JS

MET O 11 TECHNICAL NOTE NO. 51.

The spectral method of integrating the meteorological equations and some comparisons with the finite difference method

B J Hoskins* (Vacation consultant, Met O 11)

1. Introduction

In recent years, an alternative to the usual finite difference method for integrating the meteorological equations has been developed. In this paper, a description of the spectral method is given (sections 2 and 3). This is followed by a summary of some detailed comparisons of integrations using the two methods for a simple baroclinic instability problem (section 4). Possible forecast models are then compared for timing and storage requirements and other advantages and disadvantages.

* Permanent address: UK Universities Atmospheric Modelling Group, Department of Geophysics, Reading.

2. The Method

The spectral method used in baroclinic models is perhaps best illustrated by its application to the barotropic vorticity equation on the sphere:

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

where λ = longitude

μ = sin (latitude)

a = radius of earth

U = u x cos(latitude) = $-\frac{1}{a}(1-\mu^2)\frac{\partial \psi}{\partial \mu}$ (2)

V = v x cos(latitude) = $\frac{1}{a}\frac{\partial \psi}{\partial \lambda}$ (3)

absolute vorticity $\zeta = 2\Omega\mu + \nabla^2\psi$ (4)

In the spectral method a function such as ζ is represented, not by its value at a finite grid of points, but by the coefficients of a finite set of basis functions. In a periodic box, the convenient functions are products of Fourier series. On the surface of a sphere, again it is convenient to use the eigen functions of the Laplacian operator ie the spherical harmonics $Y_n^m(\mu,\lambda)$. These satisfy

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

and the orthogonality relation $\iint Y_{n_1}^{m_1} (Y_{n_2}^{m_2})^* \, d\mu \, d\lambda = 2\pi \delta_{m_1 m_2} \delta_{n_1 n_2}$

They are products of Fourier series in the EW with Legendre functions in the NS:

$$Y_n^m(\mu,\lambda) = P_n^m(\mu) e^{im\lambda}$$

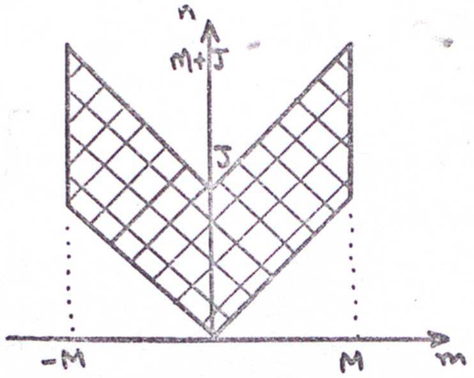
The index m is, thus a zonal wave number, n is a total wavenumber, $a/[n(n+1)]^{1/2}$ being the length scale of the harmonic.

For some purposes, n - $|m|$+1 may be considered a latitudinal wavenumber. The spherical harmonics are defined only for n $\geqslant |m|$, and Y_n^{-m} is the complex conjugate of Y_n^m. Thus a real function is represented

$$\zeta(\mu,\lambda) = \sum_{m=-M}^{M} \sum_{n=|m|}^{N} \zeta_n^m Y_n^m(\mu,\lambda)$$

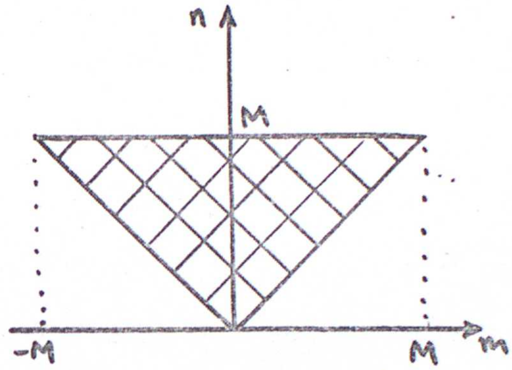
where ζ_n^{-m} is the complex conjugate of ζ_n^m and only the complex numbers ζ_n^m for $m \geq 0$ need be stored to define ζ.

In contrast with the variety of choices of finite difference grids and operators, there is really only one choice to be made in a spectral model. This is the choice of truncation. The two truncations commonly used are called rhomboidal and triangular (Fig 1).



Rhomboidal truncation (M,J)

$N = |m| + J$



Triangular truncation (M)

$N = M$

In rhomboidal truncation the same number of latitudinal modes is retained for each zonal wavenumber. In triangular truncation only scales greater than a certain value are retained. For a variety of reasons the author prefers the latter truncation. With this truncation the spectral method treats every point and every direction on the sphere in the same manner. Programs may be written so that it is a minor job to change from one truncation to the other.

It should be noted that the spherical harmonics are either even or odd about the equator, and if an integration is to be performed on one hemisphere only, then half the coefficients are identically zero and, for efficiency, need not be stored.

The spectral method was pioneered by Silberman (1954) and Baer and Platzman (1961). They used Eq. 1 to determine the interaction coefficients which described the interaction of two spherical harmonic components to produce a third. Though suggestive theoretically, this method was not suitable for the forecast problem. For anything but a severe truncation the calculation soon became unwieldy and long. Furthermore there was little possibility of the inclusion of local physical processes.

The introduction of the spectral-transform method (Orszag 1970; Eliassen et al 1970) has now removed both these obstacles to the extent that a spectral method is a serious competition with the usual finite difference methods. Substituting the series for ζ in the left hand side of Eq. 1, multiplying by $(Y_n^m)^*$ and integrating over the sphere gives

$$2\pi a \dot{\zeta}_n^m = \iint \left[-\frac{1}{1-\mu^2}\frac{\partial}{\partial \lambda}(U\zeta) - \frac{\partial}{\partial \mu}(V\zeta) \right] P_n^m(\mu)\, e^{-im\lambda}\, d\mu\, d\lambda$$

$$= \int d\mu \left[\frac{P_n^m(\mu)(-im)}{1-\mu^2} \int \underbrace{U\zeta e^{-im\lambda}}_{(A)} d\lambda + \frac{dP_n^m}{d\mu} \int \underbrace{V\zeta e^{-im\lambda}}_{(B)} d\lambda \right]$$

using integration by parts. The spectral coefficients of ζ directly (from Eq. 4) imply those of the streamfunction ψ . From Eqs. 2 and 3, these determine the spectral coefficients of U and V. The integrands (A) and (B) may be shown to be complex exponentials in λ with wavenumber between -3M and +3M. The zonal integration may therefore be performed exactly at every latitude by summation of values at at least 3M + 1 equally spaced zonal points. It may also be shown that the integrand for the NS integration is then a polynomial in μ of maximum degree 2M + 3J (rhomboidal) or 3M (triangular). Hence this integration may be evaluated exactly by Gaussian integration using at least $\frac{2M + 3J + 1}{2}$ or $\frac{3M + 1}{2}$ specially chosen latitudes.

Thus, to perform the integration, the sphere is covered with a grid composed of at least 3M + 1 equal spaced lines of longitude and $\frac{2M + 3J + 1}{2}$ or $\frac{3M + 1}{2}$ lines of latitude. The latter must be at "Gaussian" latitudes which are very nearly equally spaced. Grid point values of U, V and ζ are obtained from their spectral representations. The products $U\zeta$, $V\zeta$ are obtained by grid point multiplications and the integration performed by taking Fourier sums in the EW direction, multiplying by Gaussian weights and taking sums in the NS direction. <u>It may be shown that this method is exact for the retained modes.</u>

The basis of this method is to perform linear operations, particularly derivatives in spectral space, but to move to physical space to perform products.

The physical space is chosen so that on moving back to spectral space the retained modes describing the product are exact. The method is aided by two factors. The transformation in the zonal direction from spectral to grid and the zonal part of the integration may be performed using the extremely fast Fourier transform routines available. These are most efficient if the number of lines of longitude is a power of 2 and little worse if it is a power of 2 times a power of 3. The second factor is that the tendency of each spectral coefficient from one line of latitude may be computed and added to the accumulated values before moving on to the next latitude. Thus the grid point values at only one latitude are required at any time.

Apart from severe truncations, this transform method is quicker than the interaction coefficient method. It also allows the possibility of putting in local "physics" when at the grid-point stage.

3. The primitive equations

When proceeding to the application of the spectral method to the full primitive equations, the question arises of what vertical representation to use. Machenauer and Daley (1972) formulated a model using a spectral method for all three dimensions. However this work has encountered many problems associated with the upward propogation of waves and their amplification and representation at the top of their atmosphere. It is not at all clear that conventional layered models correctly treat the top of the atmosphere and much study will probably be performed on this problem in the next few years. It appears that in the next decade the spectral method will be efficiently applied only in the horizontal and that levels will be used in the vertical, though with some modification in upper levels. In what follows it will be assumed that σ coordinates and a finite difference scheme are used in the vertical.

Since only true scalers may be represented by a series of spherical harmonics, it is convenient to use the primitive equations in their vorticity and divergence form. This also facilitates the application of the semi-implicit time scheme (see below) and in any diagnostic study is conceptually helpful. The dry equations may then be written:

$$\frac{\partial \zeta}{\partial t} = \frac{1}{a}\frac{1}{1-\mu^2}\frac{\partial}{\partial \lambda}\mathcal{F}_V - \frac{1}{a}\frac{\partial}{\partial \mu}\mathcal{F}_U$$

$$\frac{\partial D}{\partial t} = \frac{1}{a}\frac{1}{1-\mu^2}\frac{\partial}{\partial \lambda}\mathcal{F}_U + \frac{1}{a}\frac{\partial}{\partial \mu}\mathcal{F}_V - \nabla^2\left[\frac{U^2+V^2}{2(1-\mu^2)} + \phi + R\bar{T}\ln p_*\right]$$

$$\frac{\partial T'}{\partial t} = -\frac{1}{a}\frac{1}{1-\mu^2}\frac{\partial}{\partial \lambda}(UT') - \frac{1}{a}\frac{\partial}{\partial \mu}(VT') + DT' - \dot{\sigma}\frac{\partial T}{\partial \sigma} + \kappa\frac{T\omega}{p} + Q_T$$

$$\frac{\partial \ln p_*}{\partial t} = -V\nabla \ln p_* - D - \frac{\partial \dot{\sigma}}{\partial \sigma}$$

$$\frac{\partial \phi}{\partial \ln \sigma} = -RT$$

where D is the divergence, $T = \bar{T}(\sigma) + T'$ the temperature, p_* the surface pressure, ϕ the geopotential,

$$\mathcal{F}_U = V\zeta - \dot{\sigma}\frac{\partial U}{\partial \sigma} - T'\frac{\partial \ln p_*}{\partial \lambda} + Q_U \;, \quad \mathcal{F}_V = -U\zeta - \dot{\sigma}\frac{\partial V}{\partial \sigma} - T'(1-\mu^2)\frac{\partial \ln p_*}{\partial \mu} + Q_V \;,$$

and Q_U, Q_V, Q_T are the sources and sinks in the U, V and T equations. A water vapour equation may be added and latent heat release added for Q_T.

The integration procedure is much as described for the barotropic vorticity equation. It is described in detail in Bourke (1974) and Hoskins and Simmons (1975)*. The same horizontal grid as described previously is used. This grid is insufficient to determine exactly the implied triple product in the vertical advection but various studies have shown the implied errors to be small. This should be tested again in the presence ofsteep topography.

In HS it is shown how the above equations may be integrated using a semi-implicit scheme in which the short external gravity wave modes are slowed so that a large timestep may be used. This demands only slightly more computation. Thus

* Hereafter denoted by HS

the whole computation time may be reduced by a factor of at least 6 compared with an explicit scheme for a negligible loss in accuracy (see HS). This is in contrast with finite difference models in which either the computation is much increased so that only a factor of $2\frac{1}{2}$ is realised (Gauntlett and Leslie, 1974), or second order accuracy is sacrificed by the use of a splitting method (Burridge, 1975).

4. Some comparisons with finite difference integrations

When trying to compare integration methods for the meteorological equations, the major problem to overcome is that no exact solutions of the nonlinear equations are known. One could compare (eg by RMS error of 500mb height field) full forecast models, but this allows the damaging effects of initialisation and analysis errors, and physics more suited to one model. A different approach used by the UK Universities Group (Simmons and Hoskins 1975, hereafter referred to as SH) is to take a simple situation and models with no "physics". By using increasing resolution with two different integration methods one obtains a good idea of the correct answer and thus of the errors in the various resolutions. In SH comparisons were made using as initial conditions a baroclinically unstable zonal jet plus a very small perturbation in zonal wavenumber 8. Four spectral models using triangular truncation at wavenumbers 21 and 42, and rhomboidal truncation at wavenumbers (16,15) and (32,31), and two second order finite difference models with grids of $5^{\circ} \times 3^{\circ}$ and $2\frac{1}{2}^{\circ} \times 1\frac{1}{2}^{\circ}$ were compared. In all the models a wavenumber 8 baroclinic wave grew, producing its warm and cold fronts. In SH a detailed synoptic comparison was made. Further, a spherical harmonic analysis of the finite difference results was made and the behaviour of individual wave components studied. These comparisons are all described in some detail in SH. To highlight one of the advantages of the spectral method, inclusion is made here of one figure from SH (fig. 1). This shows that the phase differences between spectral models is small whilst that between finite difference models can be large, there being indications of convergence towards the spectral phase value.

There can never be an exact equivalence made between two resolutions of the spectral and finite difference methdd because their inherent errors are different. However, from this work we may tentatively state that T42 (triangular truncation M = 42) is equivalent to second order finite difference on a $2\frac{1}{2}^{\circ} \times 1\frac{1}{2}^{\circ}$ grid, in that the medium and large scale are better treated in the spectral model, and the smallest scale better treated in the finite difference model.

5. Comparison of possible forecast models.

In order to attempt a comparison of the requirements, capabilities and drawbacks of spectral and finite difference methods for a possible forecast model a second-order finite difference model on a 1° latitude-longitude hemispheric grid with 10 layers in the vertical is considered*. Based on the comparison in the previous section, we may postulate that an "equivalent" resolution spectral model is somewhere in the region T63 or T84, the latter very likely being superior in most respects. Experiments with the T63 resolution do indeed show that it is capable of producing realistic baroclinic waves with strong fronts. This resolution uses a transform grid of 48 latitudes and 192 longitudes, equivalent approximately to a 1.9° grid. The T84 resolution requires 64 latitudes and 256 longitudes, approximately a 1.4° grid. Figures for this resolution will be given in brackets.

Total storage figures may often be misleading in that not all the arrays will have to be in core at any instant. This is particularly true for a finite difference model, but is also true for a spectral model. However, consideration must be made of the total storage requirement and of the necessary shuffling in and out of core. For the finite difference model, estimate may be made from the need for 10 arrays each of 8 x 90 x 360. This suggests a requirement of at least 2,800k words. For the spectral models the estimates from 16 arrays of $10 \times 63^2/2$ ($10 \times 84^2/2$) are 400 k (800 k) words.

* Storage and computation time for both models are approximately proportional to the number of layers and should be doubled for the whole sphere.

For the adiabatic part (dynamics) of the model, timings will be given based on the experience of the UK Universities Atmospheric Modelling Group running programs on a CDC7600. The spectral models and the smoothing in the finite difference models use a machine code fast Fourier transform. Otherwise the models are programmed in Fortran. As shown in the Appendix, T63 (T84) would require 18 (56) minutes for a forecast day, using a semi-implicit time scheme. An explicit time scheme 1^{o} finite difference model would require 296 minutes per day. Making large allowances for a semi-implicit scheme and/or a different mesh might reduce this time to nearer the T84 value. However the spectral method clearly has a huge speed advantage.

This is also true for the diabatic part of the calculation. The timing for this is proportional to the number of grid points which is a factor of 3.5 (2.0) smaller for the spectral models.

As mentioned previously, the polar point for a triangularly truncated spectral model is treated the same as every other point. The polar problem which has been so prominent in the finite difference literature is thus nonexistant. This may be graphically demonstrated by solving the barotropic vorticity equation for both a super-rotation and a Rossby-Haurwitz wave when the axis of the coordinate system is at right angles with the rotation axis. After 5 days the super-rotation is exact to within round-off error, $O(10^{-11})$ - see Fig 2a. In $11\frac{1}{4}$ days, due to the time scheme, there is a phase error of $\frac{1}{4}^{o}$ in the Rossby-Haurwitz wave integration - see Fig 2b. Other problems associated with representation on the sphere (eg different phase error at different latitudes) are also absent when the spectral method is used.

One problem that might be thought to occur with the spectral method is the representation of a strong local phenomenon by global functions. To examine this, a five layer, T42 primitive equation model was initialised with a baroclinically unstable zonal flow plus a very small local perturbation. By day 7 (see Fig 3) two depressions were growing, and by day 11 these had become very strong mature systems and other waves had been initiated. Here we comment only on the undisturbed nature of the flow on the opposite side of the hemisphere. The non-local nature of the functions is clearly not a problem here.

The major drawback at the moment with the spectral method is that there is little experience in the inclusion of physical processes in these models. Now that the transform method is used there appears to be no reason why the physics should not be included in the grid-point domain just as for a finite difference model. Indeed this has been done in Canada (Daley) and Australia (Bourke) with no obvious problems. When representing the earth's topography with spherical harmonics, there will be spurious hills in oceanic regions. Depending on the smoothness of the topography to be represented and the truncation used, these hills should be only of the order of tens of meters high, and thus of little problem. The non-local nature of the representation again arises here in that, for instance, intense very local rainfall may have effect elsewhere. However, since all the grid point fields are derived from the spectral series, fields will tend to have less features on the smallest scale than in a finite difference model, and thus trigger less physics on the smallest scale. Experience in these possible problem areas should be gained in the next few years.

The spectral-transform method is clearly logically more complicated than a simple finite difference method. Thus a program for a spectral model tends to be rather more difficult to write and to modify. However, most finite difference models include many different kinds of averaging, so that the difference between the two in practice is not large.

Spectral analysis of model output and error can be most revealing of the behaviour of the atmosphere and its poor or good representation in the model. If one uses a spectral model, this information is routinely available. If one uses a finite difference model, it can easily be obtained. However, it will usually be obtained only in special situations at best.

Some were machine oriented, but tentative comments are now given. Experience has suggested that 60 bit word length is necessary in a spectral model if, for instance, surface pressure is to remain constant within tolerable bounds in a general circulation run. The new generation of computers will be faster only if the relevant variables are in main core and this core is likely to be little larger than on the present machines. Thus the smaller total storage and less shuffling in and out of core for the spectral method may be important. Efficient implementation of the spectral method will require the development of really fast Fourier transform routines for the new generation of computers. The Distributed Array Processor concept sounds tailor made for a finite difference model, and for such a machine the computation time superiority of the spectral method may be eroded.

Finally, we note that for some purposes the spectral-transform method may be considered little different from the finite difference methods presently used on latitude-longitude grids. These calculate zonal derivatives by consideration of the functional value at three zonal points (second order), five zonal points (fourth order) or all zonal points (pseudo-spectral). The spectral method was the functional value at all points. The finite difference models do a Fourier smoothing in the zonal direction poleward of a certain latitude. Spectral models do a full spectral analysis and smoothing based on total wavelength. These improvements in derivative and smoothing techniques explain the need for a grid of less resolution in the spectral transform method.

6. Conclusion

The spectral method must now be a strong competitor with the finite difference method for integrating the metoeorlogical equations. The main advantages of the two methods described in the previous section are summarised in Table 1. The estimated comparative computation times, in particular, suggest that the spectral method could well prove superior within the next decade.

Table 1 Advantages of the two methods Spectral figures given for T63 (T84) and finite difference for a second order scheme on a 1^{o} latitude-longitude grid, both 10 layers on the hemisphere.

	Spectral	Finite difference
Representation of medium and large scale	✓	
Representation of smallest scale		✓
Storage (words)	400k(800k)	2,800k
Timing of adiabatic calculation (CP minutes per day forecast) when using a semi-implicit time scheme for the spectral method and an explicit time scheme for the finite difference method	18 (56)	296
Timing of diabatic calculation, taking finite difference time as 1.	.3 (.5)	1
Pole problem and representation of the sphere	✓	
Experience in inclusion of physics		✓

Acknowledgement

The work described here was performed in collaboration with Dr Adrian Simmons and most of the points made arose in discussion with him.

APPENDIX Derivation of timing estimates.

1. Basic fact: T63 on $\frac{1}{3}$ hemisphere, 5 layers takes 55 secs/day

   Requirement: T63 on hemisphere, 10 layers

   Estimated factors: 10 2 $\Rightarrow$ 18 min/day

   Comments: A factor 10 is probably an over estimate, though 8 would be an underestimate.

2. Basic fact: T42 on hemisphere, 5 layers takes 140 secs/day

   Requirement: T84 on hemisphere, 10 layers

   Estimated factors: 6/tstep, 2 for tstep 2 $\Rightarrow$ 56 min/day

   Comments: A doubling of resolution implies more than 4 times the computation per timestep.

3. Basic fact: $5.6^\circ \times 3^\circ$ on hemisphere, 5 layers takes 94 secs/day

   Requirement: $1^\circ \times 1^\circ$ on hemisphere, 10 layers

   Estimated factors: 5.6 x 3/tstep, 5.6 for tstep 2 $\Rightarrow$ 296 min/day.

REFERENCES

Baer, F and Plutzman, G W	1961	'A procedure for numerical integration of the spectral vorticity equation', J.Met. 18 pp 393-401.
Bourke, W	1974	'A multi-level spectral model, I. Formulation and hemispheric integrations', Mon. Weath. Rev. 102, pp 687-701
Burridge, D	1975	
Eliassen, E, Machnauer, B and Rasmussen, E	1970	'On a numerical method for integration of the hydrodynamical equations with a spectral representation of the horizontal fields', Inst. of Ther. Met., Univ. of Copenhagen, Report No 2.
Gauntlett, D J and Leslie, L M	1974	'Further development on a six-level semi-implicit primitive equation model', Progress report No 4, The GARP Programme on Numerical Experimentation, pp 36-37.
Hoskins, B J and Simmons, A J	1975	'A multi-layer spectral model and the semi-implicit method', Q.J.R.Met.Soc. 101, pp 637-655.
Orszay, S A	1970	'Transform method for calculation of vector-coupled sums; application to the spectral form of the vorticity equation', J.Atmos.Sci. 27, pp 890-895.

Silberman, I 1954 'Planetary waves in the atmosphere', J.Met. 11, pp 27-34.

Simmons, A J and Hoskins, B J 1975 'A comparison of spectral and finite difference simulations of a growing baroclinic wave', Quart.J.R.Met.Soc. 101, pp 551-565.

Figure Legends

Fig 1. The variation with height of the longitude of the maximum of the dominant vorticity component (ζ_{11}^{8}) at days 4 and 6 for two spectral and two finite difference integrations.

Fig 2. Integrations of the barotropic vorticity equation with the pole of the coordinate system on the equator of the rotating sphere. The maps are stereographic projections based on the pole of the coordinate system (one hemisphere only). The rotational north and south poles are marked.

(a) Initialised with a superrotation - an exact solution

(b) Initialised with a Rossby-Haurwitz wave. At day 11.25 the whole pattern should have moved $\frac{1}{4}$ of the way round the sphere to the west relative to the rotation axis, and the picture should appear identical with that at day 0.

Fig 3. Polar stereographic maps of the northern hemisphere for a 5 layer, Triangular 42 primitive equation integration starting with a baroclinically unstable zonal flow and a very small local perturbation. Intense local eddies develop. The ability of the spectral model to handle local disturbances on this scale is apparent. Pressure contours are drawn every 4mb.

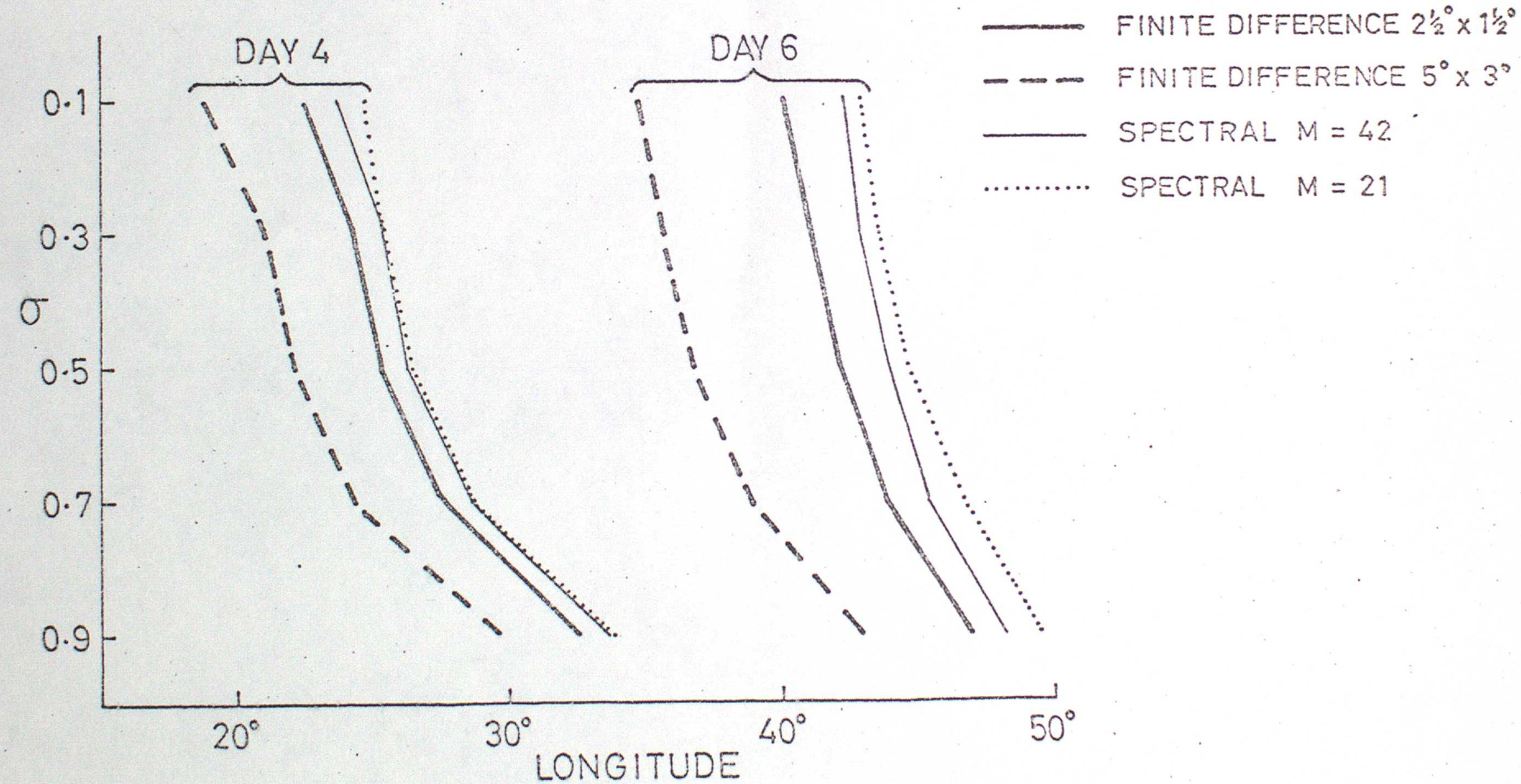
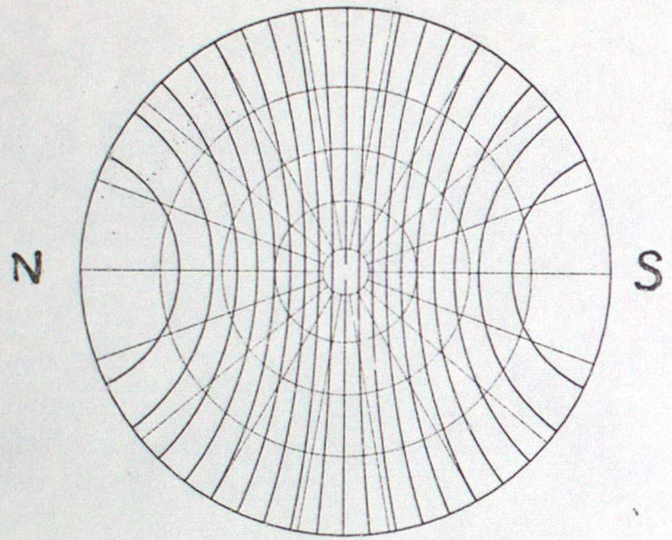


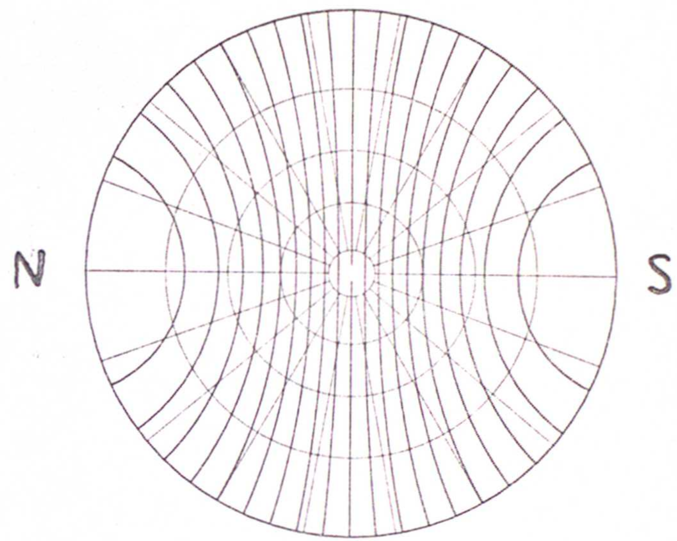
FIG. 1

The variation with height of the longitude of the maximum of the dominant vorticity component (ζ_M^{δ}) at days 4 and 6 for two spectral and two finite difference integrations.

2a

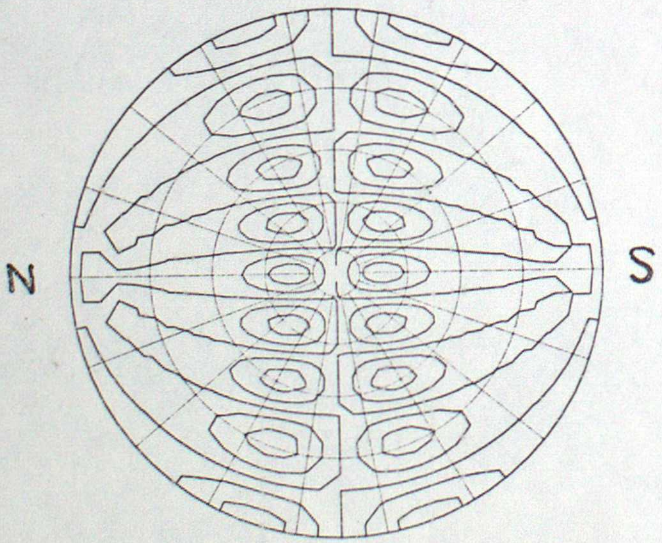


STREAM FUNCTION AT DAY 0.00

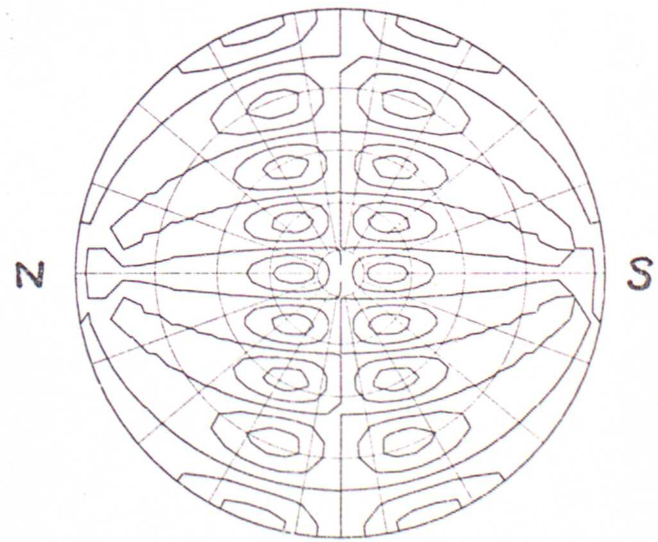


STREAM FUNCTION AT DAY 5.00

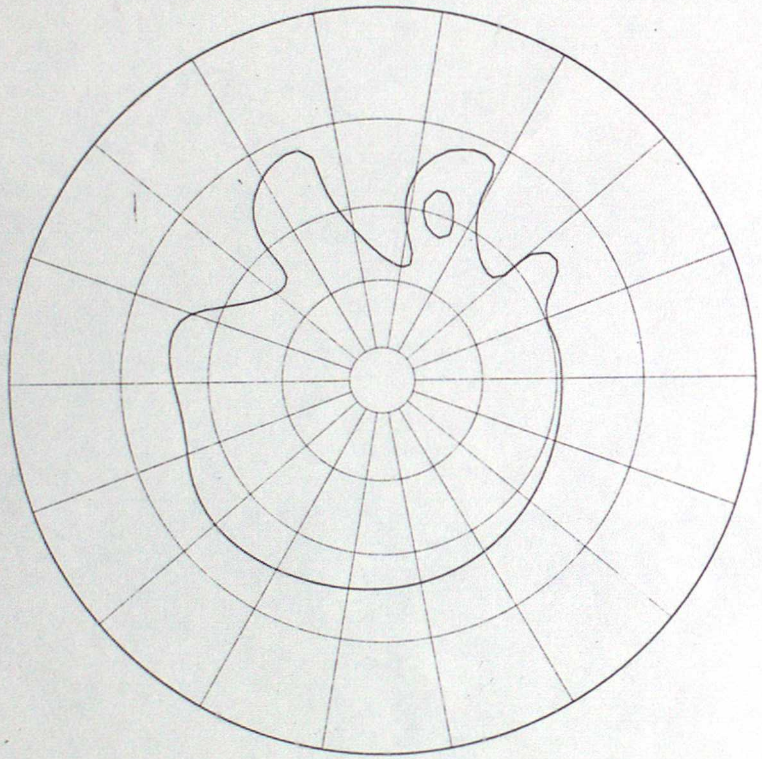
2b.



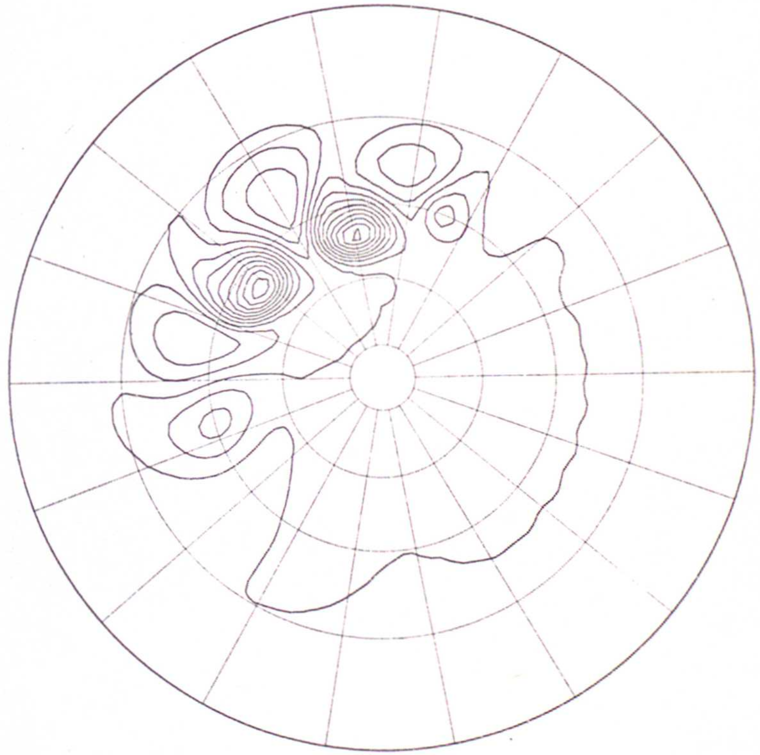
STREAM FUNCTION AT DAY 0.00



STREAM FUNCTION AT DAY 11.25



SURFACE PRESSURE AT DAY 7.00



SURFACE PRESSURE AT DAY 11.00

Fig. 3.