, respectively) vectors. Their detailed form will not be specified here; we simply note that they are computed by linearization about a basic resting state with a specified temperature profile, and for terrain-following coordinates more general than sigma a specified surface pressure (Simmons and Burridge, 1981). For this basic state, the gravity wave equations with a continuous time representation are given by

$$\frac{\partial D}{\partial t} = -\nabla^2 (\gamma T + \varepsilon \ln p_s) \tag{144}$$

$$\frac{\partial T}{\partial t} = -\tau D \tag{145}$$

$$\frac{\partial \ln p_s}{\partial t} = -\pi D \tag{146}$$

from which we obtain

$$\left(\frac{\partial^2}{\partial t^2} - B\nabla^2\right)D = 0 \tag{147}$$

where

$$B = \gamma\tau + \varepsilon\pi \tag{148}$$

The eigenvalues and eigenfunctions of the matrix B thus determine the gravity-wave phase speeds and vertical structures associated with the basic state.

Eliminating $T(t+\Delta t)$ and $\ln p_s\ (t+\Delta t)$ from (139) to (141) we obtain an equation of the form

$$(I - B\Delta t^2\ \nabla^2)\ D(t+\Delta t) = RHS \qquad (149)$$

where I is the unit matrix and RHS depends linearly on $\mathscr{D}$, $\mathscr{T}$ and $\mathscr{P}$ and on values of D, T and $\ln p_s$ at time levels t and $t-\Delta t$. In a spectral model this equation is simply solved by matrix algebra:

$$D_n^m(t+\Delta t) = (I + \frac{n(n+1)\Delta t^2}{a^2}\ B)^{-1}\ (RHS)_n^m \qquad (150)$$

Values for the spectral coefficients of $T(t+\Delta t)$ and $\ln p_s\ (t+\Delta t)$ are then computed directly from (140) and (141)

The inverse matrices in (150) need be computed once only and stored, and in general the additional computational cost per timestep of including a semi-implicit time scheme in spectral models is negligible. In previous sections we have indicated how the explicit spectral tendencies $\mathscr{D}_n^m$, $\mathscr{T}_n^m$ and $\mathscr{P}_n^m$ are calculated by quadrature. Completion of the calculation of RHS, and of the time-level t and $t-\Delta t$ contributions to the right-hand sides of (140) and (141) can be carried out in spectral space, as in the first multi-level spectral models. However, in the organisation of the high-resolution model developed at ECMWF for optimum use of peripheral storage (Baede et al. 1979), only one time level of spectral coefficients is used, and the relevant linear calculations are performed in grid-point space prior to the quadrature. A complete specification of these calculations may be found in the documentation manual of the current operational ECMWF model.

The spectral technique also lends itself to further optimisation of the time-stepping scheme. An example is given in our subsequent lecture on the design of the present ECMWF spectral model, while scope may exist for further developments, for example along the lines discussed by Orszag (1979).

3.5 Horizontal diffusion

We consider the equation

$$\frac{\partial \xi}{\partial t} = \dot{\xi} - (-1)^p K\nabla^{2p}\xi \tag{151}$$

where the second term on the right-hand side (for which p is an integer) acts as a scale-selective dissipation. Transforming to spectral space this becomes

$$\frac{\partial \xi_n^m}{\partial t} = (\dot{\xi})_n^m - \frac{Kn^p\,(n+1)^p}{a^{2p}}\,\xi_n^m \tag{152}$$

This may readily be integrated using an implicit scheme for the diffusion term:

$$\frac{1}{2\Delta t}\left\{\xi_n^m(t+\Delta t) - \xi_n^m(t-\Delta t)\right\} = (\dot{\xi}(t))_n^m - \frac{Kn^p(n+1)^p}{a^{2p}}\;\xi_n^m(t+\Delta t) \tag{153}$$

giving

$$\xi_n^m(t+\Delta t) = \frac{\xi_n^m(t-\Delta t) + 2\Delta t\;\dot{\xi}(t)_n^m}{1 + 2K\Delta t n^p\,(n+1)^p/a^{2p}} \tag{154}$$

It is thus a very straightforward matter to include a stable, linear diffusion within the spectral technique. As implemented here with an implicit scheme, the diffusion acts each timestep in a scale-selective way to reduce amplitudes of spectral coefficients calculated for t+Δt in the absence of diffusion. A variety of scale dependences other than $n^p(n+1)^p$ can be used, but dissipation may not then be local in physical space. The use of vorticity and divergence as prognostic variables also facilitates the use of a stronger damping of divergence if required.

Linear diffusion of the type illustrated here is widely used in spectral models, with apparently satisfactory results. Adoption of a non-linear diffusion as used in a number of finite-difference models requires significant programming effort and computational cost, although an example of such an application has been given by Gordon and Stern (1982).

4. THEORETICAL DIFFERENCES BETWEEN SPECTRAL AND GRID POINT MODELS

In this section, we do not aim at giving an exhaustive list of all differences between spectral and grid point models. We rather wish to concentrate on the differences which have been mentioned as potentially relevant here and there in the literature.

4.1 The pole problem

In a finite difference model on a latitude-longitude grid there is a convergence of the meridians towards the poles. As a consequence, there is a severe restriction on the timestep needed to avoid numerical instability. Many solutions have been proposed for this problem, however several led to sharp controversies since they were sometimes solving the numerical problem at the expense of some extra physical problems. The ideal solution would of course be to have an isotropic resolution, or at least to simulate it as closely as possible. This has been done at GFDL by modifying the grid near the poles (Kurihara 1965) or in other models by Fourier analysing the fields or their time tendencies and discarding shorter zonal waves in order to have an approximately constant east-west resolution. A variant of this approach has been developed at ECMWF (Burridge, unpublished) achieving a similar result through a modification of the horizontal diffusion. The results have proved very satisfactory and it is probably fair to say that "pole problems" should no longer be considered a problem for finite difference models, although care is required at the design stage.

For spectral models using spherical harmonics as expansion functions and a triangular truncation, there cannot be any pole problem since, as mentioned in Sect.2.3 the resolution is then isotropic over the sphere. However, it is worth mentioning that such is not the case for a rhomboidal truncation. There is more resolution at high latitudes, mostly in the medium part of the

spectrum (cf Sect.2.3) which may lead to a certain form of pole problem.

Furthermore spectral models using bi-Fourier expansions have a pole problem almost identical to the one in finite difference models and similar remedies (e.g. Fourier chopping) can be used.

4.2 Aliasing errors

Let us consider the simple advection equation on a circle

$$\frac{\partial u}{\partial t} = - u \frac{\partial u}{\partial \lambda} \tag{155}$$

If we consider 2N points on the circle, i.e.

$$\lambda_i = \frac{2\pi}{2N} (i-1) \quad i = 1,\ldots 2N \tag{156}$$

then the grid distance is $\Delta\lambda = \frac{\pi}{N}$ (157)

u, assumed to be periodic ($u(\lambda + 2\pi) = u(\lambda)$), can be expanded in a Fourier series.

$$u(\lambda,t) = \sum_{m=-N}^{N} u_m(t)\, e^{im\lambda} \tag{158}$$

with

$$u_m(t) = \frac{1}{2\pi} \int_0^{2\pi} u(\lambda,t)\, e^{-im\lambda}\, d\lambda \tag{159}$$

If we assume the horizontal derivatives to be computed exactly, then

$$\frac{\partial u}{\partial \lambda} = \sum_{n=-N}^{N} i\, n\, u_n(t)\, e^{in\lambda} \tag{160}$$

and $$u\frac{\partial u}{\partial \lambda} = \sum_{n=-N}^{N} \sum_{m=-N}^{N} i\, n\, u_n(t)\, u_m(t)\, e^{i(n+m)\lambda} \tag{161}$$

At the points λ_i defined by (156) we have

$$u(\lambda_i,t) \frac{\partial u}{\partial \lambda}(\lambda_i,t) = \sum_{n=-N}^{N} \sum_{m=-N}^{N} i\, n\, u_n(t)\, u_m(t)\, e^{i(n+m)\lambda_i} \tag{162}$$

n+m varies from -2N to 2N and we have only 2N points. We can thus resolve only $e^{ip\lambda}$ for p = -N,..N.

As a consequence, wavenumbers with $|p|>N$ are misrepresented (aliased) as longer waves since

$$e^{ip\lambda_j} = e^{i[p\lambda_j-(j-1)2\pi]} \tag{163}$$

and as a consequence of (156) $(j-1)2\pi = 2N\lambda_j$.

This implies

$$e^{ip\lambda_j} = e^{i(p-2N)\lambda_j} \tag{164}$$

Thus p is aliased as p-2N.

This aliasing can be a spurious source of energy in finite difference models and even if the time step is chosen such as to avoid linear instability, it can lead to non-linear instability (Phillips, 1959). In order to prevent it a solution is to filter all waves with $|p|>N$ but this becomes more complicated for two dimensional problems. A second solution is to impose certain integral constraints to be satisfied in the discrete approximations (e.g. energy conservations), inhibiting spurious growth of the amplitudes of the smallest scales: it removes the instability problem, but not the aliasing errors. However these are usually small and it is interesting to note that they are further reduced when a Fourier chopping is used to solve the pole problem (as discussed in the previous section).

Spectral models, as mentioned in Sect.3.2 and 3.3, do not suffer from aliasing errors in the computation of quadratic terms, which prevents the non linear instability mentioned above. They nevertheless generally accept aliasing from triple (or more) products and from division (as do grid point models).

Another form of aliasing errors suffered by both grid point and spectral models is what is sometimes improperly called "initial" aliasing and is very closely related to truncation errors, although not identical, and smaller in amplitude. It is easy to explain them for the orography although they also affect initial fields (whence their name). Since the actual mountains have a continuous infinite spectrum averaging or sampling is thus necessary which modifies slightly all scales, in particular the smallest retained. Fourier components for example cannot be evaluated exactly by quadrature formulae since they correspond to integrals of "polynomials" of infinite degree. This aliasing can however be reduced (almost) at will by increasing the number of points used for quadratures and by projecting back on the selected grid (this method applies also to grid point models) which leaves only the truncation error. An example is given in Fig.6 for the line of latitude 40°S analysed with 63 zonal wave numbers either from a grid of 2160 points or from one of 192 points (after a simple averaging of the orography). The resulting local differences, although small, may not be fully negligible.

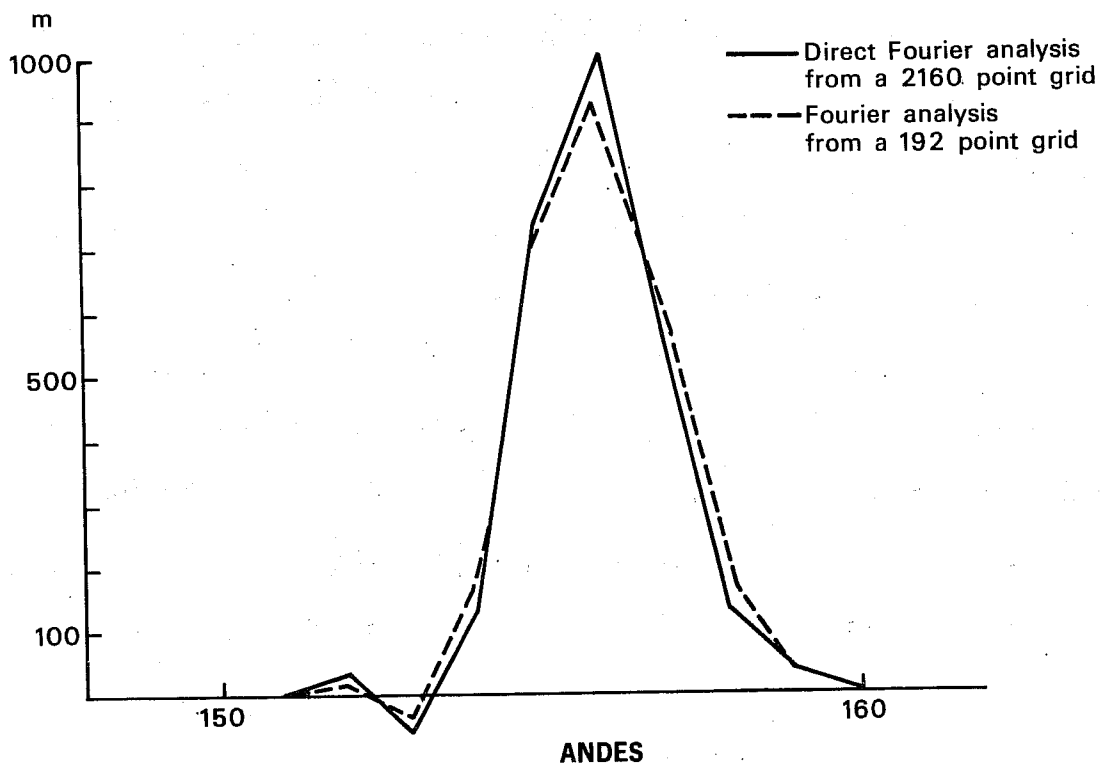


Fig. 6 Fourier analysis of the part of the circle at 40°S keeping 63 zonal wavenumbers.

4.3 Linear phase error

It is generally accepted that this is one of the important contributions to differences between spectral and grid point models. To illustrate it, let us consider the simple linear advection equation:

$$\frac{\partial u}{\partial t} = - \omega \frac{\partial u}{\partial \lambda} \tag{165}$$

Then

$$u = e^{i(m\lambda - ct)} \tag{166}$$

is a solution of (161) if

$$\underline{c = \omega m} \tag{167}$$

A second-order accurate discrete analogue of $\frac{\partial u}{\partial \lambda}$ may be

$$\frac{u(\lambda+\Delta\lambda) - u(\lambda-\Delta\lambda)}{2\Delta\lambda}$$

which implies u to be solution if

$$\underline{\tilde{c} = \omega \frac{\sin m\Delta\lambda}{\Delta\lambda}} \tag{169}$$

Comparisons of (169) and (167) shows that the discretized solution has a phase speed $\tilde{c}$ which is different from the exact one c. The absolute value is smaller, meaning that the waves move more slowly than in reality. In particular for the smallest wave representable (m=N, corresponding to $m\,\Delta\lambda = N\frac{2\pi}{N} = 2\pi$) $\tilde{c}=0$. This phase error is a function of the wavelength leading to a spurious dispersion of complex systems.

The linear phase error is of course a function of resolution ($c - \tilde{c}$ tends to zero when $\Delta\lambda$ tends to zero), but it is shown in a following lecture that it remains a significant source of error even at the comparatively high resolution used by the ECMWF grid point model (1.875 degree grid).

High order approximations for the derivatives, at least for the advection terms would also reduce the error, but at some computational expense.

4.4 Coupling errors

Although hardly ever mentioned in the literature, these errors are very similar in nature to the linear phase errors; they originate from the inexact computation of horizontal non linear differential operators and result in a misrepresentation of the non linear interactions between the various scales of motion.

Let us illustrate them with the simple non-divergent barotropic equation for a non-rotating atmosphere, following the approach by Lilly (1965).

$$\frac{\partial \xi}{\partial t} = - J(\psi,\xi) \tag{170}$$

where

$$\xi = \nabla^2 \psi \tag{171}$$

As in Sect.3.1, ξ is the vorticity and ψ the streamfunction. J is the horizontal Jacobian operator for a plane geometry:

$$J(\psi,\xi) = \frac{\partial \psi}{\partial x}\frac{\partial \xi}{\partial y} - \frac{\partial \psi}{\partial y}\frac{\partial \xi}{\partial x} \tag{172}$$

Let us make a bi-Fourier expansion of ψ.

$$\psi = \sum_{n,m} A_n^m \, e^{i(mx+ny)} = \sum_{n,m} \psi_n^m \tag{173}$$

then

$$\xi = - \sum_{n,m} (n^2+m^2) \, \psi_n^m \tag{174}$$

and

$$\begin{aligned} J(\psi,\xi) &= \sum_{n,m} n\psi_n^m \sum_{p,q} (p^2+q^2) \, p\psi_q^p \\ &\quad - \sum_{n,m} m\psi_n^m \sum_{p,q} (p^2+q^2) \, q\psi_q^p \\ &= \sum_{n,m,p,q} (pn-mq)(p^2+q^2)\psi_q^p \, \psi_n^m \end{aligned} \tag{175}$$

or in a symmetric way:

$$J(\psi,\xi) = \tfrac{1}{2} \sum_{n,m,p,q} (pn-mq)\left[p^2+q^2)-(n^2+m^2)\right] \psi_q^p \psi_n^m \qquad (176)$$

We now define

$$\delta_x F = \frac{1}{\Delta x}\left[F(x + \frac{\Delta x}{2}) - F(x - \frac{\Delta x}{2})\right]$$

$$\overline{F}^x = \frac{1}{2}\left[F(x + \frac{\Delta x}{2}) + F(x - \frac{\Delta x}{2})\right] ,$$

and replace

$\frac{\partial F}{\partial x}$ by $\delta_x \overline{F}^x$, $\frac{\partial F}{\partial y}$ by $\delta_y \overline{F}^y$, $\nabla^2 F$ by $\delta_{xx} F$ and

$J(\psi,\xi)$ by $\delta_x \bar{\psi}^x \delta_y \bar{\xi}^y - \delta_y \bar{\psi}^y \delta_x \bar{\xi}^x$

which leads to a second order scheme.

If we now make use of (173) again, we obtain

$$J(\psi,\xi) = \frac{1}{2} \sum_{n,m,p,q} \left\{\frac{\sin p\,\Delta x}{\Delta x} \frac{\sin n\,\Delta y}{\Delta y} - \frac{\sin q\,\Delta y}{\Delta y} \frac{\sin m\,\Delta x}{\Delta x}\right.$$
$$\left. \cdot \left\{R(p,q) - R(m,n)\right\} \psi_q^p \psi_n^m \right. \qquad (177)$$

where $R(p,q) = \left(\frac{\sin p \frac{\Delta x}{2}}{\frac{\Delta x}{2}}\right)^2 + \left(\frac{\sin q \frac{\Delta y}{2}}{\frac{\Delta y}{2}}\right)^2$ (178)

corresponds to the Laplacian operator.

(177) is to be compared with (176).

The errors made are clearly very similar to the ones observed for the linear phase problem (m was approximated by $\frac{\sin m\,\Delta\lambda}{\Delta\lambda}$) and evidence for the importance of linear phase errors guarantees the presence of non linear coupling errors with similar amplitude. They are such as to underestimate the interactions

between the various scales and in particular the shortest scales. They can be expected a priori to act as a reduction of the resolution. It is easy to see that their magnitude is larger than that of aliasing errors.

Evidence of their practical importance can be deduced from Williamson (1978): he showed second order schemes to behave in a way similar to fourth order schemes with coarser resolution. Since the differences between them lie in the accuracy of the discrete representation of the derivatives, and in view of the largely non linear nature of the atmosphere it supports the assertion on the effects of coupling errors made above. This is further strengthened by some results obtained at ECMWF comparing a grid point and a spectral model, showing that on average the grid point model was behaving like a spectral model with a significantly lower resolution (Jarraud et al. 1979). These coupling errors are also likely to be responsible for some of the largest differences seen between spectral and grid point models at ECMWF during the course of an extensive comparison (Girard and Jarraud, 1982) which led to the choice of the spectral technique for the ECMWF operational model, as discussed in the following two contributions to these proceedings.

Acknowledgements

The reviews by Bourke et al (1977) and Machenhauer (1979), and general discussions of the spectral method contained in the papers of Orszag, greatly assisted us in the preparation of this account of spectral techniques.

References

Baede, A.P.M., Jarraud, M. and Cubasch, U., 1979: Adiabatic formulation and organisation of ECMWF's spectral model. ECMWF Tech.Rep.No.15, pp.40.

Baer, F., 1972: An alternate scale representation of atmospheric energy spectra. J.Atmos.Sci., 29, p.649-664.

Belousov, S.L., 1962: Tables of normalized associated Legendre polynomials. Pergamon Press, New York, 379 pp.

Blinova, E.N., 1942: Hydrodynamic theory of pressure and temperature waves and centres of action of the atmosphere. Translated from Russian (1944) (Regional control offices, Second Weather Region, Pattersonfield, USA).

Bourke, W., 1972: An efficient one level primitive equation spectral model. Mon.Wea.Rev., 100, p.683-689.

Bourke, W., 1974: A multi level spectral model I: Formulation and hemispheric integrations. Mon.Wea.Rev., 102, p.687-701.

Bourke, W., McAvaney, B., Puri, K. and Thurling, R., 1977: Global modeling of atmospheric flow by spectral methods. Methods in computational physics. Vol.17: General circulation models of the atmosphere. Ed. J.Chang, Academic Press, p.267-324.

Cooley, J.W. and Tukey, J.W., 1965: An algorithm for the machine calculation of complex Fourier series. Math comp 19, p.297-301.

Courant, R. and Hilbert, D. 1953: Methods of mathematical physics. Vol.I. New York, Intersciences, 562 pp.

Daley, R., Girard, C., Henderson, J., Simmonds, I. 1976: Short term forecasting with a multi-level spectral primitive equation model. Atmosphere 14, p.98-134.

Daley, R. and Bourassa, Y. 1978: Rhomboidal versus triangular spherical harmonic truncation: some verification statistics. Atmosphere, 16, p.187-196.

Eliasen, E., Machenhauer, B. and Rasmussen, E. 1970: On a numerical method for integration of the hydrodynamical equations with a spectral representation of the horizontal fields. Rep.No.2 Institut for Teoretisk Meteorologi, University of Copenhagen.

Ellsaesser, H.W. 1966: Evaluation of spectral versus grid methods of hemispheric numerical weather prediction. J.Appl.Meteor., 5, p.246-262.

GARP publication series No.17, 1979: Numerical methods used in atmospheric models. Vol.2 WMO/ICSU Geneva, Switzerland.

Girard, C. and Jarraud, M. 1982: Short and medium range forecast differences between a spectral and grid point model. An extensive quasi operational comparison. ECMWF Tech.Rep.No.32, pp.177.

Gordon, C.T. and Stern, W.F. 1982: A description of the GFDL global spectral model. Mon.Wea.Rev., 110, 625-644.

Haurwitz, B. and Craig, R.A. 1952: Atmospheric flow patterns and their representation by spherical surface harmonics. A.F.R.L. Geophysical research paper No.14, pp.78.

Hobson, E.W. 1931: The theory of spherical and ellipsoidal harmonics. Cambridge University Press, 500 pp.

Hoskins, B.J. and Simmons, A.J. 1975: A multi-layer spectral model and the semi-implicit method. Q.J.R.M.S., 101, P.637-655.

Jarraud, M., Girard, C. and Cubasch, U. 1981: Comparison of medium range forecasts made with models using spectral or finite difference techniques in the horizontal. ECMWF Tech.Rep.No.23, pp.96.

Kurihara, A. 1965: Numerical integration of the primitive equations on a spherical grid. Mon.Wea.Rev., 93, p.399-415.

Lilly, D.K. 1965: On the computational stability of numerical solutions of time dependent non linear geophysical fluid dynamic problems. Mon.Wea.Rev., 93, p.11-25.

Machenhauer, B. and Rasmussen, E. 1972: On the integration of the spectral hydrodynamical equations by a transform method. Rep.No.3 Institut for Teoretisk Meteorologi, University of Copenhagen.

Machenhauer, B. and Daley, R., 1972: A baroclinic primitive equation model with a spectral representation in three dimensions. Rep.No.4 Institut for teoretisk meteorologie. University of Copenhagen.

Machenhauer, B. and Daley, R. 1974: Hemispheric spectral model. GARP publication series No.14, p.226-251.

Machenhauer,B. 1979: The spectral method. Numerical methods used in atmospheric models. GARP publication series No.17. p.121-275.

Orszag, S.A., 1970: Transform method for calculation of vector coupled sums: application to the spectral form of the vorticity equation. J.A.S., 27, p.890-895.

Orszag, S.A., 1974: Fourier series on spheres. Mon.Wea.Rev., 102, p.56-75.

Orszag, S.A. 1979: Spectral methods for problems in complex geometrics. Numerical methods for partial differential equations. Edited by S. Porter. Academic Press, New York.

Phillips, N.A. 1959: An example of nonlinear computational instability. The atmosphere and the sea in motion. Rossby memorial volume. New York. Rockefeller instit. press, p.501-504.

Puri, K. and Bourke, W. 1974: Implications of horizontal resolution in spectral model integrations. Mon.Wea.Rev., 102, p.333-347.

Robert, A.J., 1966: The integration of a low order spectral form of the primitive meteorological equations. J.Meteor.Soc.Jap., 44, p.237-244.

Robert, A.J., Henderson, J. and Turnbull, C. 1972: An implicit time integration scheme for baroclinic models of the atmosphere. Mon.Wea.Rev., 100, p.329-335.

Rochas, M., 1973: Fonctions sphériques. Applications à la météorologie. Note No.323 de l'EERM. Direction de la metéorologie. France.

Silberman, I., 1954: Planetary waves in the atmosphere. J.meteor. 11, p-27-34.

Simmons, A.J. and Hoskins, B.J. 1975: A comparison of spectral and finite difference simulations of a growing baroclinic wave. Q.J.R.M.S., 101, p.551-565.

Simmons, A.J. and Burridge, D.M., 1981: An energy and angular momentum conserving vertical finite difference scheme and hybrid vertical coordinate. Mon.Wea.Rev., 109, p.758-766.

Simmons, A.J., J.M. Wallace and G.W. Branstator 1983: Barotropic wave propagation and instability, and atmospheric teleconnection patterns. J.Atmos.Sci., 40, 1363-1392.

Temperton, C., 1983: Fast mixed-radix real Fourier transforms. ECMWF Tech.Memo.No.71, 18pp.

APPENDIX DEFINITION OF THE SPHERICAL HARMONICS (following Rochas, 1973)

By definition a spherical harmonic of degree n is a homogeneous function $u(x,y,z)$ of degree n, solution of the Laplace equation.

$$\Delta u = 0 \tag{A.1}$$

where Δu is the three dimensional Laplacian.

In spherical coordinates (λ,θ,r) it can be written

$$\Delta u = \frac{1}{r^2}\left[\frac{1}{\cos\theta}\frac{\partial}{\partial\theta}\left(\cos\theta\,\frac{\partial u}{\partial\theta}\right) + \frac{1}{\cos^2\theta}\frac{\partial^2 u}{\partial\lambda^2} + \frac{\partial}{\partial r}\left(r^2\,\frac{\partial u}{\partial r}\right)\right] \tag{A.2}$$

The homogeneity condition in spherical coordinates is

$$r\,\frac{\partial u}{\partial r} = nu \tag{A.3}$$

which implies:

$$u(\lambda,\theta,r) = r^n\, S_n\,(\lambda,\theta) \tag{A.4}$$

where S_n is an arbitrary function of λ and θ.

Inserting (A.4) into (A.2) gives

$$\frac{1}{\cos\theta}\frac{\partial}{\partial\theta}\left(\cos\theta\,\frac{\partial S_n}{\partial\theta}\right) + \frac{1}{\cos^2\theta}\frac{\partial^2 S_n}{\partial\lambda^2} + n(n+1)\,S_n = 0 \tag{A.5}$$

$S_n\,(\lambda,\theta)$, the solution of (A.5), is also called a spherical harmonic of degree n. It can be obtained by a classical separation method.

We look for a solution of the form:

$$S_n(\lambda,\theta) = A(\theta)\,B(\lambda) \tag{A.6}$$

(A.5) given then

$$\frac{B(\lambda)}{\cos\theta}\frac{d}{d\theta}\left(\cos\theta\,\frac{dA}{d\theta}\right) + \frac{A(\theta)}{\cos^2\theta}\frac{d^2B}{d\lambda^2} + n(n+1)\,A.B = 0 \tag{A.7}$$

which can be rewritten:

$$\alpha(\theta)\,\frac{d^2B}{d\lambda^2} + \beta(\theta)\,B = 0 \qquad (A.8)$$

The solutions are

$$B(\lambda) = a\,e^{im\lambda} \qquad (A.9)$$

Inserting (A.9) into (A.7) leads then to:

$$\frac{1}{\cos\theta}\,\frac{d}{d\theta}\,(\cos\theta\,\frac{dA}{d\theta}) + \left[n(n+1) - \frac{m^2}{\cos^2\theta}\right] A = 0 \qquad (A.10)$$

This is the so called Legendre equation and it can be shown that the solutions are the $P_n^m(\theta)$ or associated Legendre functions of first kind of order m and degree n.