Numerical studies of baroclinic instability

A thesis presented for the degree of Doctor of Philosophy

Simon S. Elliott

Jesus College, University of Oxford

Trinity Term, 1995

Abstract

Numerical studies of baroclinic instability

Simon S. Elliott, Jesus College

Submitted for the degree of Doctor of Philosophy

Trinity Term, 1995

This thesis describes two studies of baroclinic instability in a rotating fluid annulus subject to differential heating. The first part is concerned with the development of a time dependent axisymmetric numerical model. The model was formulated using the control volume finite element method and was designed to be as flexible as possible both in terms of the range of problems to be studied and the techniques used to study them. Details of these techniques are presented together with a discussion of their limitations and possible refinements and extensions which could be made.

The second part of this thesis describes a numerical study of unstable normal mode perturbations which can develop on a prescribed mean state. The growth rate and structure of these modes are examined for various background states and the relevance of these results to laboratory measurements is discussed. Evidence is presented to suggest the possible presence of hitherto unobserved baroclinic weak waves in an internally heated annulus system.

Acknowledgements

In the first place I would like to thank my supervisor and friend, Peter L. Read, for his continuous support, prompting and interest in my work. I would also like to thank David Andrews for acting as co-supervisor and offering his thoughts on various aspects of my research.

My thanks are also due to the following:

David Anderson, Tom Mullin and Alan O'Neill for their advice on the progress of my work.
Stephen Mobbs, who helped in the conception of the new annulus model, and to his student, Ian Moore for our interesting discussion at Cambridge.
Raymond Hide, for taking such an active rôle in the Chapman Club.
Stephen Lewis, my academic advisor, for his helpful suggestions.
My fellow students for their company and tolerance.
Nigel Thomas, for running the Met. Office annulus model.
Jesus College, for providing excellent food, cheap accommodation and the chance to observe the glorious spectacle of undergraduate life at Oxford.
Logica UK Ltd, for allowng me leave to complete my thesis.

My work was funded by the NERC and latterly, by a Graduate Research Grant awarded by Jesus College.

I am also immensely grateful to my family and friends for their continued love and support, and in particular to my mother and to Claire.

Bring me my bow of burning gold!
Bring me my arrows of desire!
Bring me my spear! Oh clouds unfold!
Bring me my chariots of fire!
I will not cease from mental fight,
Nor shall my sword sleep in my hand
Till we have built Jerusalem
In England's green and pleasant land.

Jerusalem, William Blake

Contents

Chapter 1

Introduction.

1.1 Motivation.

1.1.1 A new annulus model.

The atmosphere and oceans of the Earth play a major rôle in the environmental make-up of the planet. Since they are both fluids, their dynamics are primarily influenced by the same forcing factors, namely planetary rotation, differential heating, and gravity. It is apparent then that a better understanding of the precise rôle played in the dynamics of the system by these factors is fundamentally important. The same mechanisms drive the flow in laboratory systems such as the differentially heated, rotating fluid annulus (see Hide, 1953, for example), except that here the system is simplified such that we can accurately parameterise the physics, and the boundary conditions are more precisely defined. This simplicity means that we can attempt to model the rotating annulus analytically and numerically, and use these results, together with laboratory measurements, in order to gain a greater understanding of the processes taking place. The relationship between the dynamics of the ocean/atmosphere and the rotating annulus system is illustrated in Figure 1.1 (for a detailed discussion see Hide, 1977).

The annulus system is, of course, in no sense an engineering model of the global

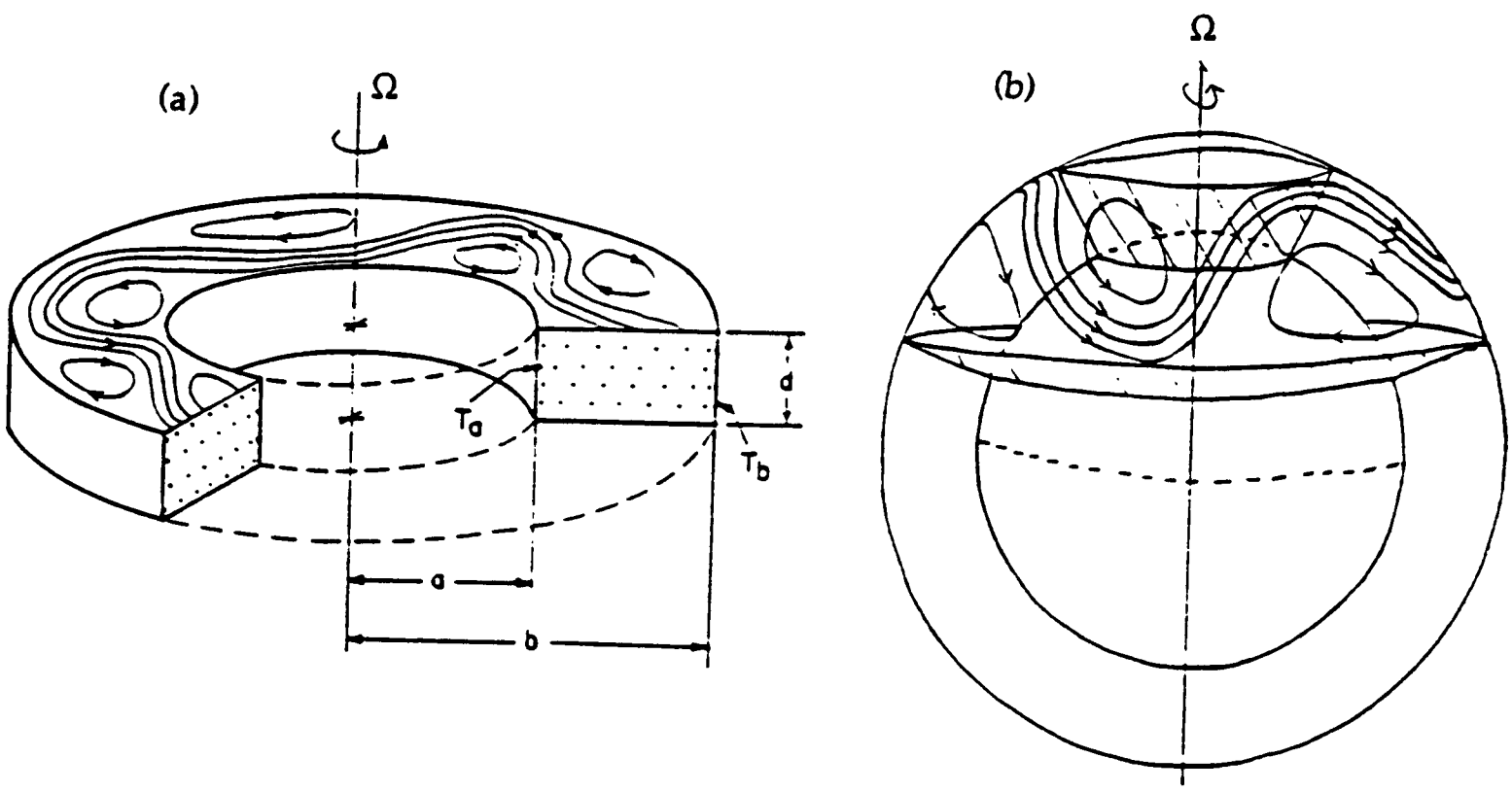


Figure 1.1: An illustration of the relationship between the ocean/atmosphere and the rotating annulus system. (a) Schematic diagram of the annulus, and (b) an equivalent schematic for ocean/atmosphere.

scale dynamics of the atmosphere or ocean. It takes no account of such processes as atmospheric chemistry or radiative transfer, but other important controlling mechanisms such as topography and the planetary vorticity gradient associated with the spherical curvature of the planet can be systematically included. It was the desire to include these and various other possible refinements into a numerical model which prompted the development of a new scheme, designed to offer greater flexibility and scope for adaptation than previous models (Williams, 1969, Farnell and Plumb, 1976, Quon, 1976 and James, Jonas and Farnell, 1981).

The decision to develop a new computer model led to additional benefits in that it gave the opportunity firstly to examine the suitability and applicability of existing but less familiar techniques (e.g. the control volume finite element method described in Chapter 2) to geophysical fluid dynamics, and secondly in that it provided an ideal testing ground from which to develop and apply some novel numerical techniques. It is important to distinguish clearly between these two, lest there be any confusion. In the first case, the

techniques have been widely tested and applied already, and their functionality is not the issue; it is more a question of their relevance to the task of modelling geophysical flows. In the second case, the new annulus model is being used as a proving ground for more "state of the art" and specially developed numerical techniques.

The development of a new model of the rotating annulus experiment provided us with a good opportunity to pursue several useful pieces of work, within the same general framework. Our most obvious aim was to be able to simulate a wider range of laboratory experiments than is possible with the existing model. These simulations could, then, not only be used to extend the parametric range over which we can obtain results, but also to search for regions in parameter space of particular interest, (perhaps where the transition between different flow type occurs), so that laboratory experiments can be concentrated there.

Our second, and rather broad ranging motivation, was to use the new annulus model as a test-bed for the development and implementation of a range of novel techniques and algorithms. We hoped to show, by example, how these methods (such as, finite elements control volumes *et c.*) could be included in a numerical model, and gain an appreciation of the benefits and detractions of the various schemes. If we were able to show that there was a definite advantage to a particular approach for modelling the annulus system, then this information might be applicable to the future development and enhancement of more complex numerical models, such as coupled ocean-atmosphere GCMs, where 'swapping in' a different time stepping scheme may not be such a tenable idea. It is, therefore, important to appreciate that although a large amount of effort may be spent in producing a code which does not fully work (as proved to be the case here), as long as the failing is understood, it is still possible to obtain valuable insight into the potential application of the numerical techniques adopted by more complex models.

1.1.2 The instability problem.

Planetary rotation and solar heating, which are the two main processes driving the atmospheric circulation, are regularly behaved phenomena with definite time scales associated with them. These time scales are fixed and well understood. However it is clear that the fluctuations in the weather exhibit no such regular behaviour, and those time scales which can be related to them are not the same as those of the forcing mechanisms. Often the weather seems to have no intrinsic periodicity about it at all.

Of course, it is possible to imagine a system forced in the same way as the atmosphere which satisfied all the necessary boundary conditions and yet was completely predictable. This clearly does not happen in the real world and this is due to the instability of 'the basic state' of the atmosphere to small wave-like disturbances. Such small disturbances are bound to occur in any large complex system and here may have orographic or anthropogenic origins, for example. In order to determine 'the basic state' it might seem sensible to take an average of a large number of observations of the system. This, however, can be misleading because the observed states will each consist of a combination of the ideal 'basic state' and the instabilities which have grown up on it. These instabilities can grow so as to alter the flow which gave rise to them and in general they will act, in accordance with Le Chatelier's principle[†], to move the system toward a more stable state. As a result, the mean measured state is often more stable than the actual basic state which should be used.

The answer to this difficulty is to side-step the problem of the uncertainty. Rather than trying to establish the ideal mean state and then examining its stability, we can simply prescribe various specific mean states and look to see which type of instability they are most susceptible to. In this way it is possible to determine which types of instability are attributable to which feature of the time mean state. The problem here is that we have tacitly assumed a linearity in the problem, where there isn't one in reality. As soon as

[†]Le Chatelier's principle, which is often applied to dynamical equilibria in physical chemistry, states that if a system at equilibrium is subjected to a disturbance or stress that changes any of the factors which determine the state of equilibrium, the system will react in such a way as to minimize that disturbance.

an unstable mode begins to grow, it will change the mean state, and this altered state may be unstable to a different instability. It does not necessarily follow, therefore, that the observed state of the atmosphere will correspond to one of the unstable modes even if the prescribed mean state is a very good approximation to the ideal one.

Early workers in this field were vitally important in showing that, by using simplified but reasonable mean states, they were able to grow unstable modes of the same type of spatial and temporal structure as those found in the atmosphere. It was in this spirit that the work described in the latter half of this thesis was undertaken. It was hoped that by using not only the quixotic mean states of previous investigations, but also actual steady states from laboratory and numerical models, it would be possible to study exactly those instabilities which determine the type of behaviour observed in the rotating annulus experiments. This process is not analytically tractable for a general background state and so a "bespoke" numerical scheme, INSTAB , was developed (see Chapters 4 and 5).

1.2 Background.

1.2.1 A general background to the annulus experiments.

The earliest experimental work on rotating fluids was purely qualitative. Fowlis (1964) describes the work of Vettin, Exner and Thomson and others from the latter part of the last century and the early part of this. Thomson appears to have been the first to suggest studying atmospheric dynamics by using differential heating in a rotating fluid. Exner and Vettin are thought to have observed axisymmetric and irregular flow respectively. This work has been reviewed by Fultz (1951).

Hide (1953, 1958) made great advances by carrying out experiments in a very controlled way on the differential heated rotating annulus system (see Figure 1.2). By carefully controlling the experimental conditions he was able to demonstrate that the stability

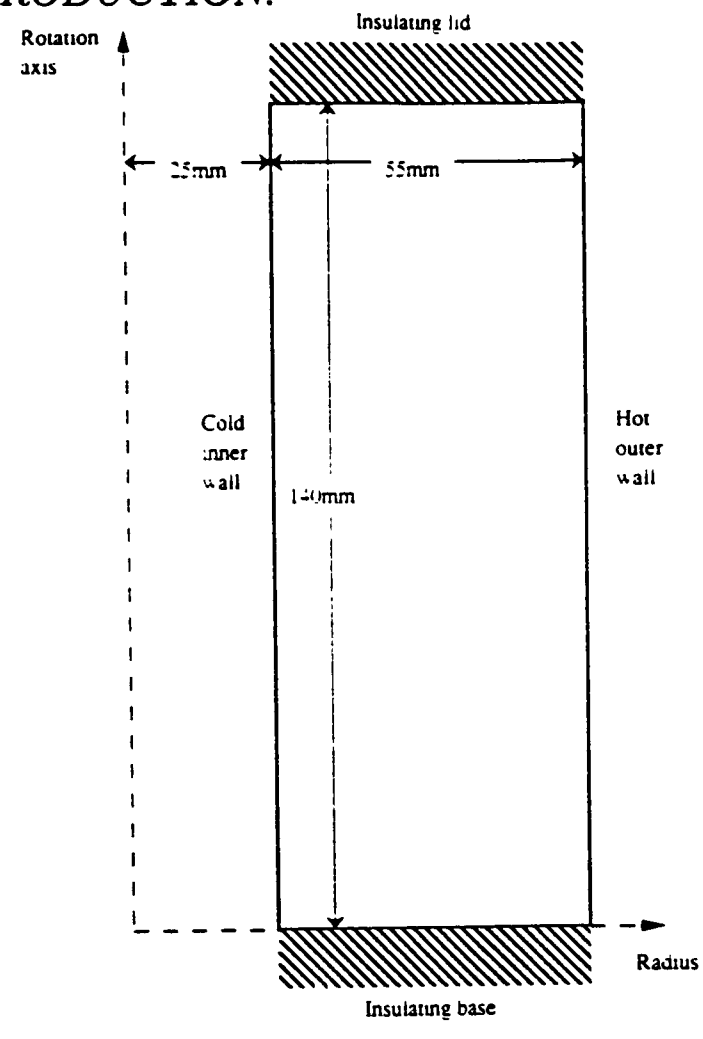


Figure 1.2: A schematic representation of a radius/height cross-section through the type of experimental set-up used by Hide (1953).

of the flow was critically dependent on a so-called "thermal Rossby number", Θ, where

$$\Theta = \frac{g\alpha\Delta T d}{\Omega^2(b-a)^2}. \tag{1.1}$$

Here, d is the depth, a and b are the inner and outer radii respectively and the other variables are as defined in Table 1.1. In its usual form the Rossby number gives a measure of the ratio of inertial accelerations to the Coriolis acceleration. Here the velocity scale is obtained from a linearization of the thermal wind relation to obtain the form for Θ given above. Hide showed that if Θ was greater than a critical value of 1.58 the flow was axisymmetric, otherwise regular waves (later shown to be baroclinic waves) were observed. The form given here for Θ is slightly different to that used by Hide, but if we assume a system with small curvature and density linearly dependent on temperature, the two are equivalent. Hide also found that at higher rotation rates two further types of flow existed; one irregular and one with vacillating waves.

It was shown by Davies (1956) that, while barotropic theory could not account for the onset of the regular waves seen by Hide, a simplified baroclinic theory gave reasonable agreement with experiment, suggesting that the waves were indeed baroclinic.

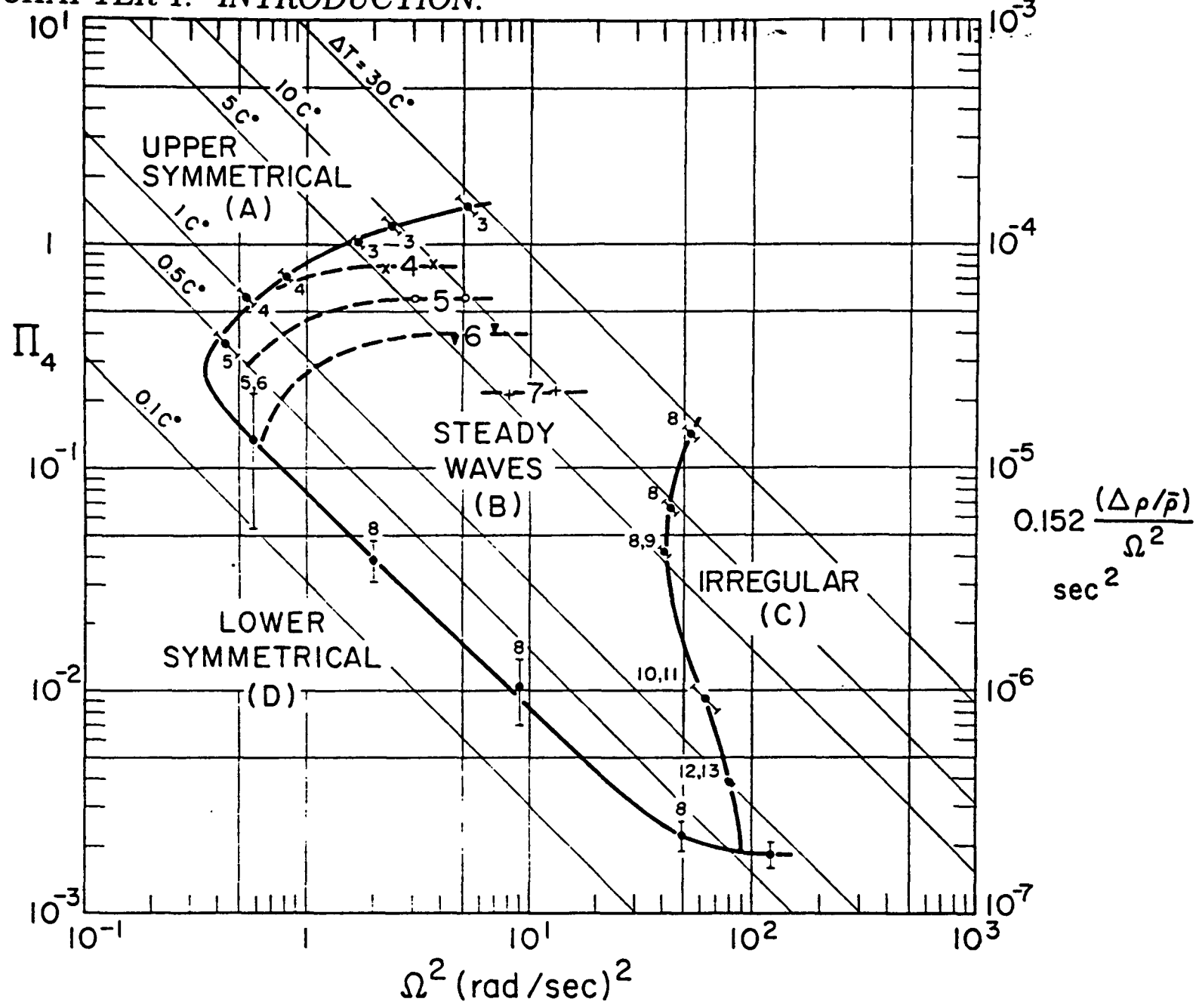


Figure 1.3: Regime diagram showing the various types of flow observed in the rotating annulus. The central anvil shaped region is the regular wave regime. For further details see Fowlis and Hide (1965), their Figure 3.

By systematically exploring parameter apace, Fowlis and Hide (1965) found that the flow in an annulus system was mainly dependent on four dimensionless parameters: Π_2, the meridional aspect ratio of the annulus, Π_4, the thermal Rossby number Θ, mentioned above, $\Pi_5 = 4\Omega^2(b-a)^5/\nu^2 d$ and Π_6 the Prandtl number[†]. The square of the ratio of the Coriolis acceleration to the viscous terms is called the Taylor number, $\tau = 4\Omega^2 L^4/\nu^2$, where L is some characteristic length scale. If we use $L^4 = (b-a)^5/d$ then Π_5 can be seen to be a Taylor number. Fowlis and Hide found that by plotting the type of flow they observed on a diagram with $\log \Pi_5$ as the x-axis and $\log \Pi_4$ as the y-axis, they were able to construct an anvil shaped region of parameter space in which regular waves were observed, as shown in Figure 1.3. They found steady axisymmetric flow both below and above this

[†]The Prandtl number is defined as the ratio of the viscous to thermal diffusivity.

region to the left of it (the lower and upper symmetrical regimes) and irregular flow to the right of it. By using a version of Eady's model of baroclinic instability, modified by others to include the effects of viscous boundary layers, they were able to obtain good agreement between experiment and theory for the onset of baroclinic waves, especially in the limit $\Pi_5 \to \infty$.

Hide continued to work in this field and Hide (1969) describes his extension of the theory of baroclinic theory to include the effect of Ekman layers on the upper and lower horizontal boundaries. He was able to deduce a stability condition on τ and found that his theoretical estimates of the critical Taylor number, τ_{crit} agreed well with experimental observations. His review paper (Hide, 1977) provides a summary of the various aspects of annulus work up to that point and includes a discussion of heat transport measurements, the effects of sloping bases and lids and details of the various types of wave vacillation which have been observed.

Since then various authors have made comparisons with processes observed in the atmosphere (for example Read, 1988a) and attention has moved toward a more detailed analysis of the observed baroclinic waves.

1.2.2 Experimental configuration.

The laboratory apparatus used to study the flows of interest consists of two co-axial cylinders mounted on a turntable, the annular gap between them being filled with fluid. The cylinders are mounted so that their common axis coincided with the rotation axis of the turntable. Different shape bases and lids have been used[†] both being typically made from thermally insulating perspex. The side walls are made of metal and cooled or heated to a constant temperature by running water at the chosen temperature through spiral channels in the wall. Although the inner wall is normally cooled in the laboratory, this choice is essentially arbitrary, since reversing the temperature difference would simply reverse the flow. Once steady wall temperatures have been established, the turntable is

[†]For details see Chapter 3.

Table 1.1: Typical values of physical parameters used in the rotating annulus experiment.

Physical parameter	Usual symbol	Typical value
Base height	z_B	0.0×10^{-3} m
Lid height	z_T	140.0×10^{-3} m
Inner radius	r_a	25.0×10^{-3} m
Outer radius	r_b	80.0×10^{-3} m
Rotation rate	Ω	0.0 – 5.0 rad sec^{-1}
Temperature difference	ΔT	4.0 deg C
Mean temperature	$\bar{T}$	19.0 deg C
Kinematic viscosity	ν	1.80×10^{-6} m^2 sec^{-1}
Thermal diffusivity	κ	1.0×10^{-6} m^2 sec^{-1}
Thermal expansion coefficient	α	3.03×10^{-4} deg C^{-1}
Mean density	$\bar{\rho}$	1.0×10^{3} kg m^{-3}
Gravity	g	9.81 m sec^{-2}

rotated anti-clockwise, so that with the applied temperature difference in the right sense, the whole system corresponds to the mid-latitude part of the Northern hemisphere of the Earth.

Typical values of the main physical parameters used in the laboratory are shown in Table 1.1 (data from Kaye and Laby, 1956).

The numerical model of the annulus described in Chapter 3, SARA , was designed to simulate exactly this type of system. Although the code was developed with a view to investigating more complex problems, this has proved infeasible in the time available and only the standard system described above has been investigated so far.

Velocity measurements in the laboratory.

Conceptually, the methods used to measure the fluid flow are very simple: the working fluid may be seeded with neutrally buoyant tracer particles which are tracked by a computer imaging system, which then reconstructs the flow. The actuality is rather more complicated and is only summarised briefly here.

Small (*i.e.* < 1 mm diameter) polystyrene beads are mixed in with the working fluid. The annulus is then illuminated by a thin horizontal sheet of light through one of five acrylic inserts in the side at different heights. A t.v. camera mounted above then follows the beads as they move through the light; the image generated can then either be viewed in real time, video taped or stored on a computer. This system is known as the Video Velocity Acquisition System or VVAS. By measuring the velocity in the radial and azimuthal directions in this way, it is possible to infer the vertical velocity from the non-divergence condition (see for example Tritton, 1989). The error in the vertical velocities calculated in this way is determined by the accuracy of VVAS and the thickness of the sheet of light, and is of the order of 1×10^{-4} m sec^{-1} (Rayer, 1992). This can be put into context by considering a typical value of vertical velocity which might be expected, for example as a result of Ekman pumping in the horizontal layers (see Greenspan, 1968, Gill, 1982 or similar.). This would be of the order of 1×10^{-5} m sec^{-1}, which is much smaller than the associated error and so very little weight can be given to empirically derived vertical velocity measurements[†]. The details of the mechanism for measuring the velocities are described in Hignett *et al.* (1985), and in some detail in Rayer (1992).

Temperature measurements in the laboratory.

The temperatures of the fluid in the annulus are typically measured using thermocouples. A ring of 32 copper-constantan thermocouples spaced equally in azimuth is positioned

[†]It is very difficult to determine the error associated with the horizontal velocity measurements because of the many sources of inaccuracy which could contribute. Rayer (1992) found a value of 0.9×10^{-4} m sec^{-1} gave reasonable results for his data analysis.

in the cavity at mid-height and mid-radius and their output continually recorded. There are two possible problems which might be anticipated using this type of measurements. Firstly, one might expect the presence of the thermocouples and the associated wires in the middle of the fluid to have a some perturbing effect on the flow. This problem has been investigated (Hignett *et al.*, 1985 for example) and found to be essentially unimportant. The second problem is that the mid-height, mid-radius location of the thermocouple ring might give temperature measurements atypical of the bulk flow. This can, however, be countered by moving the arrangement to a different height or constructing a similar ring at a different radius if this is thought to be necessary. The primary use of the temperature measurements is to analyse the motion of ordered structures (for example baroclinic waves), which can be readily identified by their signal in the temperature data, as they progress around in azimuth. If, perchance, such a feature is known only to exist between say the inner wall and mid-radius, one can imagine that the second problem described above would be important and a thermocouple ring at some smaller radius would prove useful.

1.2.3 Numerical modelling.

Together with laboratory and theoretical studies, numerical modelling makes up part of a broad range of options available for building an understanding of complex dynamical systems. In Figure 1.4, from Hide (1977), numerical studies of the annulus system would be part of box C, (Experiments on Basic Processes), together with the laboratory experiments and more theoretical work. As he shows, there is a symbiotic relationship with box D, (Atmospheric Science), the two supplying ideas and results to each other and aiding their mutual advancement. It was in this spirit that the first numerical studies of the rotating annulus work were undertaken.

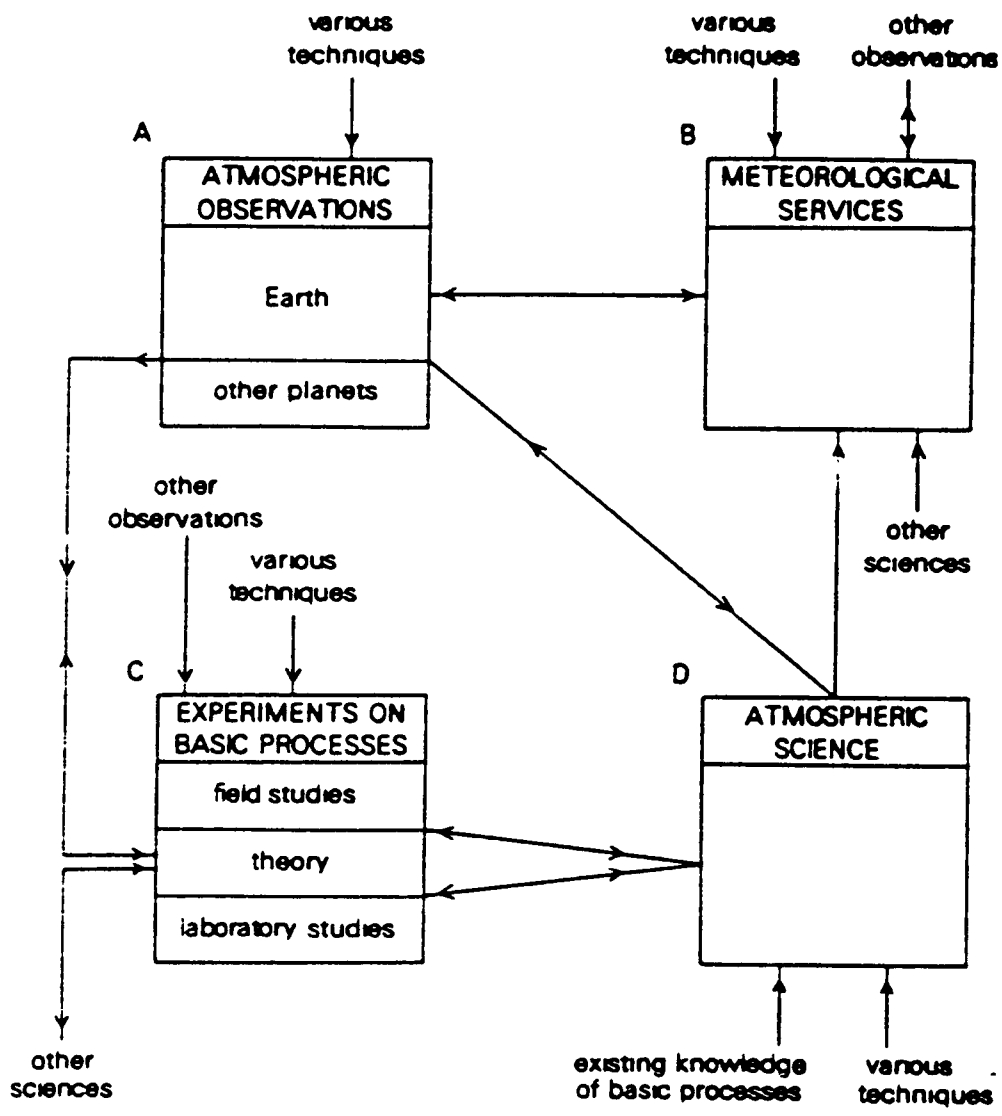


Figure 1.4: A schematic illustration of the relationship between the various fields making up meteorology, from Hide (1977).

Finite difference modelling

One of the earliest numerical models intended for accurate simulation of flow in the annulus was developed by Williams (1969). His technique involved integrating the three dimensional Navier-Stokes equations and a thermodynamic equation in cylindrical polar co-ordinates in a sector of the annulus. He did this by using a "time marching process" and solving a diagnostic Poisson equation for the pressure. The equations were cast into a second-order finite difference form in space and the time stepping was centre-differenced and so also second-order accurate. He used staggered but regularly spaced grids with various resolutions ranging from 14×14×14 to 60×34×34 points. This was a fair representation of the standard way of tackling incompressible flow problems at that time; his major advance was the novel use of spectral methods to solve the Poisson equation together with his own algorithm for the calculation of the necessary transforms. Williams found that less than 25% of the calculation time per time step was spent on the Poisson solver, meaning that for the 14×14×14 resolution version, his computer took only 2 seconds per time step. He noted some agreement between his results and the linearly unstable baroclinic waves of Eady's work.

Farnell and Plumb (1975) developed a model similar to that of Williams, but modified to allow several different boundary conditions, such as rigid or slippery (stress free) boundaries, to be used. Their model differed from that of Williams in four principal areas. Firstly, in order to get a better good agreement with laboratory measurements, they allowed ν, κ and α to vary in some simple way with temperature. Their second refinement was to use a stretched grid with fine resolution at the boundaries in order to resolve the thin boundary layers. However, the small grid spacing near the boundaries meant that in order for the scheme to remain linearly stable, a much smaller time step had to be used in accordance with the Courant-Friedrichs-Lewy condition. In order to improve this stability, they adopted the DuFort-Frankel scheme to represent the diffusion terms[†]. The

[†]The idea of the DuFort-Frankel scheme is to proceed as usual in the finite differencing of the diffusion terms, except that everywhere an operand appears, it is replaced by its average over the previous and the next time step. Although this scheme introduces a little temporal smoothing, it makes the diffusion term unconditionally stable (see, for example, Potter, 1973).

other major change they undertook was in the solution of the Poisson equation. They decided to use an iterative method (an optimised form of the alternating direction implicit method (Potter, 1973, for example) or ADI) which was not as efficient as the spectral method of Williams, but better able to cope with the complexities of a non-uniform grid.

Farnell and Plumb (1976) describe the extension of their earlier axisymmetric model to three dimensions. They found that because of the quantity of data which needed to be stored, they were only able to use 256 points in the meridional plane. The number of grid points in the azimuthal direction was constrained by their using a fast Fourier Transform to be 2^n, where $n \not> 6$ again because of the data storage problem. In practice this meant they used 16, 32 or 64 azimuthal grid points.

Quon (1976) describes a mixed model which uses finite differences in the radial and vertical directions, and a spectral representation in azimuth. This model only allowed wave-azimuthal mean flow interactions, and most of his runs include only the dominant wave. He found that cyclic reduction (or Gaussian elimination) was too complicated for the pressure and so he also used ADI.

Because of their use of a stretched grid, Farnell and Plumb (1976) were unable to use a spectral Poisson solver. Farnell (1980) showed how a spectral method could be developed for a stretched grid. For a regular grid the eigen-functions of the Laplacian operator are sines and cosines, and so by expanding the pressure field and the source function of the Poisson equation as a multiple Fourier series, the dimensionality can be reduced. However, when a stretched grid is used the eigen-functions are not analytic, but they can be determined numerically. The same process of expanding as a truncated series of eigen-functions can then be used. For problems where the grid is fixed, the eigen-functions only have to be determined at the start of the calculation and then can be used at each time step. Farnell demonstrated that this made his scheme faster and more accurate than optimized ADI when there were fewer than 40 grid points in any direction.

James, Jonas and Farnell (1981) apply this type of Poisson solver to a model which is otherwise essentially that of Farnell and Plumb. They were able to use a basic resolution of 16×16 points in the meridional plane by 64 in azimuth and could double the number

of points in any one direction for extra resolution. Their model was initialised by running a two-dimensional axisymmetric version up to steady state, and then introducing a small perturbation which could be either wave-like in the zonal direction, or a delta function at some point or a mixture of the two. Baroclinic instability then caused these waves to grow until the formation of a finite amplitude steady state. After detailed comparisons between their numerical results and results from laboratory experiments, they concluded that the numerical model reproduced the structure of the zonal flow, the dominant wave and the first harmonic reasonably well. However, other results such as the numerical and experimental regime diagrams were "not entirely similar" (*sic.*). They ascribed these differences to the limited resolution of their model, and the large amount of computer time needed for each run restricting the number of runs they could carry out.

Following on from this work, Hignett *et al.* (1985) used the same numerical model in a continuation of the comparisons between experimental measurements and numerical simulations. They also found good agreement in the zonal mean and main wave velocity fields, although this is hardly surprising. They were also able to reproduce numerically the main flow types and phenomena observed in the laboratory, such as wave number transitions and amplitude vacillation. There were, however, significant differences in the parametric location of some of these phenomena. They identified several areas where further work would be useful, several of which could be addressed by SARA , the annulus model described in Chapters 2 and 3. These included investigating the importance of model resolution, using various different numerical schemes and examining the behaviour at high rotation rate (greater than about 2 rad s^{-1}). They also suggested developing a hierarchy of simpler models to investigate how well these could simulate the full Navier-Stokes equations.

White (1986) provided a very detailed description of the finite difference schemes used by the model above (James, Jonas and Farnell, 1981) and in a companion paper Bell (1986) described its Poisson solver in similar detail. This model is in current use (some results are discussed in Rayer, 1992) and thanks to major advances in hardware, it now runs much faster and with grid resolutions up to $32 \times 32 \times 128$ (priv. comm., P. L. Read, 1993). White (1988) showed that an axisymmetric version of this model gave

values of steady-state heat flux which approached those from experiment as the meridional resolution was increased; he did, however, find a spurious eddy motion near the inner wall at an intermediate resolution and with zero rotation. This disappeared at higher and lower resolutions or if the rotation rate was non-zero.

Some typical numerical results.

For a particular apparatus and fluid, the flows which can be obtained are determined by the rotation rate Ω, and the applied temperature difference, ΔT[†]. For large values of $\Delta T/\Omega^2$, the flow is zonally symmetric, while at smaller values non-symmetric waves are obtained.

Figure 1.5 shows a typical axisymmetric flow which can be calculated from a numerical model. It can be seen that at any given radius the warm fluid is always above the cooler fluid; the system is therefore stably stratified. There is also a strong lateral temperature gradient in the body of the fluid, especially at mid-level. The zonal velocity profile shows two jets, one prograde in the upper half of the domain, and one retrograde in the lower. This flow looks rather similar to the analytical background state used by Bell (1989), indicating that his was a good approximation.

Figure 1.6 shows horizontal cross-sections through a typical regular wave number 3 flow. This flow can be thought of as a similar zonal flow to that shown in Figure 1.5 with an additional wave number 3 which is almost independent of height. It is interesting to note the agreement between the phase shift with height seen here and that predicted by the Eady model. This type of regular wave flow is termed a steady wave and its amplitude and shape do not change with time. It is also possible to obtain flows where the wave amplitude changes periodically in time (amplitude vacillation) or where the shape of the wave changes (shape vacillation). These types of behaviour can also be simulated by numerical modelling.

[†]This is equivalent to specifying a point in the regime diagram using different coordinates.

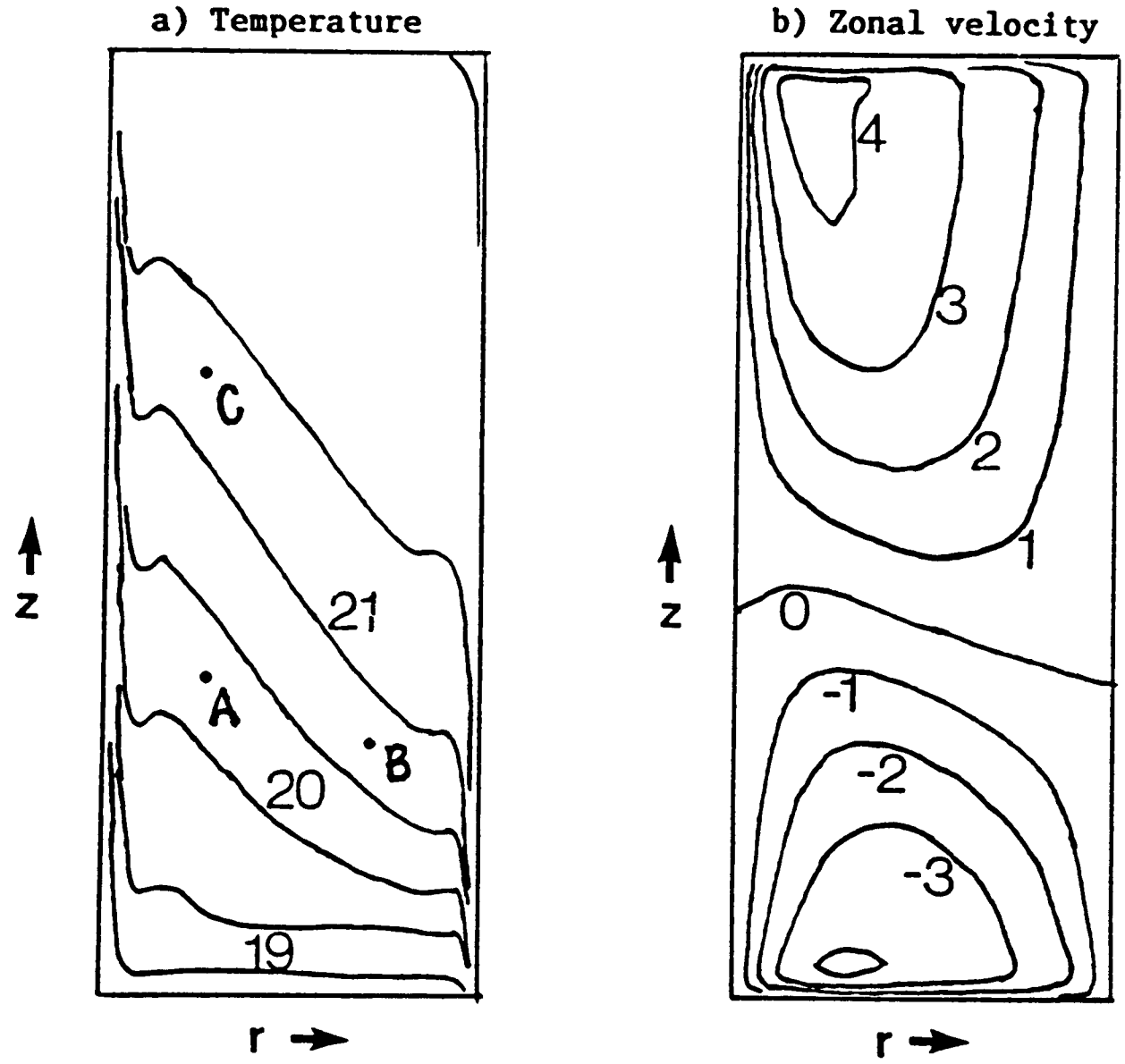


Figure 1.5: The diagram shows height/radius cross-sections of (a) temperature (deg. C) and (b) zonal velocity (mm s^{-1}) for a numerically determined axisymmetric flow. The numerical model (Hignett *et al.*, 1985) used a stretched 32 × 32 grid and the simulation was run for 4900 seconds to reach steady state. The inner and outer radii were 25 and 80 mm respectively and the height was 140 mm. The stable axisymmetric flow shown here was for a rotation rate of 4.999 rad s^{-1} and an applied temperature difference of 4.0 deg. C.

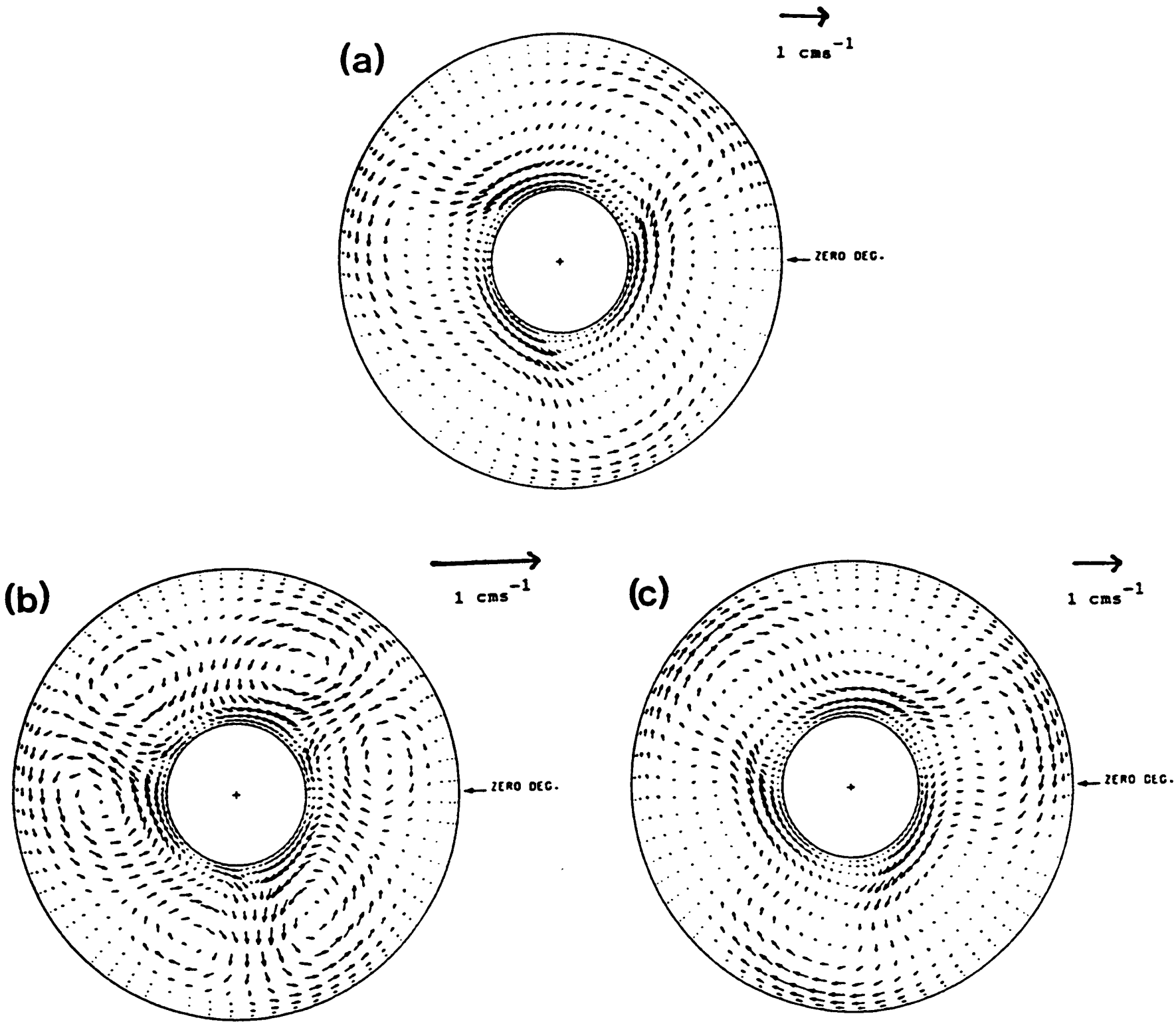


Figure 1.6: The diagram shows horizontal cross-sections of the horizontal velocity field at (a) 130.5 mm, (b) 57.3 mm and (c) 9.5 mm above the base for a numerically determined regular wave 3 flow. The numerical model (Hignett *et al.* , 1985) used stretched 32 × 32 × 64 azimuthal grid and the simulation was run for 1000 seconds from an initial state with a wave number 3 perturbation to a steady axisymmetric flow. The geometry was the same as in the previous figure, but this time the rotation rate was 1.0 rad s^{-1} and the applied temperature difference was again 4.0 deg. C.

Finite element modelling.

An alternative approach to numerical modelling of geophysical fluids is to adopt finite element methods[‡]. Finite elements methods have been widely used in engineering problems for some time; the commercial package ANSYS (Swanson Analysis Systems, 1989) for example, which has been available for over 20 years, can be used for problems involving heat transfer, fluid flow and stress analysis. Finite elements are used because of their suitability for dealing with complex boundary conditions as well as sources and sinks of heat or fluid *et c.*; these are exactly the type of considerations which are important when solving the Navier-Stokes equations, and so finite element modelling might seem ideally suited to the task.

Considering applications of finite elements to meteorology, Staniforth and Mitchell (1977) used a finite element model to integrate the barotropic free surface equations together with a semi-implicit time discretization. They used two-dimensional chapeau basis functions, which were defined as being unity at the node of interest and varying linearly to zero at each of the adjacent nodes. This gave them a convenient basis set on which to represent their functions. They used two versions of their model, one mixed-order and one fourth-order model, to generate a 48 hour forecast over North America and Canada. This they then juxtaposed with similar integrations from a spectral model, a second-order finite difference model and a mixed-order finite difference model. The results of the mixed-order finite element model were much the same as those from the mixed-order finite difference model, and considerably better than those from the second-order finite difference model. The fourth-order finite element model gave results which were more accurate than any of the others. They also found that, apart from a growing computational mode which could be filtered out, their model was stable for long time integrations and conserved both mass and energy well. Their scheme was very complicated, however, and even with the added efficiency of the semi-implicit time stepping scheme, the two finite element models were the slowest of the ones they tried, in so far as they could tell (they had some difficulty in obtaining a simple comparison because of the intricacies of the architecture of their com-

[‡]A basic description of finite element methods is given in Chapter 2.

puter.). By counting the number of operations required per time step, they estimated the finite element models to take about twice as long as their finite difference counterparts. In summary, they believed that the accuracy and flexibility which their method offered made it warrant further investigation.

They extended the model to allow variable resolution grids and in Staniforth and Mitchell (1978) they described several tests of such a scheme. They used various grids with very high resolution over a particular area of interest, varying smoothly or jumping abruptly to a more coarse resolution region around the outside. The change in computing time necessary to change to this type of variable grid model was found to be ~20%. They found that they could reproduce the forecast from a uniform high resolution model for a limited time in the high resolution part of their variables mesh and this meant that for a specific region they could reproduce the results of a high resolution model using less computer time. This highlighted the flexibility of finite element methods for problems where a non-uniform grid (such as the stretched grid in the later annulus models) is used.

The examples given here all rely on the Galerkin method (see for example Burden and Fairies, 1985) to make the error in the expansion of the dependent variables orthogonal to the basis set. A more physical technique for applying finite elements to Navier-Stokes problems is presented in Baliga and Patankar (1983). This paper and several related ones are described in some detail in Chapter 2, and only a brief outline of their method is given here. Their scheme was designed for studying *steady*[†] two-dimensional incompressible fluid flow and heat transfer in advection-diffusion problems. The salient features of their method were (i) the use of mixed interpolation for the pressure and velocity components in order to avoid the checkerboard-type instability; (ii) their use of an iterative solver which solved sequentially for the pressure and velocity fields within each iteration; and (iii) their use of the control volume finite element method. In order to insure stability they used a mixture of bi-linear and linear-exponential basis functions where the exponential part was aligned with the mean flow at that point. They used this method to solve several test problems successfully (Baliga, Phan and Patankar, 1983), and in the light of this success

[†]Our attempts to develop a time-dependent version are described in Chapter 2.

we decided to use a control volume finite element formulation in SARA (Simulation and Analysis of the Rotating Annulus), the new annulus model.

Baliga and Patankar used a pressure correction method to ensure that the velocity field was non-divergent and that all the relevant boundary conditions were satisfied. A similar scheme to that of Baliga and Patankar was adopted by Burns and Wilkes (1987) for their fluid code, FLOW3D, which has been successfully used to determine the steady state flow in a symmetrically expanding duct. Moore (1992) also used a pressure correction scheme in his model of the flow of wind laden with snow around a building on legs. Solving for pressure proved the most difficult part in developing SARA and this problem is discussed in detail in Chapter 3.

1.2.4 Investigation of the baroclinic instability.

Until the middle of this century, there had been very little attention given to the rôle of the dynamical instabilities which can develop in the atmospheric circulation. Bjerknes (1937) realised that if the atmosphere was forced to be zonally symmetric, it would be in a different state to that which was observed. He also showed that such a symmetric flow would be unstable to small perturbations and so postulated that the cyclones and anticyclones seen in the circulation were the manifestations of these instabilities.

A mathematical model of the fundamental processes driving these waves was presented in the famous paper by Eady (1949). He made several simplifying assumptions which enabled him to find an analytical solution to the equations describing the flow. These can be summarised as:

i The working fluid was non-divergent and Boussinesq (i.e. density variations could be ignored except where they give rise to buoyancy forces).

ii Viscous effects were negligible implying that the viscosity, $\nu = 0$.

iii Heat transfer by thermal dissipation was much less than that by convection so that the thermal diffusivity, $\kappa = 0$ was a valid approximation.

iv The background vertical temperature gradient (and hence the Brunt-Väisälä frequency) was positive and independent of position.

v The axisymmetric background state was assumed to be a purely zonal flow with uniform shear. This implied a constant horizontal temperature gradient consistent with the thermal wind relation (Pedlosky, 1987, for example).

vi The Rossby number (the ratio of Coriolis to inertial accelerations) was assumed to be small so that the quasi-geostrophic approximation was applicable.

vii Only small amplitude waves were considered so that linear perturbation theory was applicable.

Applying all these conditions, together with the constraint of no vertical flow at the top and bottom boundaries, Eady was able to derive an equation for the complex phase speed of the resulting waves. Despite the large number of approximations and simplifications which went into his model, the results agreed very well with observations in the atmosphere as Eady himself noted and also could be applied with equal validity to the rotating annulus system (Fowlis and Hide, 1965).

Eady's treatment of the instability problem ignored the β-effect, that is the effect of the curvature of the Earth's surface, on the dynamics. Charney (1947) used a model akin to Eady's except in that it had only a lower boundary, that the β-effect was included and the mean density was allowed to vary with height. These inclusions meant that the horizontal gradient of the background potential vorticity was no longer zero, which in turn meant that there was no longer an analytical solution to the equation governing the phase speed of the disturbances. Numerical solutions have however been obtained for Charney's model (see Pedlosky, 1987 for example), and remarkably good agreement has been obtained with Eady's model for the maximum growth rate, and structure of the unstable modes.

An example of the usefulness of this type of linear instability theory was amply demonstrated by Charney and Drazin (1961), who considered the vertical structure of forced

planetary waves; a subject of considerable importance to the conjectured coupling between events in the middle atmosphere to those in the troposphere. In their pioneering paper they deduced, by using simple profiles of the zonal mean state, their famous criterion governing whether vertical propagation can occur. This claimed that waves are trapped to their source unless the background flow is eastward (prograde) with respect to the phase speed of the waves, and not too strong.

Pedlosky (1987) (his Chapter 7) gives a very thorough discussion of instability theory in general and also describes the Eady and Charney problems in some detail.

Prompted by the structure of the zonal mean flow giving rise to the transition between the regular wave and axisymmetric regimes in the rotating annulus experiments, Bell (1989) used analytical and numerical results to examine the stability of various flows similar to those observed. He investigated the sensitivity of the transition to regular wave flow (marked by neutral wave modes) to various features in the zonal flow. Bell was restricted, however, in that the zonal mean states on which he looked for instabilities had to be analytic, for example the sinusoidal internal baroclinic jet which he used as his first example (his Chapter 2, Eqn. 2.2.20).

The concept behind INSTAB was to develop a piece of software which could use *any* background state and solve the linear instability problem numerically to determine the growth rates of any unstable modes. In particular, it was hoped that output from the two dimensional Met. Office model and from SARA could be used, enabling a stability analysis to be carried out on the computationally determined axisymmetric flows directly. If these models were run so as to simulate a region in parameter space close to the empirically determined transition to the regular wave regime, then we can expect INSTAB to predict the structure of the fastest growing (and therefore perhaps dominant) mode. The development of the instability problem in its general form is discussed in Chapter 4 and some results and their interpretation are presented in Chapter 5.

1.3 A summary of the thesis.

The first chapter of this thesis begins by outlining the motivation for the two numerical studies which have been undertaken. It then gives a summary of some of the relevant literature. The next two chapters are concerned with the development of a new annulus model; Chapter 2 describes the control volume finite element method itself and its application to our model, and Chapter 3 details the other important numerical techniques used as well as showing how well the model works and in what ways it could be improved. In Chapter 4 we develop a model for the linear instability problem in general terms for an annular geometry. The application of this model is discussed in Chapter 5 and some results are presented. Chapter 6 summarizes the results of the earlier chapters and examines possible future extensions to the work described here.

Chapter 2

The control volume finite element method.

2.1 Numerical modelling.

This chapter is concerned with the numerical methodology used for SARA . Before going on to describe this in detail, it is useful to briefly summarise the nature of computational analysis.

Despite the obvious usefulness of analytical methods, there is a wide range of problems for which they can offer no solution. For example, the classical one- and two-body problems can be solved analytically. For all but the simplest cases, however, few and many body problems can only be tackled numerically.

The finite nature of the computer means that continuous problems cannot be solved directly. Instead, the continuous variables are replaced by discrete approximations. Space dimensions are replaced by a grid of points and time is represented by a series of synoptic snap-shots. Although the computer is restricted to a finite number of unknowns, this number can be very large so that a high accuracy for solutions can be achieved. For systems such as fluids, for example, it would normally be impossible to follow each con-

stituent particle. Instead, one could either (a) use a statistical approach and solve for the macroscopic variables, or (b) consider the fluid as being made up of a number of "parcels" and let these follow the same laws as apply to the fluid (a Lagrangian scheme).

It is often possible to solve a problem using numerical methods where otherwise no solution is possible. Great care has to be taken, however, to ensure that the solution obtained is the right one. Such methods are prone to rounding errors due to the finite precision of computers, and to numerical instabilities which can arise as a result of overly simple spatiotemporal discretization (see, for example, Potter, 1973 or Smith, 1989).

In summary, algebraic analysis often relies on linearity, on symmetry, on a small number of variables and is effective for describing continuous media. Where computational methods are employed, the physical system described must be discrete and finite. The remainder of this chapter describes how the control volume finite element method can be used to discretize the equations describing the annulus problem in an heuristic and effective way.

2.2 Model design criteria.

The first step in the development of SARA was to decide what the model should be able to do and what numerical properties it should have. These choices are not only determined by the type of experiments which might be conducted using SARA and the kinds of diagnostics which one might want to examine; but also by the efficiency of the available computing facilities. Listed below are those criteria which were considered to be the most important.

- The code needed to be portable. This was particularly important because although the code was developed in Oxford, it was originally expected to be run in Leeds on a faster computer when completed. Since the start of the project, however, fast computing facilities have become available locally and although there is no longer any need to run SARA in Leeds, its portability is still as important on a local level. This portability was ensured by the use of standard FORTRAN and the inclusion

of detailed documentation.

- Library software was not used for two principal reasons. Firstly, to ensure absolute portability, and secondly because in order to be certain of the functioning of complex and original algorithms, it was very useful to have developed *all* the code "in house".

- The code was developed in such a way that alternative sections could be swapped in and out without difficulty. This modular structure meant that a simple code could be developed initially and then improved modules could be incorporated into it once they had been tested individually. It was also possible to compare the efficiency of different schemes by swapping between them and comparing the results.

- In order to give physically realistic results the code needed to be conservative. This was achieved by differencing the system equations once they have been written in a flux form.

- The numerics were made to be accurate to second order, although due to the modular form of the code, higher order algorithms could be substituted (or "swapped in") at a later stage.

- An ideal model would be very flexible in terms of its geometry, so that various effects such as a stress-free top surface, radial barriers, sloping tops and bottoms, undulating bases, multiply connected systems, open cylinders and non-cylindrical geometries could be investigated.

- It should be possible to heat the system at the walls or internally.

- It was important to make the boundary conditions as flexible as possible. Both non-slip and stress-free boundaries should be available, as well as partially insulating or conducting walls. There should also be means to include sources and sinks of the fluid in the boundaries.

- Passive tracer transport should be included in order to bring to light selected aspects of the flow.

- It was important to use an adaptable mesh which had a higher resolution in regions where the length scale of the motion becomes small, in order to resolve the fine scale motions which occur, for example, in the boundary layers.

- The code was time dependent, but steady state flows could be established by letting the model run to equilibrium.

Although not all of these ideas have been included in SARA so far, most have (see Chapter 3), and there is scope for the adoption of the others.

These considerations led us to base SARA 's design on three main numerical concepts. These were:

1. The use of a control volume finite element method for the main coding. This method was conservative, second order accurate and had the mesh flexibility required by the above criteria. The method was also known to be very stable and computationally efficient. Further details of this scheme are presented in Baliga and Patankar (1983), for example, and it is also described in detail in Section 2.4 of this thesis.

2. The Poisson solver for determining the pressure[†] is usually the rate determining step in fluid flow problems. In order to make this part of the calculation as efficient as possible, an optimized fast matrix solver was used. A description of how this works together with the reasons for the choice of the particular solver used in SARA (a conjugate gradient solver), are given in Appendix A.

3. The code was made explicit in velocity in order to simplify the coding. The pressure correction method, described later in this chapter, would require a semi-implicit scheme (Patankar, 1980) for the velocity if it were implemented.

[†]Whether one uses a vorticity/stream function formulation or, as in SARA , a pressure/velocity formulation, there is necessarily a diagnostic Poisson equation to be solved at each time step.

2.3 Finite differencing and finite elements.

2.3.1 Finite differencing.

In order to represent a physical variable in a numerical scheme, it is necessary to discretize and store the value of the variable on a grid of points. This applies equally in space, where a mesh with either one, two or three is used, and in time. The larger the number of grid points in a given region, the greater the spatial or temporal resolution, but the larger the storage demands on the computer system. A further consideration is that the run time of the code at each time step is a function of the number of spatial grid points (it is approximately proportional to N^3, where N is the number of spatial grid points), and when run time is expensive this imposes an additional constraint on resolution. The value of each variable is calculated at every grid point each time step, by solving the equations which describe the evolution of the system. These equations will in general involve the derivatives of the variable in one or more directions and these are readily calculated from the values on the grid. For example, consider the x derivatives of a function φ at the n^{th} grid point along the x direction. Then using finite differences and a uniform grid

$$\frac{\partial \varphi}{\partial x} = \frac{\varphi_{n+1} - \varphi_{n-1}}{2\Delta x} \tag{2.1}$$

or, for the second derivative

$$\frac{\partial^2 \varphi}{\partial x^2} = \frac{\varphi_{n+1} - 2\varphi_n + \varphi_{n-1}}{(\Delta x)^2} \tag{2.2}$$

where the subscripts refer to the value at that grid point and Δx is the (constant) grid spacing. This kind of scheme is commonly used (see for example Potter, 1973) and is said to be second order accurate (or simply second order). This is because comparison with a Taylor expansion about the point at which the derivative is being calculated shows an error in the approximation proportional to $(\Delta x)^2$. Because Δx is small (i.e. $\ll 1$) this means that the finite difference approximation gets more accurate as the grid resolution is increased. Higher order derivatives can be calculated in a similar way and finite difference forms of the vector calculus operators *div, grad, curl* and ∇^2 can also be constructed .

2.3.2 Finite elements.

Using the method described above, systems of differential equations can be cast into a difference form and suitable algorithms formulated for their solution. A second technique, which is in principle not dissimilar to finite differencing, is the use of finite elements. This is a fairly common method in some areas of science (for example, stress analysis and thermal modelling problems in engineering), but remains less commonly used for modelling geophysical systems. One important advantage of the finite element method over finite differencing is its ease of applicability to complex boundary conditions. The task of modelling fluid flow in the standard right cylindrical annulus experiment introduces rather straightforward boundary conditions, but if different topography were used or fluid source/sinks in the walls, the problem would become rather more complex. Gradient boundary conditions are inconvenient to apply using finite differences because they require "ghost cells" outside the physical domain. In itself this does not pose any particular problems, however, the function's value on the boundary is not stored explicitly, and must be inferred by interpolation. Finite element methods are much better suited to gradient boundary conditions and source/sink terms. This is particularly true for the control volume finite element method used in SARA .

As with finite differencing, the values of each variable are held at discrete points, called nodes. The regions formed by the lines joining the nodes are the elements from which the method takes its name (see Figure 2.1). The next step is to use piecewise polynomials of fixed degree to approximate each variable over the whole region. These must be chosen such that the approximation is continuous across elements. With triangular nodes it is normal to use polynomials linear in both directions,

$$\varphi(r,z)_{element} = Ar + Bz + C, \tag{2.3}$$

or for rectangular nodes bilinear functions can be used,

$$\varphi(r,z)_{element} = Ar + Bz + Crz + D. \tag{2.4}$$

The constants A, B, C and D can be uniquely determined from the three/four nodal values of the elemental triangle/rectangle. These shape functions are determined at each

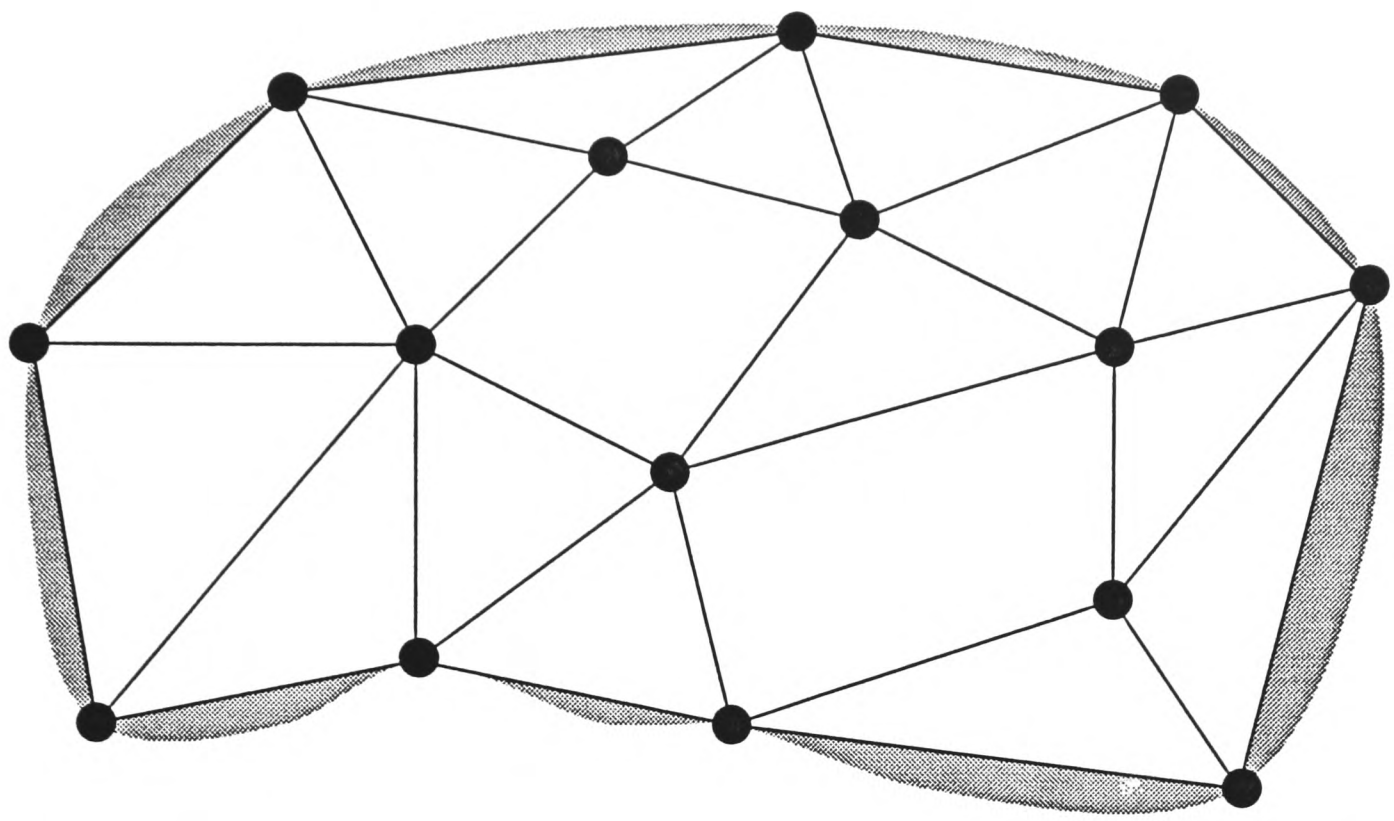


Figure 2.1: The diagram shows triangular and square elements filling up space and the nodes at their vertices.

time step for every element and for each of the variables stored by the model, namely the radial, zonal and vertical velocity components, temperature and pressure.

In the case of a three dimensional model, it is easiest to consider a rectangular cartesian coordinate system, although the same ideas apply in cylindrical polar coordinates away from the vertical axis. If space is first broken up into rhomboidal elements, each with six faces and eight vertices, two sets of approximating polynomials suggest themselves. The first is a tri-linear function,

$$\varphi(x,y,z)_{\text{element}} = Ax + By + Cz + Dxy + Exz + Fyz + Gxyz + H, \qquad (2.5)$$

where the constants are determined by the eight vertices. Alternatively, if the rhomboid is further split into five tetrahedra, then functions linear in each direction can be used

$$\varphi(x,y,z)_{\text{element}} = Ax + By + Cz + D \qquad (2.6)$$

in a similar way. For all of these interpolation polynomials, the amount of computer time taken to calculate the constants is approximately proportional to the square of the number of nodes. If we compare triangular with rectangular elements, and observe that the rectangles are each made up of two triangles, then

For triangular elements: $2 \times 3^2 = 18$ time units

For rectangular elements: $4^2 = 16$ time units

or, in three dimensions, for tetrahedral and rhomboidal elements

For tetrahedral elements: $5 \times 4^2 = 80$ time units

For rhomboidal elements: $8^2 = 64$ time units

where there are five tetrahedra making up each rhomboid. It is clear then that using fewer elements, each with more nodes, is computationally more efficient. This is not the only factor, however, which needs to be considered when choosing which shape elements should be used. The more complicated elements have a preferred direction associated with their shape functions and this spoils the symmetry and isotropy of the system. In addition, the shape functions associated with the simpler elements are much easier from which to obtain gradient information, as described below. In the two dimensional axisymmetric model, SARA , we use triangular elements, and in a three dimensional version of the code, tetrahedral elements would be used.

Other kinds of shape functions can be used such as locally aligned, linear-exponential functions (see Section 2.5). Although these have some advantages, the bi-linear shape functions are the simplest to work with and give fitted functions which are continuous across all the elements. An important advantage of finite elements becomes apparent when we consider the calculation of the derivatives of a function. Since an analytical form is known for the function at all points, the derivative can simply be obtained from the shape function. For example, within the element considered previously,

$$\frac{\partial \varphi}{\partial r} = A \tag{2.7}$$

and

$$\frac{\partial \varphi}{\partial z} = B. \tag{2.8}$$

Variational methods.

The conventional method used to obtain the discretized equations when using finite elements is based on the calculus of variations. To understand this method fully, the reader should have a good knowledge of this aspect of calculus and is referred to a standard text on the subject (for example Lanczós, 1966). Here, however, a general understanding of the main features of this type of method is all that is required as, in the event, a variational formulation was not adopted for the purposes of constructing SARA . The reason for this is discussed below.

The calculus of variations shows that solving certain differential equations is equivalent to the minimization of a related integral, known as the *functional.* This equivalence is the known as the variational principle. If the functional is minimized with respect to the nodal values of the dependent variable (*i.e.* velocity, temperature or pressure), the resulting conditions give the required discretization equations. In addition to its algebraic and conceptual complexity, the main drawback to this formulation is its limited applicability, since a variational principle can not always be found for the differential equations of interest.

If the set of n basis functions being used are the linearly independent set $\phi_1, \phi_2, \ldots, \phi_n$ then these must all satisfy the boundary conditions individually. The functional to be minimized is then

$$I[\phi] = I\left[\sum_{i=1}^{n} c_i \phi_i\right]. \tag{2.9}$$

For a minimum of the functional to occur it is necessary for

$$\frac{\partial I}{\partial c_j} = 0, \tag{2.10}$$

for each $j = 1, 2, \ldots, n$. This minimization is accomplished using either the Rayleigh-Ritz method or the Galerkin method depending on the specific problem being considered. For some cases they are equivalent (see Burden and Faires, 1985). Both these methods involve dealing with large amounts of numerical algebra which is slow and prone to inaccuracy.

2.4 The control volume finite element method.

SARA uses a finite element formulation, together with the control volume technique. This was originally developed for use with problems arising in fluid dynamics and heat transfer. Control volumes provide an alternative way of using finite elements without needing to involve the use of the variational methods described above. The use of control volumes with finite differencing was described in great detail by Patankar (1980), and details of the extension of this method to finite element modelling (hereafter referred to as CVFEM) can be found in Baliga and Patankar (1983). Although the work presented there was concerned with steady state problems, it was hoped to use the same general methodology for the study of the time dependent flow in the rotating annulus.

It is easiest to describe the use of control volumes in a two dimensional rectangular geometry. The exact application in SARA is slightly different because of the cylindrical geometry, which introduces metric terms. The principal, however, is exactly the same. Each point in the discretized space (see the description of the grid structure and stretching in Chapter 3), with the exception of those on the boundaries of the calculation domain, is at the vertex of several adjacent elements (see Figure 2.2a). The corners of the control volume associated with this node are the mid points of the lines joining it to each of its immediate neighbours, and the centroids of the elements around it. This method of constructing the control volumes was introduced by Winslow (1966). For the case of the particular grid used in SARA this gives rise to twelve-sided irregular control volumes. Figure 2.2b illustrates one such control volume resulting from the use of a regular grid; the twelve sides of the control volume appear to be only six as the regular grid makes each pair of sides, bc-cd, de-ef, *et c.* run into each other. This method for constructing polygonal control volumes can be used with any choice of finite elements, whether triangles or rectangles or a mixture of the two. This system has several useful features; the control volumes do not overlap, their boundaries are never along elemental boundaries and, collectively, they fill the computational domain completely. This makes conservation of momentum and mass easier to ensure, which is very important for the results to be physically meaningful.

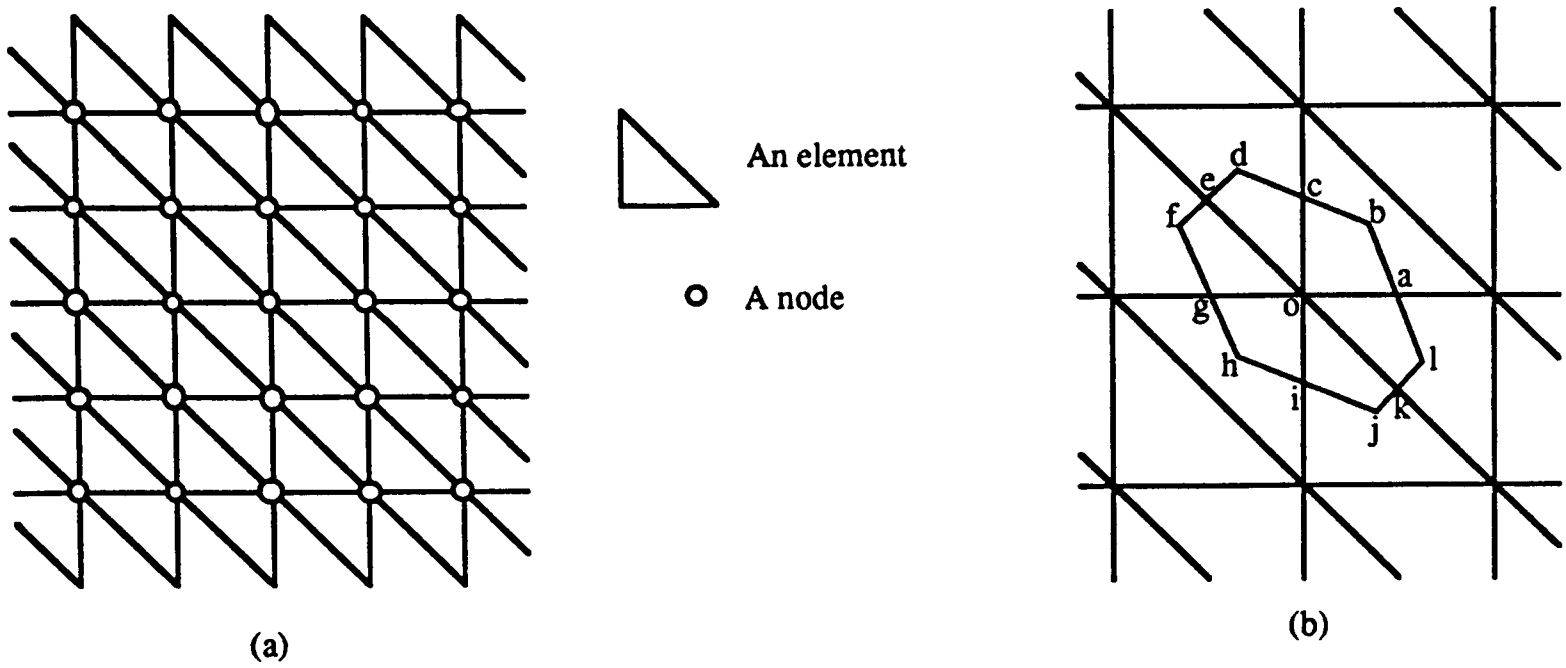


Figure 2.2: Details of the nodes and elements used in SARA and their associated control volumes: (a) The triangular elements shown here on a regular grid for simplicity, and (b) the control volume for node O. The six sides are an artifact of the regular grid, in general the control volumes will be twelve sided.

The computational procedure for the advection and diffusion of any scalar variable, such as temperature, is dependent upon forming integral equations in terms of fluxes. The conservation equation for a scalar field, χ, can be written

$$\frac{\partial \chi}{\partial t} + \mathbf{u} \cdot \nabla \chi = \kappa \nabla^2 \chi + S_\chi, \tag{2.11}$$

where $\mathbf{u}$ is the two dimensional velocity field, t is time, κ is the diffusion coefficient for χ and S_χ is the source term for the creation or annihilation of χ at the node under consideration. If we define the flux vector $\mathbf{J}$ as

$$\mathbf{J} = \mathbf{u}\chi - \kappa(\nabla \chi), \tag{2.12}$$

then

$$\nabla \cdot \mathbf{J} = \mathbf{u} \cdot \nabla \chi + \chi(\nabla \cdot \mathbf{u}) - \kappa \nabla^2 \chi, \tag{2.13}$$

and the conservation equation can be written

$$\frac{\partial \chi}{\partial t} = -\nabla \cdot \mathbf{J} + S_\chi \tag{2.14}$$

since $\nabla \cdot \mathbf{u} = 0$ for non-divergent flow. If this conservation equation is then integrated over the area of the control volume associated with the node of interest, and the divergence

theorem is applied, the rate of change of χ can be written as

$$\frac{\partial \chi}{\partial t} = \frac{1}{A} \oint_C \mathbf{J} \cdot \hat{\mathbf{n}} dl + S_\chi \tag{2.15}$$

where A is the area of the control volume, C is the control volume perimeter, $\hat{\mathbf{n}}$ is a unit normal and dl is along C. The integral is then calculated around C by using Simpson's Rule (see Press, Flannery, Teukolsky and Vetterling, 1989) along each of the straight lines ab, bc, cd *et c.* (Figure 2.2b). **J** is evaluated at each point along the integration path from the co-efficients of the relevant shape function, and its outward component is evaluated by forming the scalar product of **J** and $\hat{\mathbf{n}}dl$. When a bi-linear form is used for the shape functions, it is only necessary to use three points for the Simpson's Rule approximation to be exact. In general, however, more points should be used to obtain sufficient accuracy, or a higher order integration scheme such as one of the Newton-Cotes closed formulae. Source terms such as the Coriolis accelerations in SARA can simply be added on at this stage. Boundary conditions can be applied once they have been cast into a flux form. For example if the flux, **J**, of χ through an adjacent element is known, then this value can be used along the appropriate sides of the control volume rather than calculating the normal integral. If there is no flux of χ into a given element then there will be no contribution to the integral along the path through that element.

The rate of change of χ is then determined for this node and the procedure must then be repeated for the other nodes. These values are stored for use by the time stepping scheme, once the diagnostic variables have been determined. The procedure described here involves twelve simple line integrations for each independent variable, at each internal node and at each time step. It is clear then that small changes in the complexity of the integration scheme used at this stage will greatly alter the overall computation time for each time step. If, for example, five points were used for the line integrations (a sensible minimum if we were to use linear-exponential shape functions), there would be an increase in run time of $\approx 70\%$ for this part of the model.

2.4.1 Pressure correction.

Although the technique described above is suitable for the prognostic equations, a different approach needs to be taken to solve the diagnostic equation for the pressure at each time step. Many authors (such as Patankar and Baliga) use an iterative scheme known as pressure correction, which generates a pressure field consistent with the boundary conditions and an incompressible fluid. There follows a description of how a pressure correction algorithm could be used in SARA ; for a more detailed account of pressure correction the reader is referred to Patankar (1980).

Where the control volume method is used for discretizing the advection diffusion equations for the velocity components, the resulting equations for each node can be written as

$$a_i\varphi_i = \sum_{\text{nb}} a_{\text{nb}}\varphi_{\text{nb}} + S + (p_B - p_A)A_i. \tag{2.16}$$

Here i refers to the node we are considering, nb to the neighbouring nodes, φ is the velocity component, a is determined from the shape functions, S is the associated source term, p_B and p_A are the pressure values used in determining the appropriate part of the pressure gradient, and A_i is the cross-sectional area on which the pressure force acts. The pressure term is not included as part of the source term because it proves to be more convenient in this form when deriving the pressure correction equation (Eq. 2.27).

In SARA there is a momentum equation similar to Eq. 2.16 for each of the radial and vertical components. In order to solve these for the velocity, it is necessary either to know or to guess a pressure field. Since, if this is a guess, it will not in general be the correct guess, the resulting velocity field will not be consistent with the incompressibility of the flow. If we let p^* be the guessed pressure field and u^* and w^* be the resulting velocity components, then we can say

$$a_i u_i^* = \sum_{\text{nb}} a_{\text{nb}} u_{\text{nb}}^* + S + (p_W^* - p_E^*)A_{ih} \tag{2.17}$$

and

$$b_i w_i^* = \sum_{\text{nb}} b_{\text{nb}} w_{\text{nb}}^* + S + (p_B^* - p_T^*)A_{iv}. \tag{2.18}$$

We have used A_{ih} and A_{iv} as the horizontal and vertical cross-sectional areas respectively, and the subscripts E, W, T and B to denote adjacent reference points in the radial and vertical directions. If we now say that the correct solution is the current guess plus some correction (denoted by a prime) then we have

$$p = p^* + p' \tag{2.19-a}$$

$$u = u^* + u' \tag{2.19-b}$$

$$w = w^* + w'. \tag{2.19-c}$$

Now subtracting Eq. 2.17 from the radial form of Eq. 2.16 then

$$a_i u_i' = \sum_{\text{nb}} a_{\text{nb}} u_{\text{nb}}' + (p_W' - p_E') A_{ih}. \tag{2.20}$$

At this stage it is customary to discard the first term on the right hand side of Eq. 2.20. There are several reasons for this which are discussed in some depth by Patankar (1980). Some of the most important are outlined later in this section. Now

$$a_i u_i' = (p_W' - p_E') A_{ih} \tag{2.21}$$

or

$$u_i' = \frac{A_{ih}}{a_i}(p_W' - p_E'). \tag{2.22}$$

Eq. 2.22 is called the velocity correction equation and when written as

$$u_i = u_i^* + \frac{A_{ih}}{a_i}(p_W' - p_E') \tag{2.23}$$

it shows how the velocity field resulting from the guessed pressure field should be corrected to give the incompressible velocity field. One can write an equation similar to Eq. 2.23 for the vertical velocity component:

$$w_i = w_i^* + \frac{A_{iv}}{b_i}(p_B' - p_T'). \tag{2.24}$$

To complete the problem it is now only necessary to find the pressure corrections.

This is accomplished by using the continuity equation in cylindrical-polar coordinates which, for axisymmetric flow, is

$$\frac{1}{r}\frac{\partial(ru)}{\partial r} + \frac{\partial w}{\partial z} = 0. \tag{2.25}$$

Integrating Eq. 2.25 over the control volume we then get an equation of the form

$$\frac{1}{r_{\text{mid}}}(r_E u_E - r_W u_W)\Delta z + (w_T - w_B)\Delta r = 0. \tag{2.26}$$

where r_{mid} is the mean of r_E and r_W. We can now substitute for u and w from the velocity correction equations, Eqs. 2.23 and 2.24, into Eq. 2.26 to get the pressure correction equation:

$$\frac{1}{r_{\text{mid}}}\left\{r_E\left[u_E^* + \frac{A_{Eh}}{a_E}(p'_P - p'_E)\right] - r_W\left[u_W^* + \frac{A_{Wh}}{a_W}(p'_W - p'_P)\right]\right\} +$$
$$\left\{\left[w_T^* + \frac{A_{Tv}}{b_T}(p'_P - p'_T)\right] - \left[w_B^* + \frac{A_{Bv}}{b_B}(p'_B - p'_P)\right]\right\} = 0. \tag{2.27}$$

Since the starred velocity components are known from the guessed pressure field, the only unknowns are the pressure correction terms themselves. Eq. 2.27 can also be written as

$$p'_P = \frac{1}{\alpha_P}(\alpha_W p'_W + \alpha_E p'_E + \alpha_B p'_B + \alpha_T p'_T) \tag{2.28}$$

for each node where all the αs are known. Once in this form we can apply the pressure correction equation to each node and impose the appropriate boundary conditions. This set of simultaneous equations can then be solved using some appropriate matrix technique (see Appendix A) for the pressure corrections. The new pressure at each node can be calculated by adding the correction term to the initial guess, p^*. The velocity components can be updated by using the pressure correction terms in the velocity correction equations, Eqs. 2.23 and 2.24. The corrected values for the pressure can be used as the guesses for the next iteration, and the whole procedure repeated until satisfactorilly small corrections are needed.

There remains the issue of the rather forceful way in which the first term from the right hand side of Eq. 2.20 was discarded. This was justified for the following reasons:

- If this term were included, the velocity correction at a given node would be expressed in terms of its neighbouring nodes. Each of these in turn would also be expressed in terms of their neighbours, so that the velocity correction at each node would be a function of all the others. This is clearly the case in the physical sense. It is far more numerically expedient, however, to neglect these effects. In short, omitting this term allows a much simpler (more direct) method of solution.

- The pressure correction scheme described here is based on the SIMPLE algorithm, or Semi Implicit Method for Pressure Linked Equations (see, for example, Patankar 1980). If the $\sum_{nb} a_{nb} u'_{nb}$ term were retained the system would be fully implicit, and the pressure corrections at some nearby node would affect the velocity correction there and hence the velocity correction at the node of interest. This effect is omitted and so the scheme is only semi-implicit.

- It is, of course, imperative that the procedure converges, and that it converges to the correct answer. Fortunately it happens that there is convergence when the term is dropped (Patankar, 1980). The converged solution does satisfy the incompressibility constraint and so the omission of $\sum_{nb} a_{nb} u'_{nb}$ is irrelevant as far as the final answer is concerned.

- It is often necessary to use some under-relaxation of the pressure correction in order to ensure convergence. This means that only some fraction of the calculated pressure correction is added on to the guessed pressure to get the corrected value. So

$$p = p^* + \varsigma_P p' \tag{2.29}$$

where ς_p is the under-relaxation factor[†] for the pressure. The velocity correction equations are not under-relaxed because this would mean that the corrected velocities would not satisfy the continuity equation (Eq. 2.25).

Pressure correction has been successfully applied to many steady state problems in fluid dynamics, see for example Baliga *et al.* (1983); Burns and Wilkes (1987); Moore (1992); Patankar (1980); and Schneider and Raw (1987a,b). For SARA , however, we decided to use a more traditional Poisson equation method to solve the diagnostic equation for the pressure at each time step; this method is described in Appendix 1. There were three main reasons for not choosing pressure correction:

1. The Poisson equation method is simpler. Because of this simplicity it is well suited

[†]The under-relaxation factor can, in principle, take any value $0 < \varsigma_P \leq 1$. The optimum choice is problem dependent; Baliga states a value of 0.8 as widely applicable and Moore (1992), suggests a range of 0.80 to 0.96.

to the use of a fast matrix solver, which was one of the main concepts on which SARA 's design was based.

2. The pressure correction method tends to converge to a solution which satisfies the boundary conditions and is consistent with the continuity equation. However the annulus flow is time dependent and, although the boundary conditions remain fixed, the flow passes through a succession of quasi-static states before spinning up to a final steady state. It was feared that, by using the pressure correction method, only the final steady state would be calculated, and no spin-up would be simulated.

3. The pressure correction method entails the iterative application of a complex algorithm at each time step. This is costly in terms of computation time. The fast direct solver which could be used with the Poisson equation method is, for a wide range of problems, much quicker since there is only one iteration. Even the fast semi-direct solvers would be quicker because they converge in a very few iterations and each iteration is itself less involved.

It is vital to the functionality of the numerical scheme used in SARA that the velocity field is non-divergent. This constraint is fundamental to the conservation properties of the code. We impose non-divergence by using the vertical gradient of the vertical velocity field to diagnose the radial velocity field; the exact method is described in the following chapter. The vertical velocity field is obtained from a prognostic equation involving the vertical gradient of the pressure. This is the *only* part which the pressure plays in determining the state of the system at each time step, the next pressure field then being diagnosed from a Poisson equation before integrating the next time step.

There are, then, two distinct phases at each time step in SARA in which (i) the velocity components are calculated subject to the imposed boundary conditions, and (ii) the pressure is then diagnosed from a Poisson equation. The fact that there is such a clear division is a main part of the reason why this type of scheme cannot work (priv. comm., A. Fradkov, 1993 and E. Suli, 1993).

> *There is an inherent coupling between the velocity field and the pressure field in that, for given velocity boundary conditions, the pressure is uniquely determined by the velocity, subject only to an arbitrary gauge. There are no independent boundary conditions on the pressure.*

In SARA we use a matrix formulation to solve the Poisson equation for the pressure, but whatever method is employed, it is always necessary to impose some kind of boundary condition on the pressure in order for the problem not to be ill-posed. Since we are not free to impose boundary conditions (apart from a gauge), those which are used will be artificial, and can ultimately lead to a spurious pressure field.

In the Met. Office annulus model (Williams 1969, Hignett *et al.* 1985, White 1986), prognostic equations are solved for all the velocity components and the temperature; the non-divergence condition is then used to enable a Poisson equation for the pressure to be set up. The boundary condition used for the pressure is that the normal gradient should be equal to the appropriate component of the Navier-Stokes equation (Williams, 1969). At each time step, the divergence of the flow field (due to rounding errors) is computed and compensated for in the source terms of the pressure equation at the next time step. Although this formulation is satisfactory for the Met. Office model (and, indeed, there would appear to be no problems caused by the pressure terms), the same would not be true for SARA. This is because to the absence of a staggered grid necessitating the implementation of mixed interpolation (see Section 2.5.1), which, in turn, leads to spurious pressure driven flows at nodes adjacent to the boundaries. This problem is discussed further at the end of Chapter 3.

The pressure correction method is not subject to these difficulties, as the velocity components and pressure field are solved for simultaneously, requiring only one set of boundary conditions. Both mixed and equal-order interpolation have been used successfully with pressure correction (Baliga and Patankar 1983, and Prakash and Patankar 1985).

2.5 Refinements to the control volume finite element method.

The control volume finite element method, as described earlier, can be considered as a basic working version. There have been several improvements and modifications to the scheme since the initial idea was formulated, some of which have been made use of in SARA and some of which might perhaps be adopted at a later stage.

2.5.1 Mixed interpolation.

Special care has to be taken when dealing with the pressure equation. Problems can arise because the pressure does not appear explicitly in the momentum equations, only pressure gradients. This means that if the pressure is interpolated using linear shape functions, only the differences in value between the pressure stored at alternate grid points are used in the model equations. Because of this there is effectively no difference between a pressure field which is constant in space, and a "checker-board" pressure field such as the one shown in Figure 2.3. This type of pressure distribution is not physical and so must not be allowed to develop.

With finite differences, the solution is often to use staggered grids for the various variables such as the Arakawa B-, C- and E-grids favoured by oceanographers. This is, however, not an option when using finite elements because the element boundaries will not in general be aligned with the coordinate axes.

The problem is avoided in SARA by using a grid with half the resolution of the main grid on which to store the pressure values. Control volumes are constructed in exactly the same way for each node, the pressure nodes being located at alternate nodes of the primary grid used for the velocity and temperature values. This scheme is a simplified adaptation of the mixed interpolation scheme used by Baliga and Patankar (1983). This simplification was possible because of the ordered nature of the spatial grid used for SARA . The whole domain discretization and control volume construction routine is computationally very

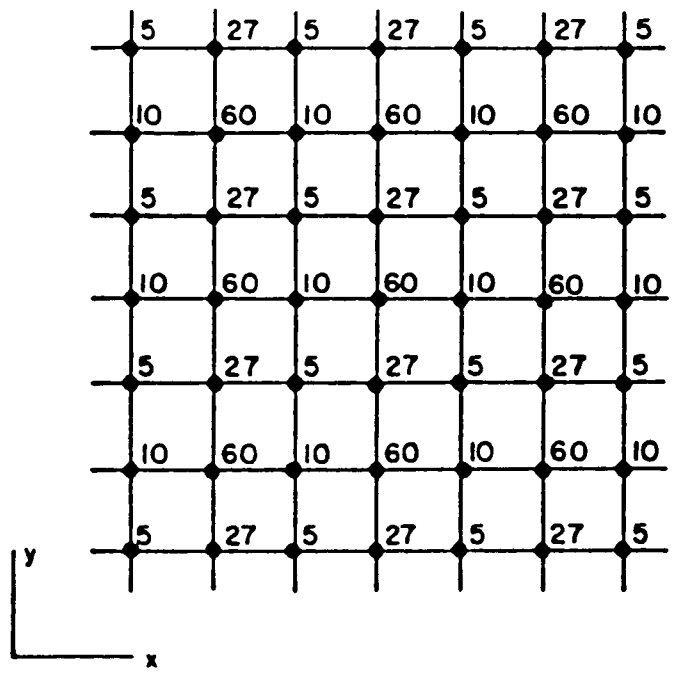


Figure 2.3: A pathological example of checker-boarding in a pressure field. The pressure gradient would be calculated as zero, despite the fact that it is obviously not. This type of field can arise because the de-coupling between pressure points has no effect on the velocity calculation since it still produces the correct pressure gradient values.

simple and so can readily be adapted to deal with different numerical domains and hence different physical geometries.

2.5.2 Linear-exponential basis functions.

For problems in which conduction plays an important rôle, such as thermal forcing in SARA , the use of linear basis functions for the finite elements could lead to unrealistic oscillatory solutions, which could ultimately cause iterative schemes to diverge. This is particularly true when either or both of the velocity components vary rapidly in one of the coordinate directions[†].

One solution is first to define a local coordinate system for each element. This is chosen such that the X axis is parallel to the average flow vector across that element, and the Y axis is orthogonal to it. Linear-exponential basis functions can then be used, where the exponential variation is along the average flow direction and the linear variation

[†]This is exactly the case in the corners of the computation domain in SARA .

perpendicular to it:

$$\varphi = A \exp\left\{\frac{\bar{\rho}\hat{U}X}{\Gamma}\right\} + BY + C, \tag{2.30}$$

where $\bar{\rho}$ is the mean density, $\hat{U}$ is the magnitude of the mean velocity across the element, and Γ is the diffusion coefficient for φ. Further details of this method are given by Baliga and Patankar (1980).

2.5.3 Using an upwind formulation.

In order for any scheme to be unconditionally conservative, whatever flux leaves one control volume must enter the neighbouring one. If this is the case then the total flux within a control volume must have some negative contributions. It is possible that the net outward flux could be greater than the net inward flux and in this case the total discretization coefficient could be negative. This can lead to values of the transport variable becoming greater/smaller than the largest/smallest value specified by the boundary conditions, even when there are no sources/sinks.

Using an upwind or donor-cell method (Potter 1973, for example) it is possible to ensure that the discretization coefficient is always positive so that these under- or overshoot problems cannot occur. We can ensure that this is the case by requiring that flux of any transport variable out of a control volume is calculated using its value at the node at the centre of the control volume which it is leaving. This is the idea behind upwinding. If we consider two points μ and ν then, if the flux from μ to ν if positive, we use the value at μ, and conversely, if it is negative we use the value at ν.

Upwinding schemes applied to CVFEMs are described in Prakash (1987), and Schneider and Raw (1986). In SARA we have not used any upwinding, In the light of the excellent results achieved by these authors in test problems, however, it would seem a very sensible strategy to adopt. It is also worth noting that the implementation of upwinding schemes usually increases the numerical diffusion perpendicular to the mean flow (Prakash, 1987).

2.5.4 Equal order velocity-pressure interpolation.

The mixed order interpolation scheme, described above (Section 2.5.1), is effective, but has some drawbacks. The pressure grid will tend to be rather coarse and so will not be very accurate for problems in which the pressure varies on a small scale. It will also be necessary to use two sets of finite element domain discretizations and two sets of control volumes, one being for the pressure and one for the other variables. Although the numerical cost is negligible for the determination of the finite elements for the pressure (due to the pressure nodes always being coincident with nodes on the primary grid), there is no such commonality to capitalize upon when determining the pressure control volumes. It was exactly these kind of considerations which prompted the development of equal order schemes such as those of Prakash and Patankar (1985), and Prakash (1986). These authors use the same nodal distribution for pressure, velocity and temperature. Despite the obvious advantages of this type of scheme for some applications, there would be little benefit from using an equal order scheme in SARA once the difficulties with the pressure term have been solved. This is because the regular grid structure used makes the application of two sets of finite elements and two sets of control volumes relatively straightforward.

Chapter 3

SARA: A numerical model of the rotating annulus.

We begin this chapter by presenting the equation set used for the rotating annulus model, together with the relevant boundary conditions. We then go on to describe the various integration schemes used for these equations. The model has undergone a three phase validation, and this process is then described and illustrated with some salient results.

3.1 The structure of the grid.

The grid used was two dimensional (i.e. a radial/vertical cross section of the annulus) because we have assumed (for simplicity) that the flow is axisymmetric in formulating the equation set (Eqs. 3.16 to 3.20). In order to resolve the side-wall and end-wall boundary layers, the grid can be stretched so as to give increased resolution near the boundaries. The stretching was accomplished using independent mapping functions in the radial and vertical directions based on those used by Farnell and Plumb (1976). These were of the form

$$\zeta(r) = r_0 \tanh(\alpha_r r), \tag{3.1}$$

and

$$\eta(z) = z_0 \tanh(\alpha_z z), \tag{3.2}$$

where $\zeta(r)$ and $\eta(z)$ are the transformed radial and vertical grids, r_0, α_r, z_0 and α_z are constants which define the degree of stretching and interval of the grid, and r and z are zero at mid-height, mid-radius. Since the modulus of the hyperbolic tangent function is always less than 1, it was possible to non-dimensionalize the the problem with respect to the half width/height of the annulus. This enabled us to drop the scaling terms r_0 and z_0, so that each mapping function was determined by a single stretching parameter, α_r or α_z. It was then a simple matter to re-dimensionalize the grid after it had been stretched. The effect of $\alpha_{r,z}$ on the stretching of the vertical grid is illustrated in Figure 3.1. The upper plots show $\zeta(r)$ for values of $\alpha_{r,z}$ of 1, 2 and 3, and the lower plots show how the horizontal grid lines would look in each case. Each set has seventeen grid lines (as in SARA), although for $\alpha_{r,z} = 3$ the grid is compressed to such an extent near the boundary, that they become indistinguishable in the figure. Other mapping functions can be used to stretch the grid, such as

$$\rho(r) = r_0(\alpha_r r)^{1/3}. \tag{3.3}$$

The method described above is preferable, however, because whatever value is used for $\alpha_{r,z}$, the transformed grid is always between -1 and 1, and is therefore simple to rescale to the computational domain.

When SARA is run the program prompts the user for a horizontal and a vertical boundary layer thickness to be resolved. Since the physical dimensions of the space to be gridded and the number of grid lines are already specified, we can define a reference grid interval using these, assuming that the grid is uniform. If the user specifies a boundary layer thickness greater than this reference length, then a regular grid is used. Otherwise, the ratio of the boundary thickness specified by the user to the reference grid interval gives a number between 0 and 1 (the stretching ratio), determining how stretched the grid will need to be. A look-up table is then used to determine the appropriate value of $\alpha_{r,z}$ (see Figure 3.2). The boundary grid interval generated will never be greater than that specified by the user, however it may be marginally smaller. This is a result of using a look-up table for the scaling and can't be practically avoided.

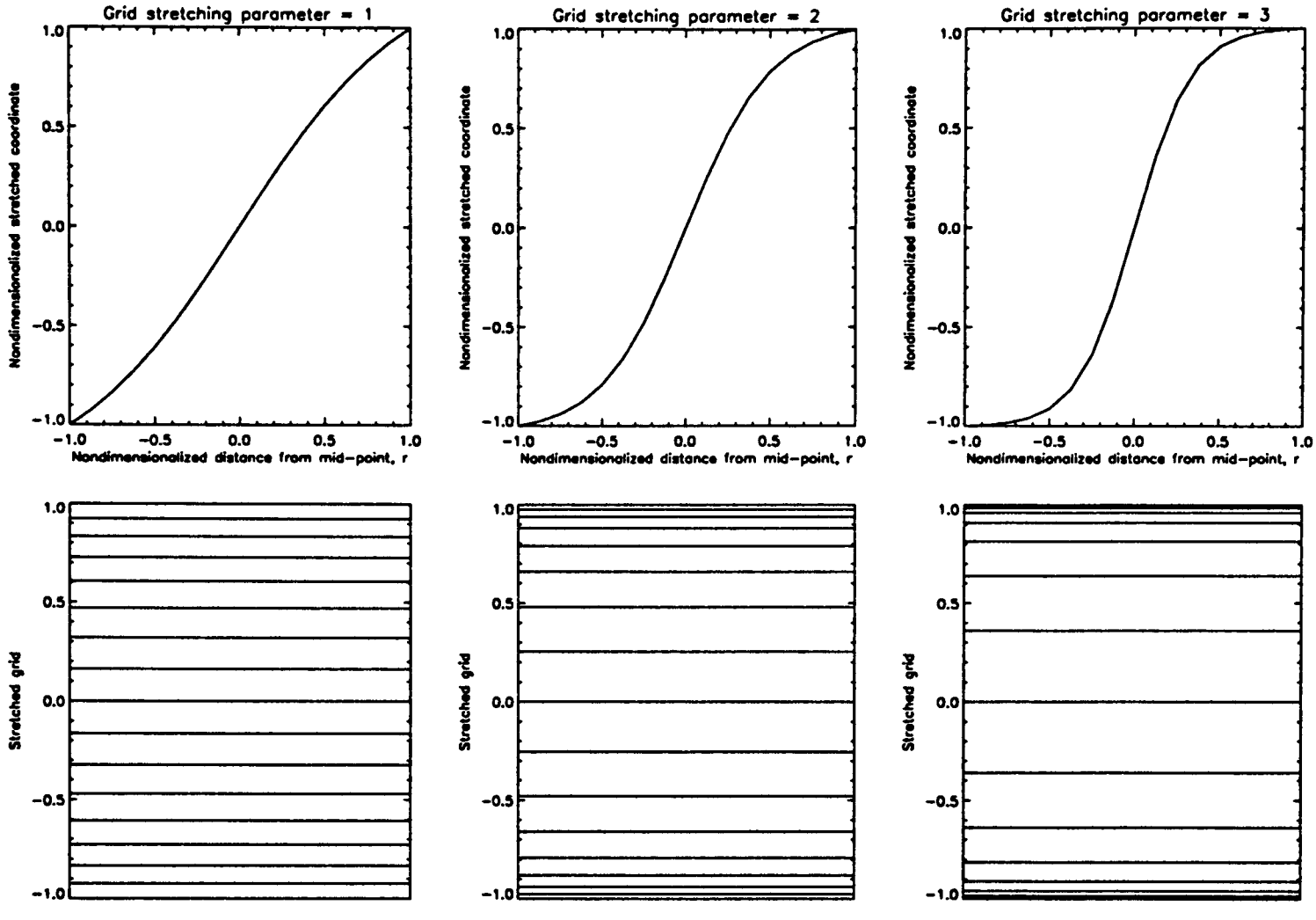


Figure 3.1: An illustration of the effect of $\alpha_{r,z}$, the stretching parameter, on the grid structure. $\alpha_{r,z}$ was 1 for the left hand plot, 2 for the centre plots and 3 for the right hand plots. Each figure has the same number of grid lines, shown in the lower figures, although some are so close as to be unresolvable.

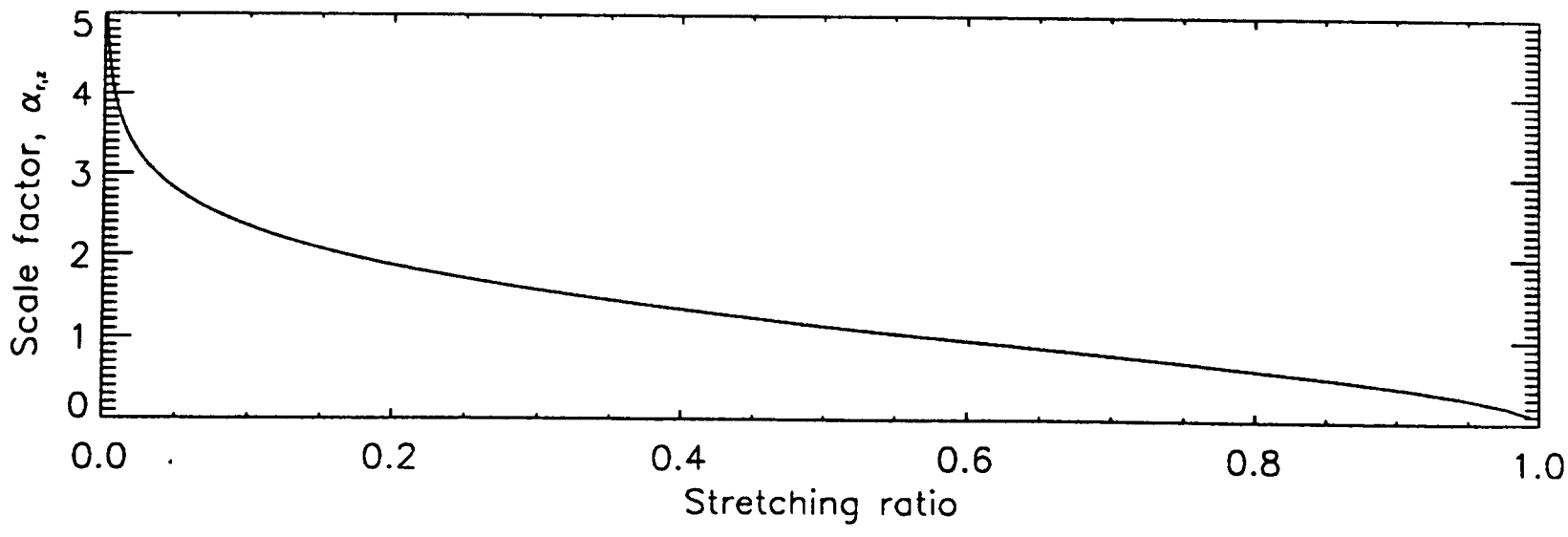


Figure 3.2: The variation of the scale factor, $\alpha_{r,z}$ used in the grid stretching mapping function with the stretching ratio (see main text). When the stretching ratio is small, the scale factor changes very rapidly.

This method of determining the structure of the stretched grid is based on the method used in the existing Met. Office annulus model, described by Farnell and Plumb (1975). The main difference between the two schemes is that, whereas in SARA the user is prompted for the boundary layer thicknesses, Farnell and Plumb calculated their thicknesses by assuming the presence of Ekman and thermal or Stewartson boundary layers on the end and side walls respectively. SARA could easily be adapted to include such a scheme, although some care must be taken over stretching the vertical grid when rotation is turned off and the Ekman number becomes undefined.

3.2 The equation set.

The rotating annulus system can be defined in terms of five variables: the temperature, the pressure and the radial, zonal and vertical velocity components. In order to simplify the equations, we can make the following assumptions:

i that the Boussinesq approximation holds (*i.e.* that density variations can be ignored, except where they are coupled with gravity to produce buoyancy forces),

ii that the density varies linearly with temperature, and

iii that the kinematic viscosity and thermal diffusivity are independent of temperature.

The first assumption can be justified by scaling arguments (see for example Rayer, 1992). With this approximation, the continuity equation is the same as for an incompressible fluid. The second and third approximations are simplifications which can be improved in later versions of SARA to describe the effect of temperature more accurately[†]. The equations describing the system can then be written in cylindrical polar coordinates as,

$$v_t + uv_r + \frac{v}{r}v_\phi + wv_z = -\frac{1}{\bar{\rho}}\frac{P_\phi}{r} - u(2\Omega + \frac{v}{r}) + \nu\left[\frac{1}{r}\frac{\partial}{\partial r}(rv_r) + \frac{v_{\phi\phi}}{r^2} + v_{zz} + \frac{2u_\phi}{r^2} - \frac{v}{r^2}\right] \quad (3.4)$$

[†]It is a relatively simple matter to have the density, kinematic viscosity and thermal diffusivity varying quadratically with temperature. This is the technique adopted by the Met. Office annulus model.

$$w_t + uw_r + \frac{v}{r}w_\phi + ww_z = -\frac{1}{\bar{\rho}}P_z + g\alpha(T - \bar{T}) + \nu\left[\frac{1}{r}\frac{\partial}{\partial r}(rw_r) + \frac{w_{\phi\phi}}{r^2} + w_{zz}\right] \tag{3.5}$$

$$T_t + uT_r + \frac{vT_\phi}{r} + wT_z = \kappa\left[\frac{1}{r}\frac{\partial}{\partial r}(rT_r) + \frac{T_{\phi\phi}}{r^2} + T_{zz}\right] \tag{3.6}$$

$$\frac{1}{r}\frac{\partial}{\partial r}(ru) + \frac{v_\phi}{r} + w_z = 0 \tag{3.7}$$

$$\nabla^2\left(\frac{P}{\bar{\rho}}\right) = 2\Omega v_r + 2\Omega\frac{v}{r} - \frac{2\Omega u_\phi}{r} + g\alpha T_z \tag{3.8}$$

$$-\frac{uu_r}{r} - u_ru_r - uu_{rr} - \frac{2v_ru_\phi}{r} - \frac{vu_{r\phi}}{r} - \frac{wu_z}{r} - wu_{rz} + \frac{2vv_r}{r}$$
$$-\frac{uv_{r\phi}}{r} - \frac{v_\phi v_\phi}{r^2} - \frac{vv_{\phi\phi}}{r^2} - \frac{2w_\phi v_z}{r} - \frac{wv_{\phi z}}{r} - \frac{u_\phi v}{r^2} - \frac{uv_\phi}{r^2}$$
$$-2u_zw_r - uw_{rz} - \frac{vw_{\phi z}}{r} - w_zw_z - ww_{zz}$$

where r, ϕ and z are the radial, zonal and vertical co-ordinates, u, v and w are the radial, zonal and vertical velocity components, T is the temperature, P is the pressure, Ω is the rotation rate, ν is the kinematic viscosity, $\bar{\rho}$ is the mean density, g is gravity, α is the thermal expansivity of water, κ is the thermal diffusion coefficient and $\bar{T}$ is the local mean temperature. The subscripts denote partial derivatives.

For axisymmetric flow, all the zonal derivatives are zero and so can be dropped from the equations. We can then write Eqs. 3.4 to 3.8 as

$$v_t + uv_r + wv_z = -u(2\Omega + \frac{v}{r}) + \nu\left[\frac{1}{r}\frac{\partial}{\partial r}(rv_r) + v_{zz} - \frac{v}{r^2}\right] \tag{3.9}$$

$$w_t + uw_r + ww_z = -\frac{1}{\bar{\rho}}P_z - g\alpha(T - \bar{T}) + \nu\left[\frac{1}{r}\frac{\partial}{\partial r}(rw_r) + w_{zz}\right] \tag{3.10}$$

$$T_t + uT_r + wT_z = \kappa\left[\frac{1}{r}\frac{\partial}{\partial r}(rT_r) + T_{zz}\right] \tag{3.11}$$

$$\frac{\partial}{\partial r}(ru) = -r\frac{\partial w}{\partial z} \tag{3.12}$$

$$\nabla^2\left(\frac{P}{\bar{\rho}}\right) = 2\Omega v_r + 2\Omega\frac{v}{r} + g\alpha T_z \tag{3.13}$$

$$-\frac{uu_r}{r} - u_ru_r - uu_{rr} - wu_{rz} - 2u_zw_r$$
$$-\frac{wu_z}{r} - w_zw_z - ww_{zz} - uw_{rz} + \frac{2vv_r}{r}$$

The pressure terms above refer to the dynamic perturbation from hydrostatic pressure. This is so that it is easier to produce diagnostic plots, which would otherwise be dominated by the hydrostatic part. It also serves to make the numbers used in the matrix solver for the poisson equation much smaller, which helps with the numerical stability and accuracy.

3.2.1 The equations in flux form

If we define the 2-dimensional operators

$$\nabla_s \phi = \frac{\partial \phi}{\partial r}\hat{\mathbf{r}} + \frac{\partial \phi}{\partial z}\hat{\mathbf{z}} \tag{3.14}$$

and

$$\nabla_s \cdot \eta = \frac{1}{r}\frac{\partial}{\partial r}(r\xi) + \frac{\partial \zeta}{\partial z} \tag{3.15}$$

where $\eta = \xi\hat{\mathbf{r}} + \zeta\hat{\mathbf{z}}$, then Eqs. 3.9–3.13 can be written as

$$\frac{\partial v}{\partial t} = -\nabla_s \cdot (\mathbf{u}v - \nu\nabla_s v) - u(2\Omega + \frac{v}{r}) \tag{3.16}$$

$$\frac{\partial w}{\partial t} = -\nabla_s \cdot (\mathbf{u}w - \nu\nabla_s w) - \frac{1}{\bar{\rho}}\frac{\partial P}{\partial z} - g\alpha(T - \bar{T}) \tag{3.17}$$

$$\frac{\partial T}{\partial t} = \nabla_s \cdot (\kappa\nabla_s T - \mathbf{u}T) \tag{3.18}$$

$$\frac{\partial}{\partial r}(ru) = -r\frac{\partial w}{\partial z} \tag{3.19}$$

$$\nabla^2\left(\frac{P}{\bar{\rho}}\right) = -\nabla_s \cdot \left[\mathbf{u}\cdot\nabla_s\mathbf{u} + 2\Omega \wedge \mathbf{u} - g\alpha(T - \bar{T})\right] \tag{3.20}$$

3.3 Boundary conditions.

In order to specify the problem completely, boundary conditions have to be imposed for each of the Eqs. 3.16 to 3.20. A no-slip boundary is assumed for the velocity components, so that

$$u, v, w(r = r_a, r = r_b, z = z_B, z = z_T) = 0. \tag{3.21}$$

These conditions always hold and so, once the velocity values have been set to zero for the nodes on the boundary at the first time step, they never need to be recalculated. This

reduces the number of nodes at which the velocity equations have to be solved from $m*n$ (for a grid with m radial and n vertical grid points) to $(m*n)-2m-2n+4$, so if the grid is 17 grid points square as in SARA , then the number of velocity calculations is reduced from 289 to 225 nodes, a saving of about 25%.

The temperature is held constant on the vertical boundaries, providing one of the main driving mechanisms in the model. The top and bottom boundaries are perfect insulators and so the isotherms must intersect at right angles to the boundaries. In a similar style to Eq. 3.21 the temperature boundary conditions can be written

$$T(r = r_a) = T_{cold}, \tag{3.22-a}$$

$$T(r = r_b) = T_{hot}, \tag{3.22-b}$$

$$\frac{\partial T}{\partial z}(z = z_B, z = z_T) = 0. \tag{3.22-c}$$

As with the velocity components, this is very simple to apply to the vertical boundaries since, once the temperature has been specified at the first time step, it never has to be recalculated. Some subtlety is needed, however, to evaluate the temperature on the top and bottom boundaries from Eq. 3.22-a. In a normal finite difference scheme "ghost" cells are often included outside of the physical region being modelled. These can be used to apply gradient boundary conditions by setting the value in the ghost cell to the necessary value. This is not applicable to the control volume finite element method (or CVFEM) as it would involve constructing control volumes around boundary nodes, which would be half in and half out of the calculation domain. The method used in SARA solves the problem of dealing with boundary nodes by not using CVFEM for them at all. The temperature is evaluated at all the internal nodes as normal and then the two values nearest to the boundary are used to determine the temperature on the boundary nodes. Since any function can be well approximated by a quadratic in the region of a turning point the vertical temperature profile is assumed to be quadratic near the top and bottom boundaries. Consider a point on the bottom boundary where the temperature needs to be calculated and the two nodes immediately above it where the temperature is known. If 1 and 2 denote the nodes above the boundary we have that

$$T(z_B) = T_B, \quad \text{to be evaluated} \tag{3.23-a}$$

$$T(z_1) = T_1, \tag{3.23-b}$$

$$T(z_2) = T_2. \tag{3.23-c}$$

The quadratic form for the temperature profile is

$$T = az^2 + bz + c \tag{3.24}$$

and so

$$\frac{\partial T}{\partial z} = 2az + b. \tag{3.25}$$

Now, since the vertical temperature gradient is zero (from Eq. 3.22-a)

$$\left.\frac{\partial T}{\partial z}\right|_{z=z_B} = 0. \tag{3.26}$$

Then substituting from Eq. 3.26 into equation 3.25

$$b = -2az_B, \tag{3.27}$$

and if $z_B = 0$ then $b = 0$, and so

$$T = az^2 + c. \tag{3.28}$$

Substituting T_1 and T_2 at z_1 and z_2 into Eq. 3.28, and solving these equations simultaneously gives

$$a = \frac{T_2 - T_1}{z_2^2 - z_1^2} \tag{3.29}$$

$$c = T_1 - \frac{(T_2 - T_1)z_1^2}{(z_2^2 - z_1^2)}, \tag{3.30}$$

and since $z_B = 0$ Eq. 3.24 gives

$$T_B = c = T_1 - \frac{(T_2 - T_1)z_1^2}{(z_2^2 - z_1^2)}. \tag{3.31}$$

This method is described for a non-uniform grid but is equally applicable (and simpler) on a regular grid. A similar calculation is performed at the top boundary and in this way the insulating boundary condition can be applied.

The pressure equation (3.20) is solved by first transforming it into a matrix problem and then using a specially developed fast solver as described in Appendix A. It is necessary to specify boundary conditions for the pressure in order to construct this matrix. It was

shown in Chapter 2 that there are *no* independent pressure boundary conditions, and so the matrix must be constructed using artificial constraints. Since the pressure we are considering is the departure from hydrostatic, a first approximation to the pressure perturbation on the boundaries is

$$P(r = r_a, r = r_b, z = z_B, z = z_T) = 0 \tag{3.32}$$

i.e. that the pressure at the boundaries is the total pressure which would be there in the absence of any dynamical or thermal forcing. This approximation can be justified by considering the pressure equation in component form, (Eq. 3.13), with the velocity components and the vertical temperature gradient set to zero (due to the no-slip boundary condition, the insulating lid and base, and the isothermal side walls) at the boundaries. We then have that

$$\nabla^2 \left(\frac{P_{boundary}}{\bar{\rho}} \right) = 0, \tag{3.33}$$

which implies no short scale variation of the pressure on the boundaries. Since the gauge of the pressure field is arbitrary, it can be chosen so that the value at the walls is always small, compared to the hydrostatic pressure (which, using $P \sim \bar{\rho} g z$, has a typical value of $\sim$ 1000 Pa). The non-hydrostatic part is, therefore, negligible on the boundaries, at least to a first approximation, and so Eq. 3.32 could be applied.

An alternative choice of pressure boundary conditions (as adopted by the Met. Office model, for example), is to calculate the normal pressure gradient from the relevant component of the Navier-Stokes equation (Bell 1986, or Williams 1969), with the velocity set to zero (for no-slip boundaries). This gives, for the top and bottom boundaries,

$$-\frac{1}{\bar{\rho}} P_z = g\alpha(T - \bar{T}) - \nu w_{zz}, \tag{3.34}$$

where $w_r = 0$ and so is dropped. Similarly, for the side wall boundaries, we have

$$-\frac{1}{\bar{\rho}} P_r = -\nu \left[\frac{1}{r} \frac{\partial}{\partial r} (r u_r) \right], \tag{3.35}$$

where $u_{zz} = 0$ and so is dropped.

These boundary conditions are consistent with zero flow at the boundaries, and are used in preference to the hydrostatic boundary condition in the current version of SARA .

Since the model can not work while the pressure and velocity remain uncoupled, there is little benefit to be gained from refining the boundary conditions any further. It is important, however, to consider the way we calculate the pressure term, when interpreting the results from SARA (see the discussion of the tertiary validation results later in this chapter).

3.4 Outline of solution scheme.

There are two types of equations to be solved at each time step in SARA , the prognostic equations (Eqs. 3.16, 3.17 and 3.18) and the diagnostic equations (Eqs. 3.19 and 3.20). The diagnostic equations are solved at each time step using the most recently calculated values of all the unknowns and so need to be solved after the prognostic equations. Since the radial velocity is determined from the vertical velocity only, Eq. 3.19 is solved before Eq. 3.20 so that the most recent value of *all* the variables can be used to determine the pressure. The order in which the prognostic equations are solved is not important because the zonal and vertical velocity components and the temperature are all integrated in time using tendencies calculated from the value of the variables at the previous time step.

The solution scheme for each time step might be represented as:

$$
\begin{aligned}
i)\quad & \left.\frac{\partial v}{\partial t}\right|_i && \models && u_i, v_i, w_i \\
ii)\quad & \left.\frac{\partial w}{\partial t}\right|_i && \models && u_i, w_i, T_i, P_i \\
iii)\quad & \left.\frac{\partial T}{\partial t}\right|_i && \models && u_i, w_i, T_i \\
iv)\quad & v_{i+1} && \models && v_{i-1}, v_i, \left.\frac{\partial v}{\partial t}\right|_i \\
v)\quad & w_{i+1} && \models && w_{i-1}, w_i, \left.\frac{\partial w}{\partial t}\right|_i \\
vi)\quad & T_{i+1} && \models && T_{i-1}, T_i, \left.\frac{\partial T}{\partial t}\right|_i \\
vii)\quad & u_{i+1} && \models && w_{i+1} \\
vii*)\quad & w_{i+1} && \models && u_{i+1} \\
viii)\quad & P_{i+1} && \models && u_{i+1}, v_{i+1}, w_{i+1}, T_{i+1}
\end{aligned}
$$

where $\models$ has been used to mean "is calculated using", and the subscripts denote the time step index. Components *i*, *ii* and *iii* are evaluated using the control volume finite element

method discussed in Chapter 2. Components *iv*, *v* and *vi* refer to the time stepping scheme, which is described later in this chapter (Section 3.5). The diagnosis of the radial velocity using the continuity equation is described by component *vii*; component *vii** refers to the second pass of the edge-weighted cross-over relaxation scheme which is used in the present version of SARA (see Section 3.6). The control volume finite element method is again used for component *viii*, together with a linear matrix solver in order to diagnose the pressure from the divergence of the Navier-Stokes equation, Eq. 3.20 (see Appendix A).

3.5 The time stepping scheme.

A common practice for the time differencing of advection/diffusion problems is to use a two-step formulation. These methods achieve stability by centre-differencing the time stepping equation. This is accomplished by using the data from the two previous time steps to work out its value at the next. This can be written as

$$\varphi_{i,j}^{n+1} = \varphi_{i,j}^{n-1} + 2\triangle_t \left(\frac{\partial \varphi}{\partial t}\right)_{i,j}^{n}. \tag{3.36}$$

The Lax-Wendroff and leap-frog formulations are common examples of two-step schemes and can both be described by an equation such as Eq 3.36. The essential difference between them is in the spatial representation of the $\frac{\partial \varphi}{\partial t}$ term, and since it is for precisely this term which the control volume finite element method is used, neither are applicable in full to SARA .

Although two step schemes are very commonly used, they are nevertheless subject to those limitations which affect all advection algorithms. The problems arise because advection is an inherently non-linear process and von-Neumann analysis of the scheme in 3.36 shows that it is subject to numerical diffusion and dispersion (Potter, 1973). The diffusion is usually small compared to the physical diffusion and so does not pose a problem. Numerical dispersion can only be limited by using a higher order scheme; this would be very computationally intensive and would necessitate the storage of data from several time steps (four for fourth order). Another problem common to all schemes of

this kind is the growth and eventual dominance of a "numerical mode" in the solution. This mode arises because of decoupling between adjacent time steps and is not dissimilar to the checker-board instability which was discussed in Chapter 2. The numerical mode oscillates about the physical solution and grows in time until it eventually dominates it (Potter, 1973). It can be eliminated by periodically smoothing the solution out over the last two time steps and continuing the integration with this smoothed solution replacing the last two time steps. This smoothing is typically carried out every seventeen or so time steps, but any prime number will avoid aliasing of any natural periodicities in the system. This is a crude concept but it worked fairly well. Without this smoothing, SARA became dominated by the numerical mode very quickly (it grows exponentially in time), typically causing the code to crash after ~ 100 time steps.

There are two problems with periodically smoothing out the numerical mode. Firstly, the smoothed scheme is no longer exactly second order accurate, although as the interval between smoothing operations is increased, the solution tends to second order accuracy. Secondly, it is a rather *staccato* process, like the ticking movement of the second hand on some watches compared to the smooth sweep of a Rolex. We wanted a time stepping scheme for SARA which suppressed the numerical mode, but which at the same time was a continuous process which could be applied at every time step.

This is achieved by mixing a little of the current state of the system with its predicted value at each time step (priv. comm., S. D. Mobbs 1992). If we consider φ as representative of any of the prognosed variables, and drop the spatial indices, i and j, we can write this mixing as

$$\varphi_{\mathcal{I}}^{n+1} = \varphi^{n-1} + 2\Delta_t \left(\frac{\partial \varphi}{\partial t}\right)^n \tag{3.37}$$

$$\varphi^{n+1} = (1-\epsilon)\varphi_{\mathcal{I}}^{n+1} + \epsilon\varphi^n \tag{3.38}$$

where ϵ is small and represents the fractional amount of mixing of the states, and the subscript $\mathcal{I}$ refers to an intermediate value. This can be written directly as

$$\varphi^{n+1} = (1-\epsilon)\varphi^{n-1} + \epsilon\varphi^n + (1-\epsilon)2\Delta_t \left(\frac{\partial \varphi}{\partial t}\right)^n . \tag{3.39}$$

For convenience, we will use f to mean $\frac{\partial \varphi}{\partial t}$, the term evaluated using finite elements, in the following discussion. If we use second order Taylor expansions for φ^{n-1} and φ^{n+1} in

the above expression, we find

$$\left(\widetilde{\frac{\partial\varphi}{\partial t}}\right)^n = \frac{(2-2\epsilon)}{(2-\epsilon)} f^n - \frac{\epsilon(\Delta_t)}{(4-2\epsilon)} \left(\frac{\partial^2\varphi}{\partial t^2}\right)^n, \tag{3.40}$$

where the term on the left hand side is tilded to distinguish it from f^n. When ϵ is zero, we recover the normal two step scheme, but as its value is increased to introduce some temporal mixing, there is a corresponding increase in the first order error in the representation of the time derivative.

We now seek to adapt the scheme so that rather than being multiplied by $(2-2\epsilon)/(2-\epsilon)$ as in Eq. 3.40, the first order time derivative is unaffected by the amount of mixing. We stipulate that

$$\left(\widetilde{\frac{\partial\varphi}{\partial t}}\right)^n = f^n - \varepsilon(\Delta_t) \left(\frac{\partial^2\varphi}{\partial t^2}\right)^n. \tag{3.41}$$

Next we use the second order Taylor expansions for φ^{n-1} and φ^{n+1} to find

$$2\Delta_t \left(\widetilde{\frac{\partial\varphi}{\partial t}}\right)^n = \varphi^{n+1} - \varphi^{n-1}, \tag{3.42}$$

and

$$(\Delta_t)^2 \left(\frac{\partial^2\varphi}{\partial t^2}\right)^n = \varphi^{n+1} - 2\varphi^n + \varphi^{n-1}. \tag{3.43}$$

On substituting these into Eq. 3.41 we find

$$\varphi^{n+1} = \left(\frac{1+2\varepsilon}{1-2\varepsilon}\right)\varphi^{n-1} - \frac{4\varepsilon}{(1-2\varepsilon)}\varphi^n + \frac{2\Delta_t}{(1-2\varepsilon)} f^n, \tag{3.44}$$

or

$$f^n = \left(\frac{\partial\varphi}{\partial t}\right)^n = \frac{(1-2\varepsilon)\varphi^{n+1} + 4\varepsilon\varphi^n - (1+2\varepsilon)\varphi^{n-1}}{2\Delta_t}. \tag{3.45}$$

This is the time stepping scheme which is presently used in SARA . It entails the storage of one more level of data than a standard two-step scheme, but by choosing a suitable value for ε there is no need for occasional mixing steps. If $\varepsilon = 0$ in Eq. 3.44, then it can be simplified to $\varphi^{n+1} = \varphi^{n-1} + 2\Delta_t f^n$ which is the standard two-step forward difference scheme discussed above (Eq. 3.36). Similarly, setting $\varepsilon = -1/2$ we recover $\varphi^{n+1} = \varphi^n + \Delta_t f^n$ which is the unconditionally unstable one-step forward difference scheme.

The time stepping scheme used in SARA is quasi centre differenced in that, as $\varepsilon \to 0$, it tends to a formally centre differenced scheme. The time stepping is, therefore, quasi

second order accurate in Δ_t. It is, however, still subject to the Courant, Friedrich Lewy condition since the maximum speed of the flow of information cannot exceed the mesh speed. This effectively limits the maximum time step to ~ 0.5 s for a typical flow pattern and grid spacing using SARA .

During normal running of SARA , a small (negative) value of ε is used to mix between successive time steps. A value of -0.02 has proved suitable. This time stepping scheme is particularly useful because it allows provision to be made for the problem of initializing the first time step. Because the value of $\varphi^{n=-1}$ is not known, most multi-level schemes have to employ a special differencing technique for the first time step. However, since the scheme described here can be made forward differenced by changing the value of ε, this problem can be avoided. We can use the same time stepping scheme at every time step by simply changing the value of ε to -0.5 for the first time step[†]. If this still proves too inaccurate, the time step itself can be made smaller for the first few iterations. This did not, however, appear to be necessary for SARA .

3.6 Edge-weighted cross-over relaxation.

The radial velocity is diagnosed from the vertical velocity, via the continuity equation (Eq. 3.12). This equation has no term for the zonal velocity component because the flow is, by definition, axisymmetric and the zonal derivatives of all the variables are zero. The radial velocity is calculated on each vertical level separately using a line-by-line method. This is non-trivial, however, since the problem is over-determined; the radial derivative term associated with the radial velocity is known at every internal node (the source terms), and the radial velocity has to be zero on both side walls (the boundary conditions). Since each horizontal level is a separate one dimensional problem with only a first order differential operator, there can be only one independent boundary condition for each equation, at each level. Several methods are available for solving this type of

[†]This principle can be improved on slightly by letting ε change gradually from -1/2 to its normal time step value smoothly, over the first few time steps.

problem, such as the linear and nonlinear shooting algorithms (see, for example, Burden and Faires, 1985) or relaxation methods (see Press *et al.*, 1989). The scheme described here is a form of shooting algorithm; it is fast and simple, and is the scheme used in SARA.

The vertical velocity gradient is calculated for each of the internal nodes, and, by starting from one of the side boundaries where the velocity is zero, a forward difference scheme is used to work across to the other boundary (Eq. 3.46). There is no guarantee that such a scheme will give zero velocity for the boundary grid point at the opposite side, and in general it does not. The same method can be applied starting from the other boundary and working back to the start. These two solutions, the forward and backward solution, are equally incorrect but can be used to evaluate a good approximation to the actual solution. This is done by using each method up to within one node of the boundary and then taking an average of the two solutions as the final answer. It can be shown that each solution gets less accurate as it gets further from the start, and so some kind of weighted averaging has to be used to favour whichever of the two solutions is most accurate at a particular point. The forward difference scheme can be written as

$$\tilde{f}_i = \tilde{f}_{i-1} + (x_i - x_{i-1})g_i, \tag{3.46}$$

$$g_i = \left.\frac{\partial f}{\partial x}\right|_i \tag{3.47}$$

where f is the function we are solving for, x is the space co-ordinate, subscripts denote grid indices and $\tilde{f}$ is the calculated approximation to f. In terms of solving Eq. 3.12, $f_i = r_i u_i$ and $g_i = -r_i \frac{\partial w_i}{\partial z}$. Consider a Taylor expansion of f around the i^{th} grid node

$$\bar{f}_i = \bar{f}_{i-1} + (x_i - x_{i-1})g_i - \frac{(x_i - x_{i-1})^2}{2!}g_i' + \frac{(x_i - x_{i-1})^3}{3!}g_i'' + \cdots \tag{3.48}$$

where $\bar{f}$ is the real value of f and primes denote derivatives. The difference between the real and the calculated value of f is the error, ϵ in the approximation, so

$$\epsilon_i = \left|\bar{f}_i - \tilde{f}_i\right| = \left|-\frac{(x_i - x_{i-1})^2}{2!}g_i' + \frac{(x_i - x_{i-1})^3}{3!}g_i'' + \cdots\right| \tag{3.49}$$

If the Taylor expansion is carried out as before, but at the $(i+1)^{th}$ node, we have

$$\bar{f}_{i+1} = \bar{f}_i + (x_{i+1} - x_i)g_{i+1} - \frac{(x_{i+1} - x_i)^2}{2!}g_{i+1}' + \cdots \tag{3.50}$$

but we can substitute the real solution for the i^{th} node, with the approximate solution, $\tilde{f}_i$, and the associated error, giving

$$\bar{f}_{i+1} = \tilde{f}_i + \epsilon_i + (x_{i+1} - x_i)g_{i+1} - \frac{(x_{i+1} - x_i)^2}{2!}g'_{i+1} + \cdots \tag{3.51}$$

It is apparent from equation 3.51 that the error at any grid node includes the error at all the previous nodes in the calculation, so that

$$\epsilon_n^\sigma = \left|\sum_{i=1}^{n} \epsilon_i\right| = \left|\sum_{i=1}^{n} \left[-\frac{(x_i - x_{i-1})^2}{2!}g'_i + \frac{(x_i - x_{i-1})^3}{3!}g''_i + \cdots\right]\right|, \tag{3.52}$$

or

$$\epsilon_n^\sigma = \left|\sum_{i=1}^{n} \left[\sum_{j=2}^{\infty} -1^{(j-1)}\frac{(x_i - x_{i-1})^j}{j!}g_i'^{(j-1)}\right]\right|, \tag{3.53}$$

where

$$g'^{(j)} = \frac{\partial^j g}{\partial x^j}, \tag{3.54}$$

and ϵ_n^σ is the total error up to that point. If only the leading term in the error is used for each grid point then equation 3.53 shows that the error grows linearly[†] with spatial index, i.

In light of this, the averaging of the forward and backward solutions to the continuity equation, $\overrightarrow{\overrightarrow{u_i}}$ and $\overleftarrow{\overleftarrow{u_i}}$, was such that the weighting of each drops off linearly as the solutions moved away from their start points. This is illustrated in Figure 3.3. We can write the averaged solution for the radial velocity, $\bar{u}$, as

$$\bar{u}|_{i=2,N-1} = \left(\frac{i-2}{N-3}\right)\overleftarrow{\overleftarrow{u_i}} + \left(\frac{N-1-i}{N-3}\right)\overrightarrow{\overrightarrow{u_i}}\ . \tag{3.55}$$

It is important to consider what this scheme will and will not do. It *will*, for example, seek to calculate a radial velocity distribution consistent with the boundary conditions which, together with the vertical velocity, satisfies the continuity equation, but it *will not* always generate a non-divergent flow. This is not as unfortunate as it might seem and can be readily explained. Consider a system in which the fluid is at rest and in which there is no temperature variation. Now, if at some instant we impose some vertical velocity at the centre of the fluid (quite how is beside the point), we might then think of using edge-weighted cross-over relaxation to calculate the ensuing radial velocity distribution. The resulting velocity field is shown in Figure 3.4 (a). It is apparent that fluid is being moved

[†] $\epsilon_n^\sigma \propto (n)$

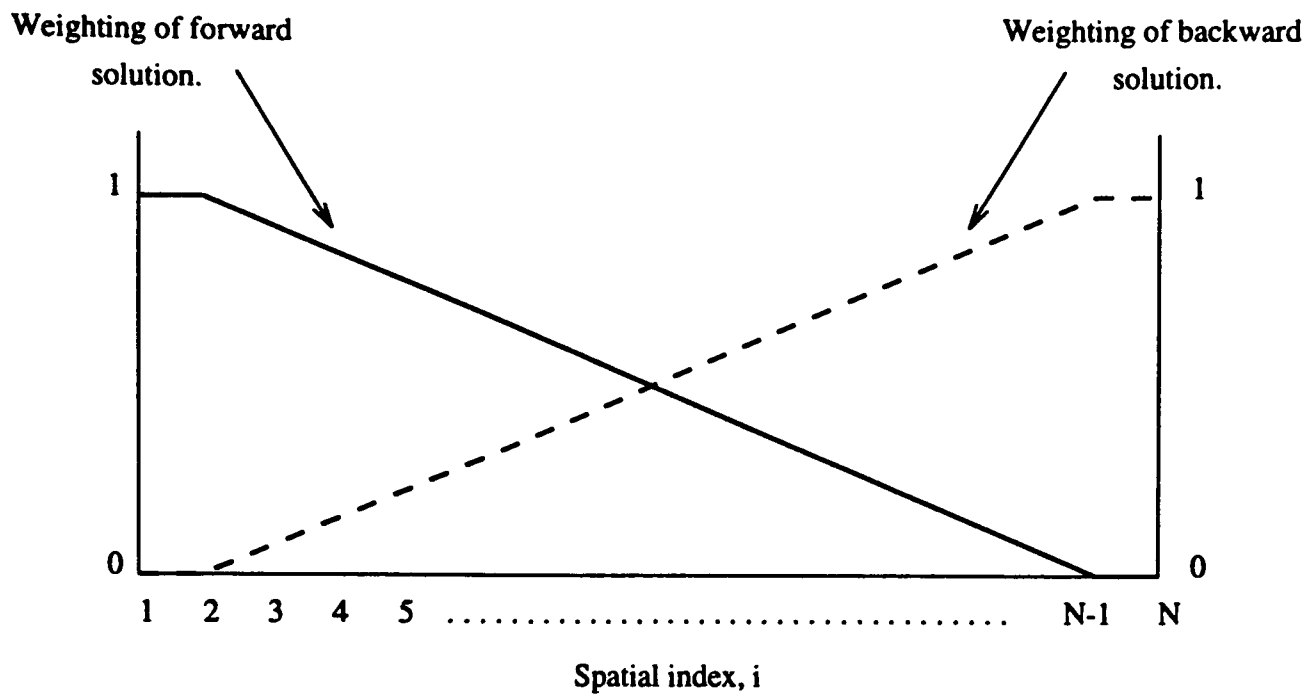
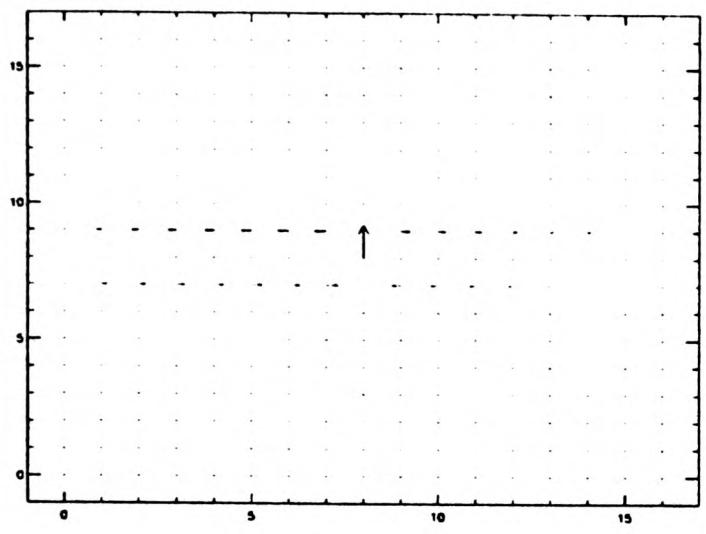


Figure 3.3: The linear variation of the weighting used for the forward and backward solutions to the continuity equation for the internal nodes. The boundary points are at $i = 1$ and $i = N$.

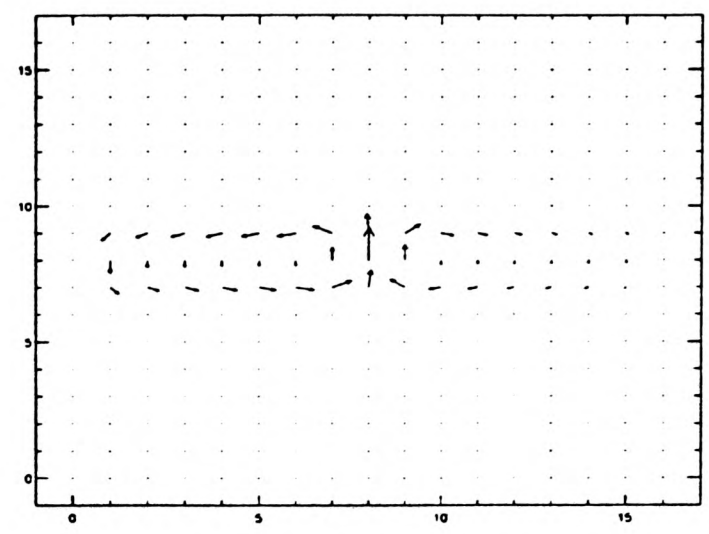
out towards the boundaries above mid-height, and being sucked in below mid-height, but the only vertical velocity is the rising at the centre. This is obviously incorrect as it implies a net upward flow for the whole domain. In order to improve the non-divergence of the flow, we can take the radial velocity which has just been calculated and use it in the edge-weighted cross-over relaxation scheme to find a new estimate of the vertical velocity field. The new velocity field is shown in Figure 3.4 (b), and the two re-circulating cells look rather more physically plausible; there is no exact solution *per se*. The left/right asymmetry in the solution is a result of the metric terms arising for the use of cylindrical polar coordinates (the axes are re-scaled radius and height). Both passes of the scheme have made an improvement in the continuity of the velocity field, and so we might think of reiterating the whole process, alternately using the latest estimates of the radial and vertical velocity field, until the solution converges to some "correct" solution. Figures 3.4 (b) and (c) show the velocity field after a further two passes. The solution can be seen to be settling down to a fully non-divergent state. However, two points must be noted:

1. This scheme tends to lead to the generation of relatively strong velocity components parallel to the boundaries at grid points near the boundary. This is due to the line-by-line method used in solving the equation, and can subdued by the judicious use

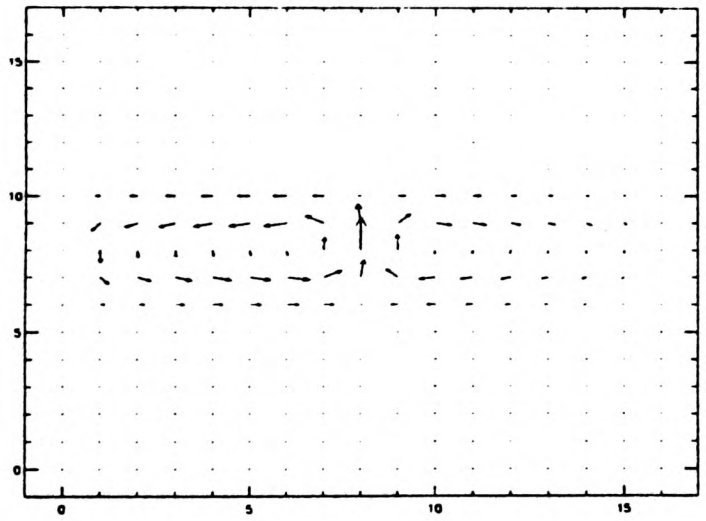
(a) One pass.



(b) Two passes.



(c) Three passes.



(d) Four passes.

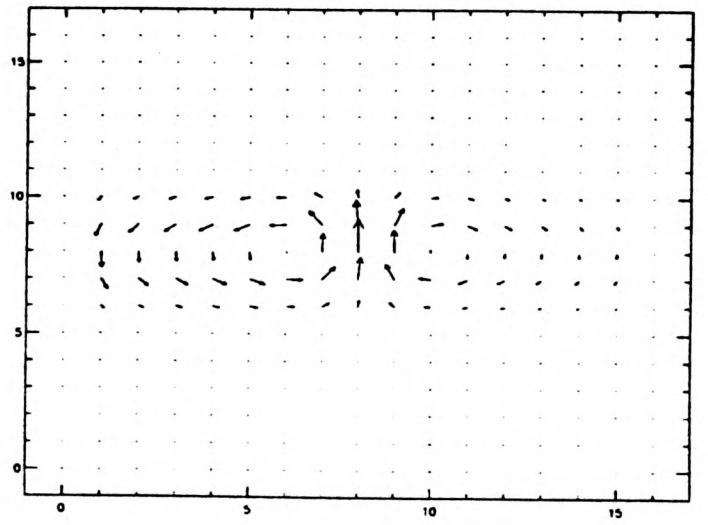


Figure 3.4: Multiple passes of the edge-weighted cross-over relaxation scheme can be used to try and force the system toward a non-divergent velocity field. Here, the initial velocity was a delta function in the vertical at the centre of the domain. (a), (b), (c) and (d) correspond to one, two, three and four passes. Note the development of a downward flow near the inner and outer walls (see main text).

of an artificial viscosity term (see Chapter 6).

2. If a pressure correction scheme (as described in the previous chapter) is adopted, then the the edge-weighted cross-over relaxation will not be needed for the calculation of the velocity, as pressure correction always gives a non-divergent velocity field.

The present version of SARA uses two passes to calculate the velocity at each time step. There would seem to be little point in slowing the code down by using more passes when the pressure terms constitute the main source of inaccuracy.

The same method, edge-weighted cross-over relaxation, can be used to calculate the meridional stream function from the radial and vertical velocity components. The stream function is a useful diagnostic, especially for investigating the radially overturning part of the flow. It is possible to solve two different equations for the stream function, $\psi_r = w$ and $\psi_z = -u$. The first equation is solved line-by-line along rows of grid points, and the second up and down the columns of grid points. In principle both methods should give the same result, and so either is suitable. In practice, they are slightly different and so both are calculated and their mean used as the diagnostic. The averaged stream function can be compared with the meridional velocity vectors (see for example, the validation section later in this chapter), and seen to be in very good agreement.

Although in a version of SARA using pressure correction, this scheme would not be needed for model dynamics, it would still be useful for producing meridional stream function diagnostics.

3.7 Validation.

Before the model can be used in earnest it has to be validated. This is achieved by using the model to solve problems for which the solution is already known, and ensuring that there is agreement between what is expected and what is observed. When any discrepancy has been accounted for and eliminated, the model can be assumed to be working properly

and the solutions it generates can be assumed right. During its development, SARA has undergone a three phase validation, the third phase of which is still being carried out at the time of writing.

3.7.1 Primary validation.

The first stage in the validation procedure was to make sure that the control volume finite element routine was working properly, before the rest of the model was constructed around it. Since this part of the code deals with diffusion and fluxes, it seemed appropriate to set up a diffusion problem as a test case. A hypothetical situation was simulated in which an annular cavity was filled with water. The inner and outer radii were set to 9.000 and 9.055 m respectively. This meant that the effect of curvature was sufficiently small to be ignored in the analytical solution. Solving for the analytical solution was then quite simple, whereas for a solution in fully cylindrical geometry it would have been rather more complex (Crank, 1956). The water was, however, not allowed to flow, so that the only way heat could be transferred was by diffusion. The system was assumed to be initially isothermal at a temperature of 21 degrees and then at one instant both the side walls were set to 17 degrees; the top and bottom boundaries were perfect insulators. The primary reason for the choice of this problem was that it was easy to obtain an analytical solution (Eq. 3.60) for the temperature distribution as it decays with time to an isothermal system at 17 degrees.

The heat equation in the absence of advection can be written

$$\frac{\mathrm{d}C}{\mathrm{d}t} - D\nabla^2 C = 0 \tag{3.56}$$

where C is the temperature difference from the applied temperature at the side walls, and D is a thermal diffusion coefficient. By assuming that the radial, vertical and temporal parts of the solution are separable, it is possible to solve Eq. 3.56 subject to the boundary conditions

$$C = C_0 \quad x_{min} \leq x \leq x_{max}, \quad t = 0 \tag{3.57}$$
$$z_{min} \leq z \leq z_{max}$$

$$C = 0 \quad x = x_{min}, x = x_{max} \quad t \geq 0 \tag{3.58}$$

$$\frac{\partial C}{\partial z} = 0 \quad z = z_{min}, z = z_{max} \quad t \geq 0 \tag{3.59}$$

where C_0 is the temperature of the initial isothermal state. This is most easily done by using a Fourier series representation of the vertical and radial parts and an exponentially decaying form for the time dependant part (see, for example Crank, 1956). Applying the normal procedure then gives the following solution to Eq. 3.56

$$C = \frac{C_0}{\pi} \sum_{m=1}^{\infty} \left[\frac{1 + (-1)^{m+1}}{m} \right] \sin \left[\frac{m\pi x}{x_{max} - x_{min}} \right] \exp \left[\frac{-m^2 \pi^2 D t}{(x_{max} - x_{min})^2} \right] \tag{3.60}$$

where $x = 0$ is the inner wall. This solution is shown by the dotted curves in Figure 3.5 for values of D and t to represent heat flow through "solid" water for 100, 200, 500 and 1000 seconds. The solid lines in Figure 3.5 show the solution obtained using the control volume finite element code from SARA for the same problem. The crosses show the position of the radial grid points, the grid having been stretched so as to have a spacing of less than 1 mm at the boundary. In both these figures, the radial distance from the inner wall was used as the horizontal axis. The agreement between the analytical solution and model results was excellent in each case. In the light of this we can conclude that the CVFEM part of the model has been correctly coded.

3.7.2 The radial overturning problem.

The second phase in the validation process was to make sure that the advection terms in the problem were being calculated properly. To do this we considered the radially overturning flow which is always present in the annulus, but which is the only flow present in the absence of rotation. SARA was run with the rotation rate set to zero so that this would be the only flow to develop.

The model was set up with the inner cylinder being cold (17 degrees) and the outer cylinder being hot (21 degrees), with the top and bottom boundaries thermally insulating. The fluid was initially at 19 degrees before the wall heating/cooling was turned on. By this stage there was no longer any need to stay in rectangular coordinates and so a cylindrical polar coordinate system was adopted. When the differential heating was first applied

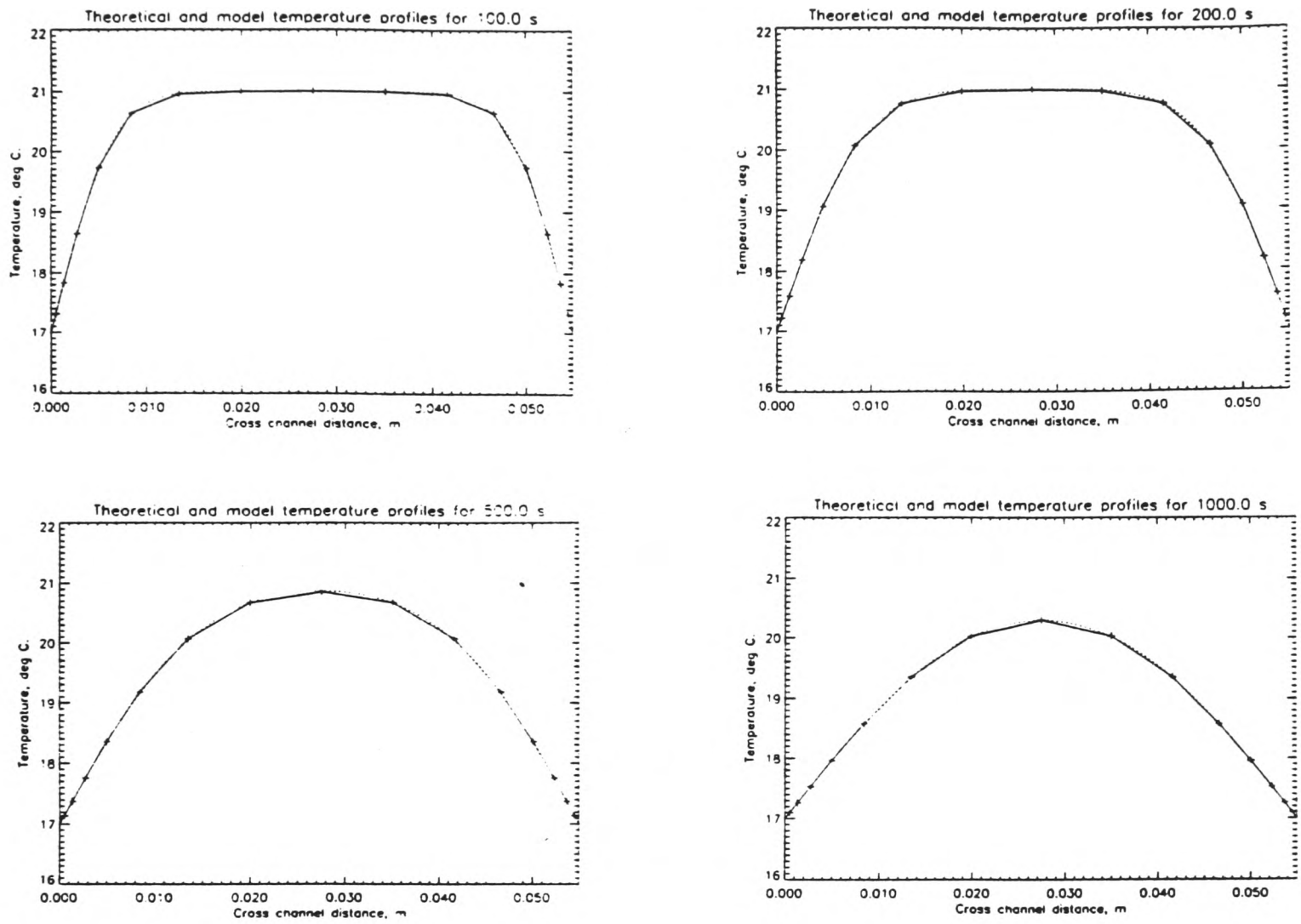


Figure 3.5: Validation results for thermal diffusion in the absence of advection. The four plots show temperature (deg. C) as a function of cross channel distance (m) after 100, 200, 500 and 1000 s. In each plot the dotted line shows the analytical solution, and the solid line shows the model results. The crosses show the the radial position stretched grid points.

heat began to diffuse into the fluid close to the outer cylinder and out of the fluid close to the inner cylinder. The buoyancy force caused the warm fluid to rise and the cooler fluid to sink close to the boundaries, and this in turn drove an inward flow across the top of the annulus and an outward flow across the bottom. The overall effect is known as radial over-turning and has been the subject of much study (see, for example, Gill, 1966; McIntyre, 1968; Hignett *et al.*, 1981; and Read, 1986a).

If it is assumed that the thermal boundary layers are convective (*i.e.* buoyancy driven) and much thinner than the annular gap, that the cross-sectional aspect ratio is of order unity and that the ratio of inner to outer radius is also of order unity, then a scale analysis can be used on the zonal vorticity equation in the boundary layer (McIntyre, 1968). When this is done, a thermal boundary layer thickness, l can be derived from a balance of forces between buoyancy and thermal dissipation

$$l = \left(\frac{\nu\kappa(z_{max} - z_{min})}{g\alpha\Delta T} \right)^{\frac{1}{4}}, \tag{3.61}$$

where ν is the kinematic viscosity, κ is the thermal diffusivity, g is gravity, α is the thermal expansivity and ΔT is the applied temperature difference. If the relevant values for SARA are used in this expression, we can expect a thermal boundary layer thickness of a few millimetres.

Figure 3.6 shows a typical velocity plot from SARA for the radial over-turning problem. The figure shows a velocity vector at each of the regularly spaced grid points (or nodes) used for the run, and the general structure of the circulation can be clearly seen. The calculation domain (grid points 0 to 16) extended radially from 0.025 m to 0.080 m, and vertically from 0.000 m to 0.140 m. We can see that the flow is concentrated in a thermal boundary layer which is about two or three grid spacings wide. This corresponds to a physical width between about 6 and 9 mm which is in good agreement with the value determined by the scale analysis. We were able to infer from this agreement that the basic mechanism for velocity calculation was working as expected, and that the CVFEM was being used correctly to advect the heat around in the fluid.

There are two further points of interest which are illustrated by Figure 3.6. Firstly, there is a general downward motion in the fluid which can be seen everywhere except near

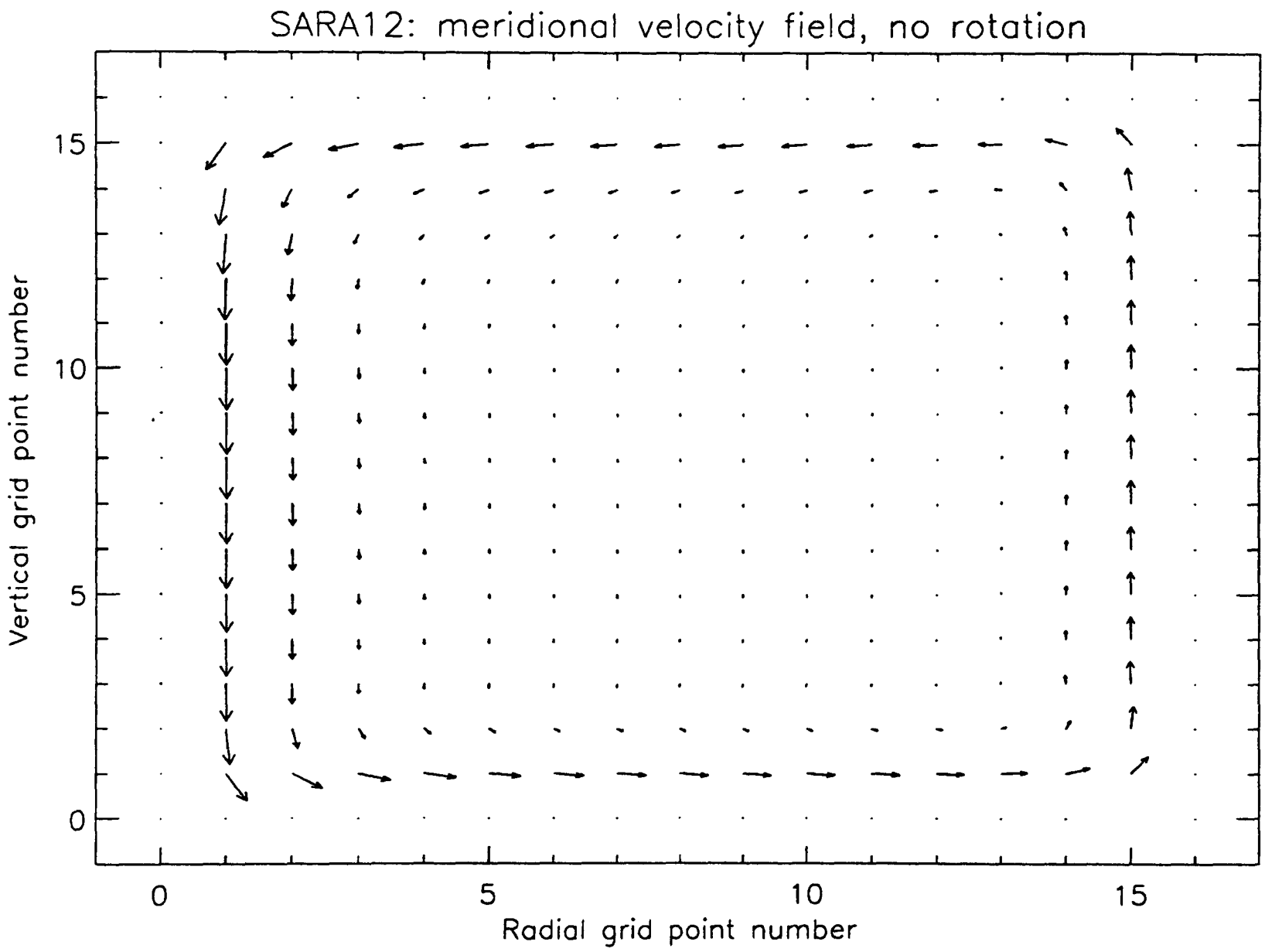


Figure 3.6: Validation results for axisymmetric radial over-turning. The model was used with 17 grid points in each direction, 0 and 16 being on the boundaries. The largest arrow corresponds to a velocity of -0.875 mm/s. The rising and sinking motion takes place in narrow thermal boundary layers.

the outer wall, and is particularly noticeable in the return flow close to the horizontal boundaries. This extraneous motion is due to the way in which the buoyancy forcing is calculated in the vertical momentum equation, Eq. 3.17. The mean temperature, $\bar{T}$, is a constant for the fluid and it is the difference between the temperature at some point and this value which determines the buoyancy forcing at that point. The value of $\bar{T}$ is essentially arbitrary, because as long as it remains constant, the pressure equation, Eq. 3.20, will generate vertical pressure gradients to balance the system out. However, SARA uses artificial gradient pressure boundary conditions (for an explanation of this, see Section 2.4.1), and the pressure grid is a lower resolution than the velocity grid. These two factors combine with the result that there can be no variation in the vertical pressure gradient on the scale of the velocity grid, near the horizontal boundaries. This rather subtle problem would not arise if the pressure correction method was incorporated into SARA , because it would not then be necessary to impose the artificial boundary conditions.

The remaining point about Figure 3.6 is to note the horizontal asymmetry of the flow. This is due to the cylindrical geometry of the annulus system, and the fact that the grid used for this example had a regular spacing in radius. This meant that the elements nearer the outer wall had a larger effective volume and so less fluid would flow through each one; hence the smaller velocity arrows and the slightly narrower thermal boundary layer.

3.7.3 Tertiary validation.

In this section, some results from SARA are used to illustrate the workings and shortcomings of the model.

Since we know that the pressure field is going to be the main problem, it is more fruitful to ignore the terms in the model where the pressure is used, and to investigate the development of the resulting hydrostatic system. Although these runs can not be expected to produce realistic results, they will enable us to see how well the rest of the

model works. Because SARA is an axisymmetric code, there is no pressure gradient term in the zonal momentum equation (Eq. 3.16). Also, because of the use of the continuity equation for the determination of the radial velocity, there is no explicit radial pressure gradient term in the model equations. The only place where the pressure occurs in the system equations is in the vertical momentum equation, Eq. 3.17, and so omitting the pressure (or, rather, assuming that the gradient of the perturbation pressure is zero) will have its effect initially on the calculation of the vertical velocity.

Figures 3.7, 3.8 and 3.9 show the results of three runs of SARA with the vertical pressure term dropped. The three runs used identical parameters except for the rotation rate, which was 0.0, 0.5 and 1.0 rad/s respectively. The artificial viscosity parameter referred to in the caption is described in Chapter 6, but its value was the same for all the runs presented in this chapter, and it is sufficient here to know that the velocity field was slightly smoothed. In the run which produced the flow shown in Figure 3.7 the rotation rate was zero and so no zonal flow developed. In the next two runs the rotation produced a pair of zonal jets, one prograde near the lid and one retrograde near the base. The structure of the zonal flow did not vary with the rotation rate, only the magnitude. This was because the code was only run for a small number of time steps (so that the flow did not become unstable), and the dominant term in the zonal momentum equation (Eq. 3.16) over this time scale was the Coriolis acceleration, which is proportional to the rotation rate. The rotation does, of course, also force the radial flow via the Coriolis force; this does not happen explicitly in SARA because the radial velocity is inferred from the vertical velocity and the non-divergence condition.

Figures 3.10, 3.11 and 3.12 show a further three runs of SARA which are all for the same rotation rate (0.5 rad/s), but for increasing numbers of time steps. Figure 3.8 can also be considered part of this set of results, and would come after Figure 3.10. The time step used for these runs was 0.1 seconds. In Figure 3.10 the control volumes associated with the grid nodes nearest the walls pass through the boundary elements. The shape functions in these elements have the temperature varying linearly with radius, and so integrating across the control volume leads to numerical diffusion, and the grid points adjacent to the vertical boundary will gradually adjust their temperature in the corresponding sense.

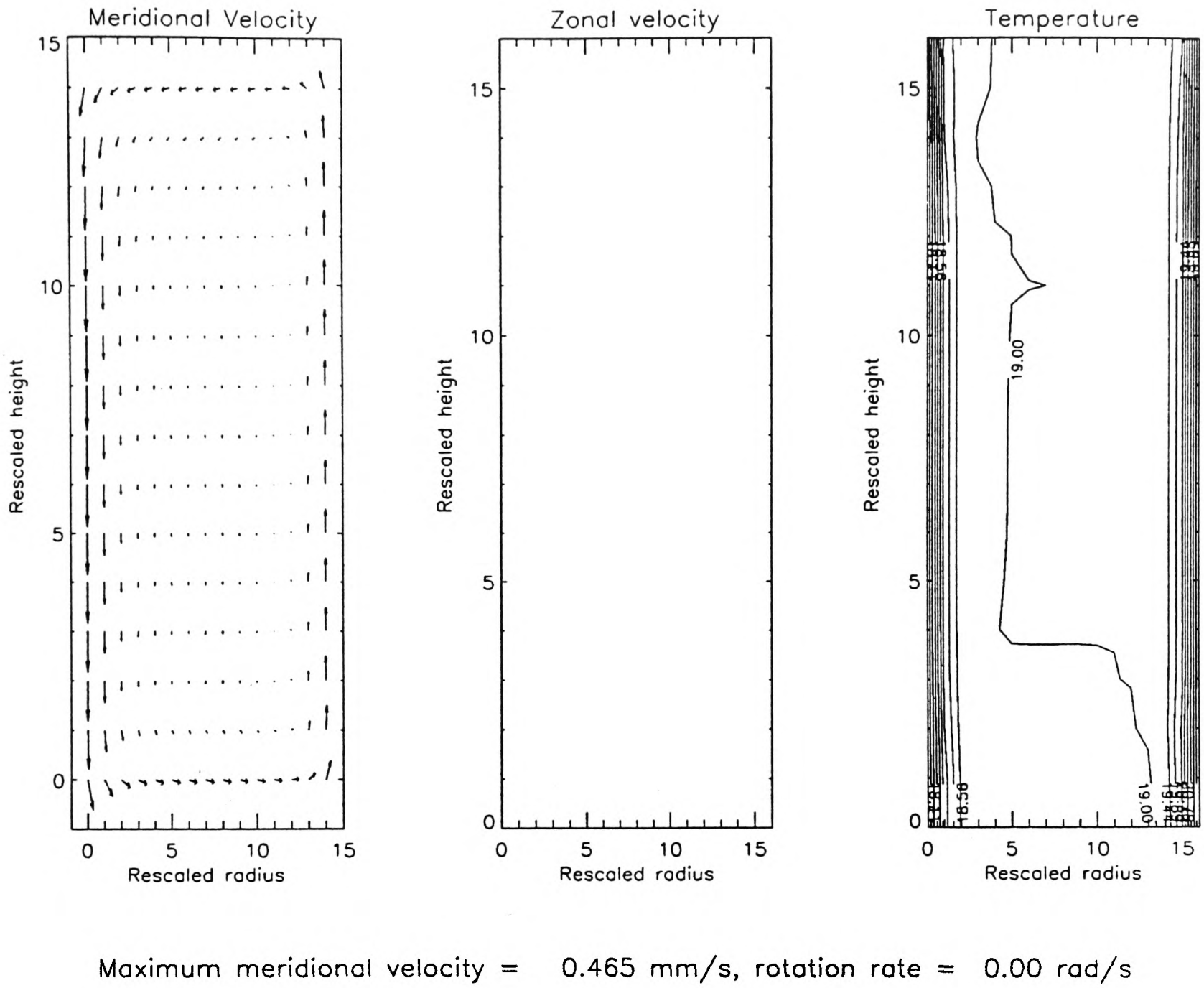


Figure 3.7: Validation results from SARA. This run was for 15 time steps with an artificial viscosity parameter of 0.2. As the rotation rate is zero, there is no zonal flow.

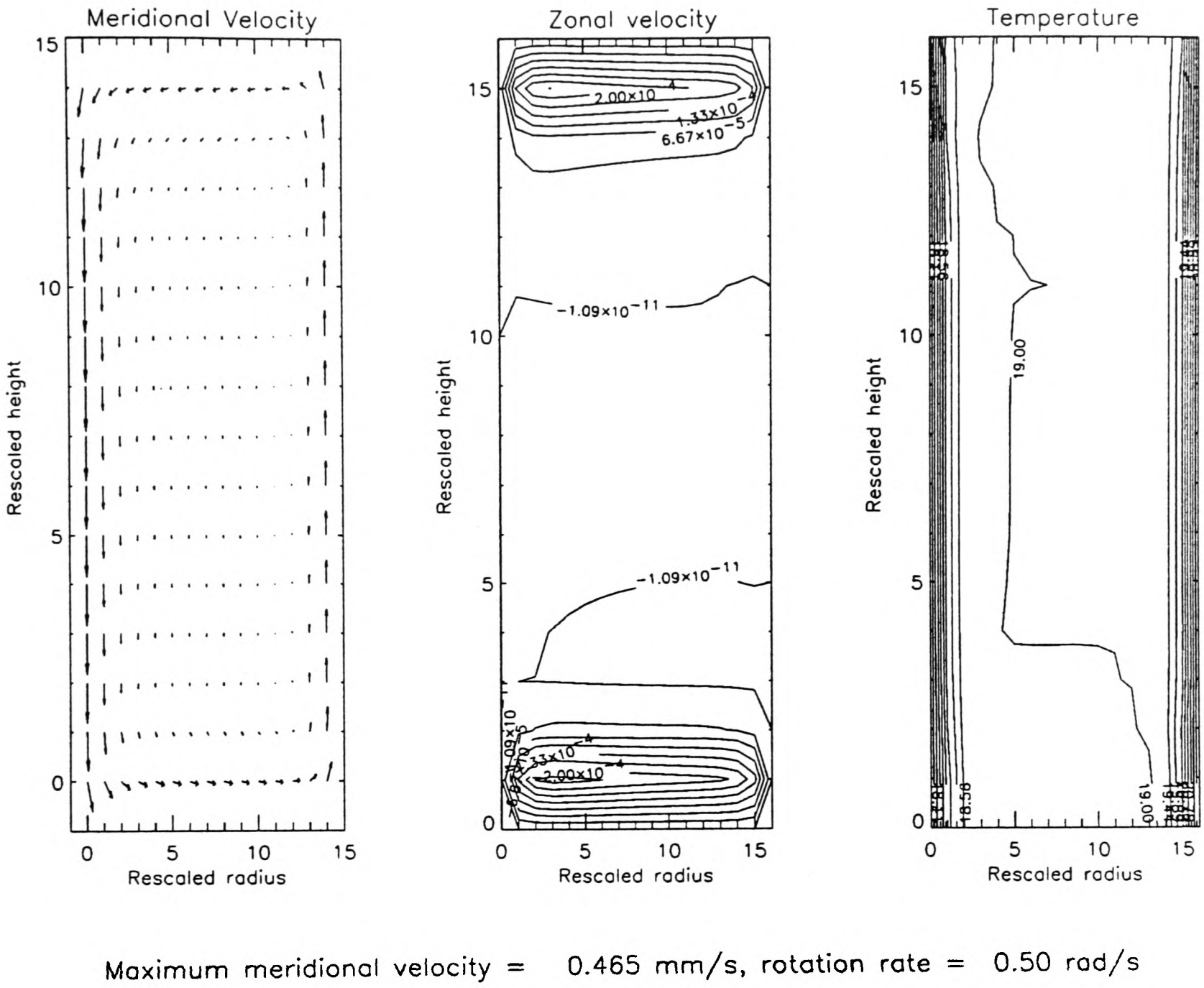


Figure 3.8: Validation results from SARA. This run was for 15 time steps with an artificial viscosity parameter of 0.2. The results are the same as for the previous case except that two opposing zonal jets have been formed, near the top and bottom boundaries.

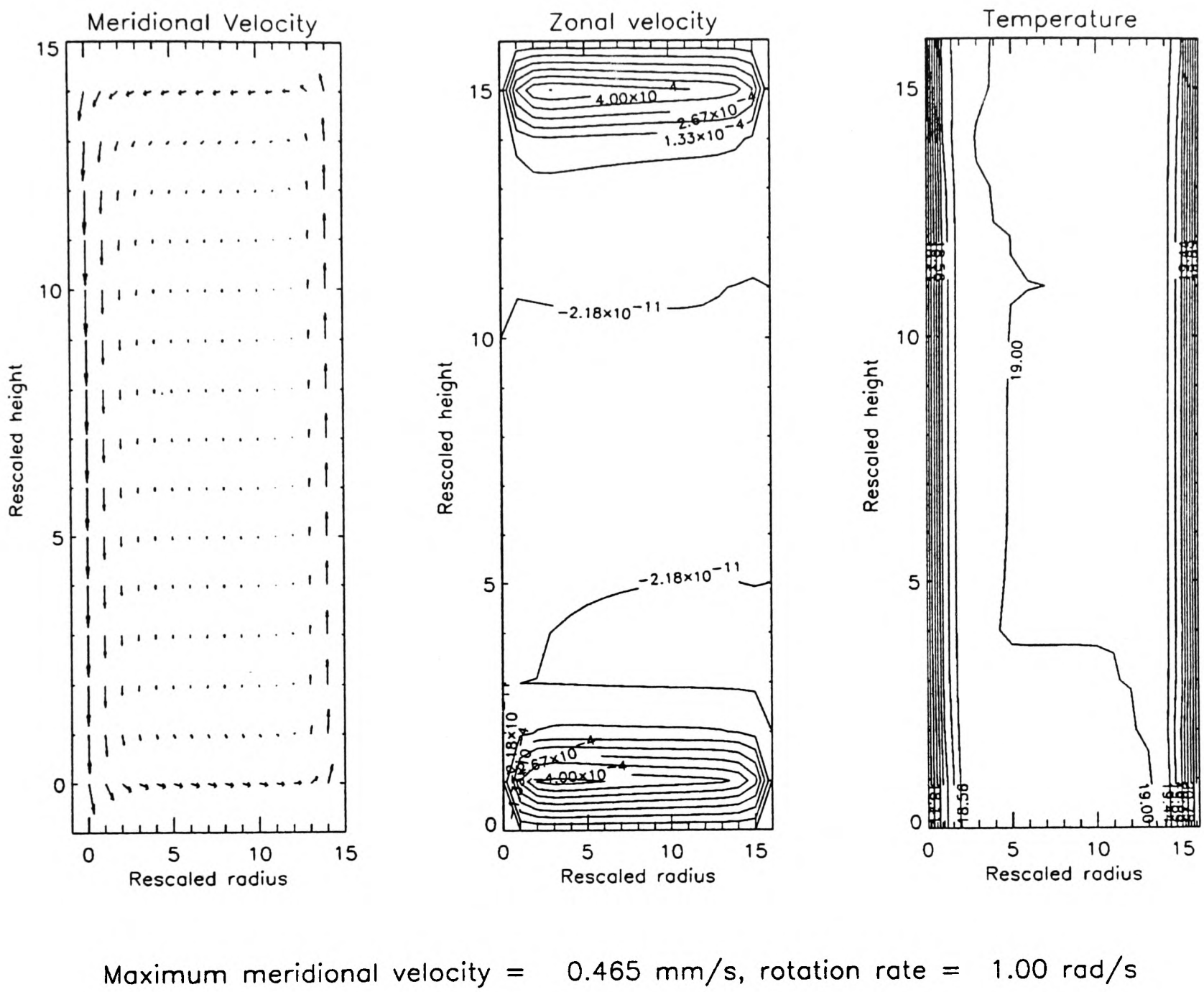


Figure 3.9: Validation results from SARA. This run was for 15 time steps with an artificial viscosity parameter of 0.2. The higher rotation rate used in this run induces stronger jets, but their structure is unaltered from the lower rotation rate run.

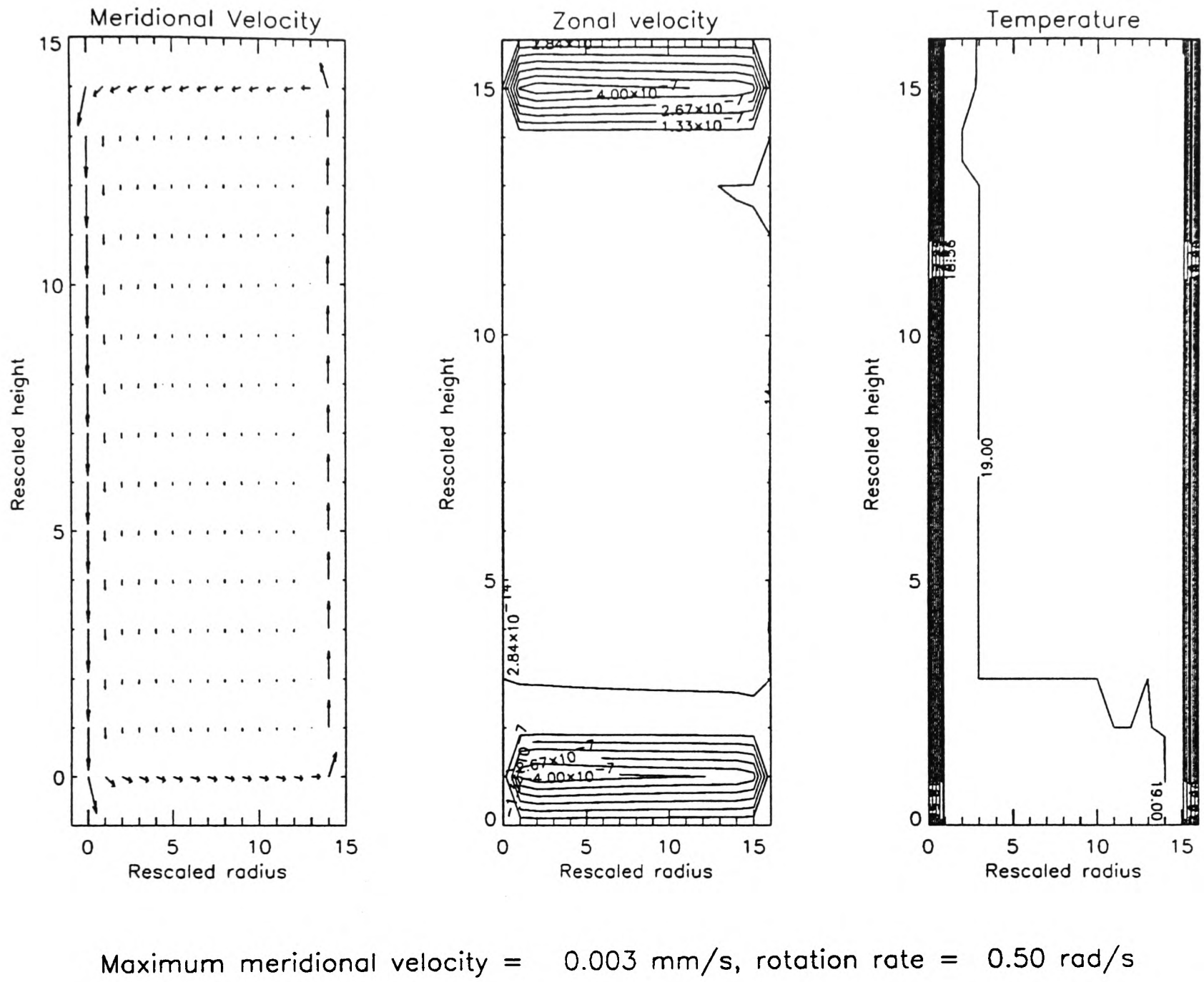


Figure 3.10: Validation results from SARA. This run was for 5 time steps with an artificial viscosity parameter of 0.2. At this early stage, the flow is only present close to the boundary because the heat has not yet diffused away from the wall.

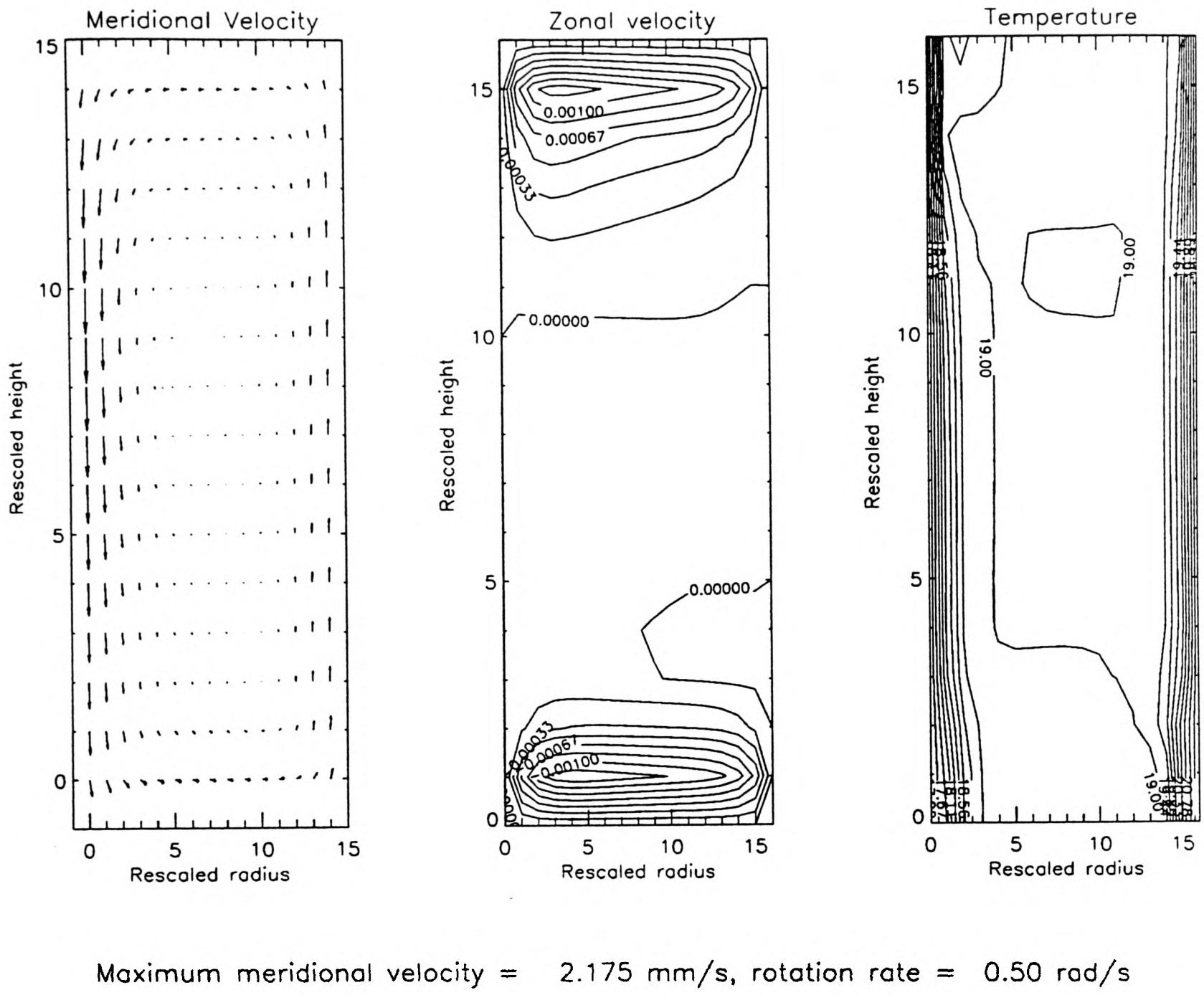


Figure 3.11: Validation results from SARA. This run was for 25 time steps with an artificial viscosity parameter of 0.2. The meridional flow has spread in from the boundaries due to the artificial viscosity, and the heat has begun to diffuse in from the walls.

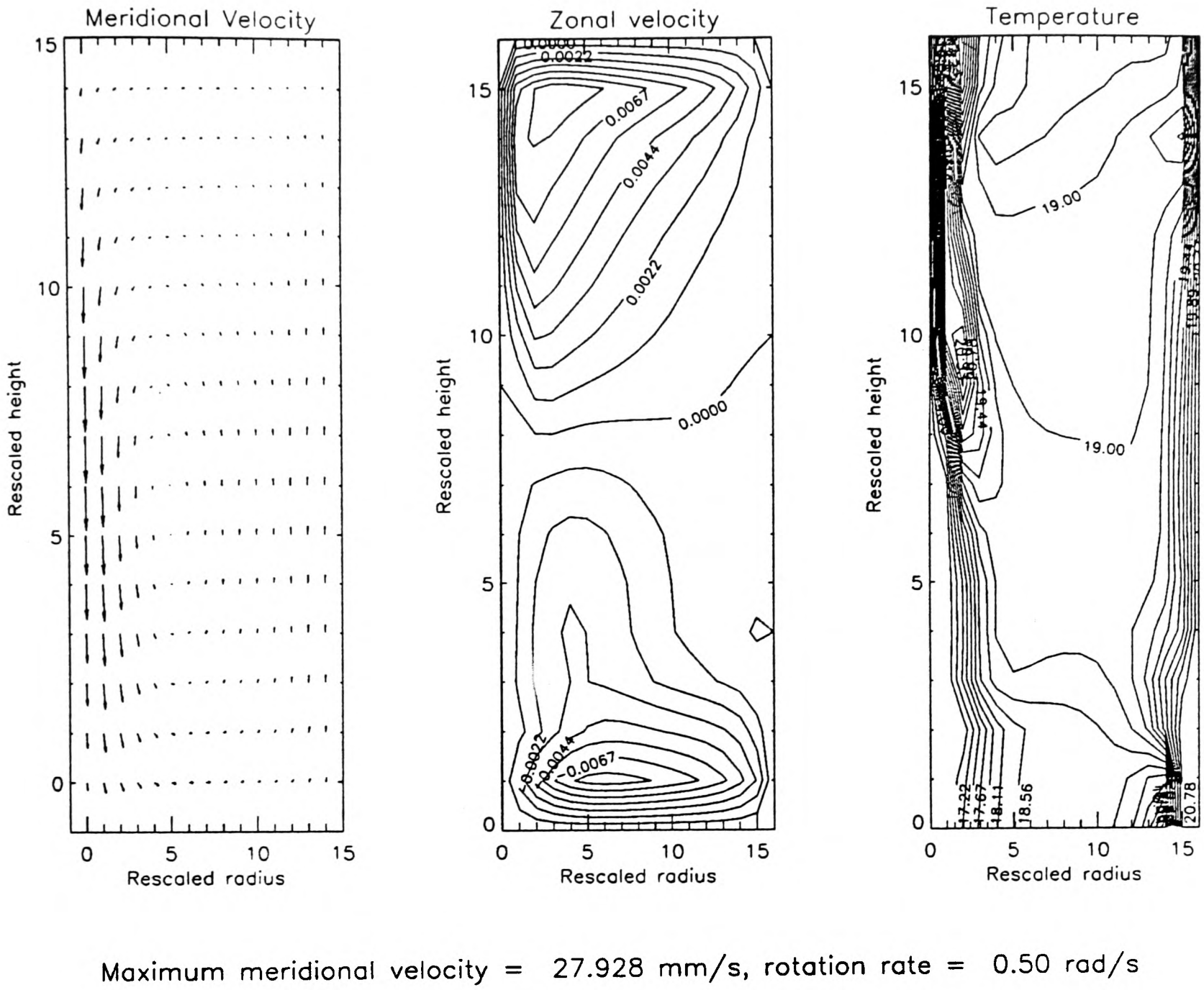


Figure 3.12: Validation results from SARA. This run was for 35 time steps with an artificial viscosity parameter of 0.2. The flow has become unstable as shown by the temperature contours and the unrealistically large downward flow near the inner wall.

Using a stretched grid will not avoid this numerical diffusion, but it will limit the range in r over which it is felt. As the fluid near these boundaries warms up or cools down, the buoyancy force drives a radial overturning flow as described in the previous section. As the simulation is run for longer periods, the numerical diffusion smooths out the temperature field slightly more, and the thickness of the thermal boundary layer begins to increase a little, due partly to the artificial viscosity and partly to the real viscosity. The system does, however, become unphysical (and ultimately unstable) after a short while because of the absence of the necessary pressure gradient terms to balance the vertical advection terms. This can be seen from Figure 3.12.

The use of a pressure correction scheme would be particularly helpful here because, as there is no need for an explicit continuity equation, both the radial and vertical momentum equations can be used with their source terms, and the Coriolis term could be included in the radial equation directly.

It is interesting, if only for the sake of completeness, to examine the effect of keeping the pressure term and watching SARA to see what goes wrong. Figure 3.13 shows the results from a run at a rotation rate of 0.9 rad/s for 15 time steps, but this time the pressure has been left in and normal gradient boundary conditions have been used. The meridional circulation is dominated by two counter-rotating cells separated by a strong outward flow at about 100 mm from the base. This outward flow produces two corresponding zonal jets but the velocity of these are an order of magnitude smaller than that of the meridional flow. It is the pressure field which forces this strong meridional flow. In Figure 3.14 the same meridional velocity field is shown with the pressure field overlaid as contour lines. Near the outer wall there are large vertical gradients in the pressure field; these produce much larger terms than the thermal forcing in the vertical momentum equations and so the structure of the vertical velocity field is dominated by the pressure distribution. In order to satisfy the non-divergence criterion near the outer boundary, strong outward radial flow is produced. Although this flow obviously does not agree with experimentally observed results, we would not expect it to because of the errant pressure field. It is reassuring, however, to note that although the pressure field is not correct, SARA has generated a velocity field which is consistent with the pressure field and which satisfies

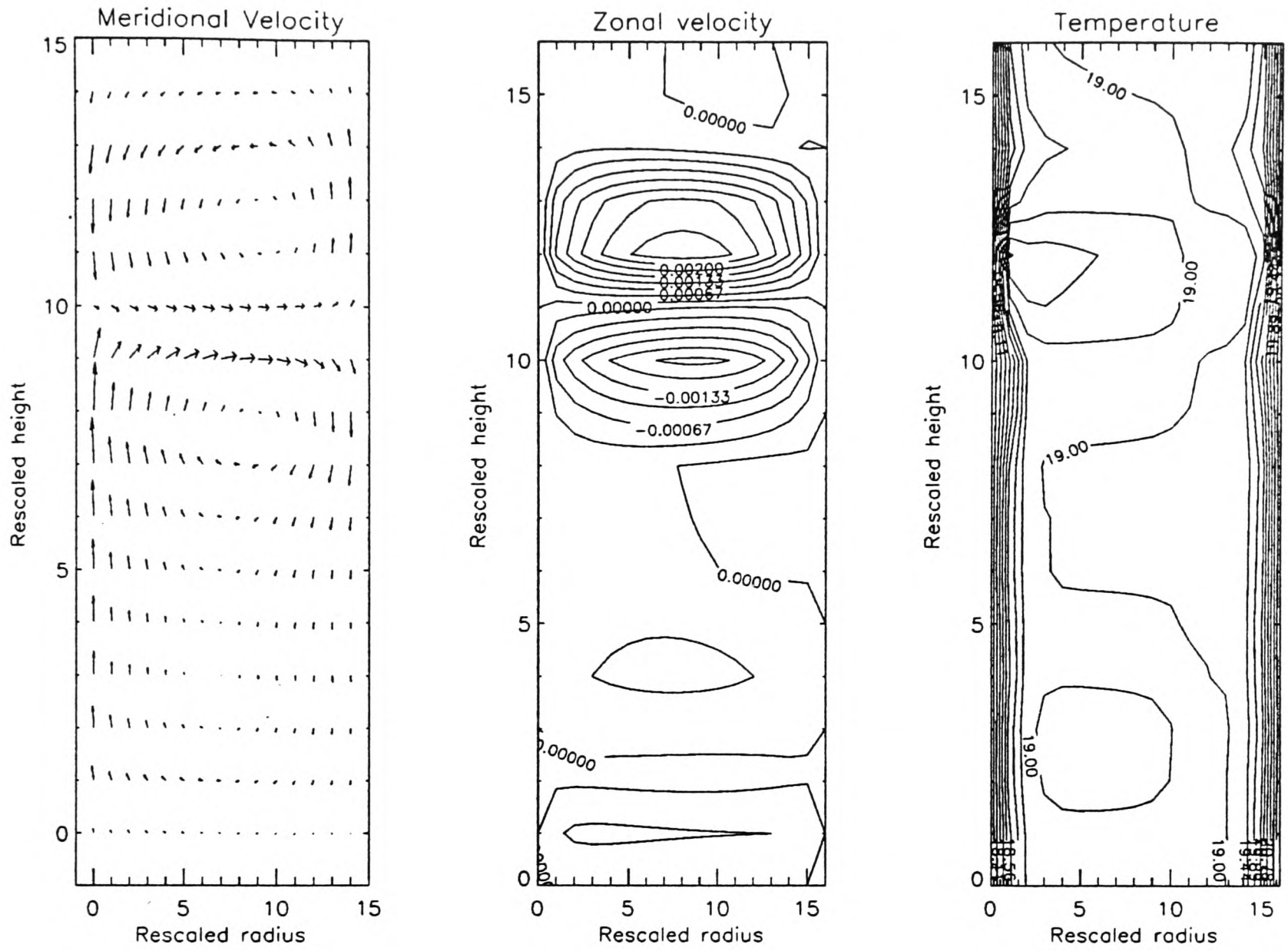


Figure 3.13: Validation results from SARA. This run was for 15 time steps with an artificial viscosity parameter of 0.2. Here the model was run with the errant pressure terms included; the pressure gradients are large and drive the outward radial flow near the top of the annulus.

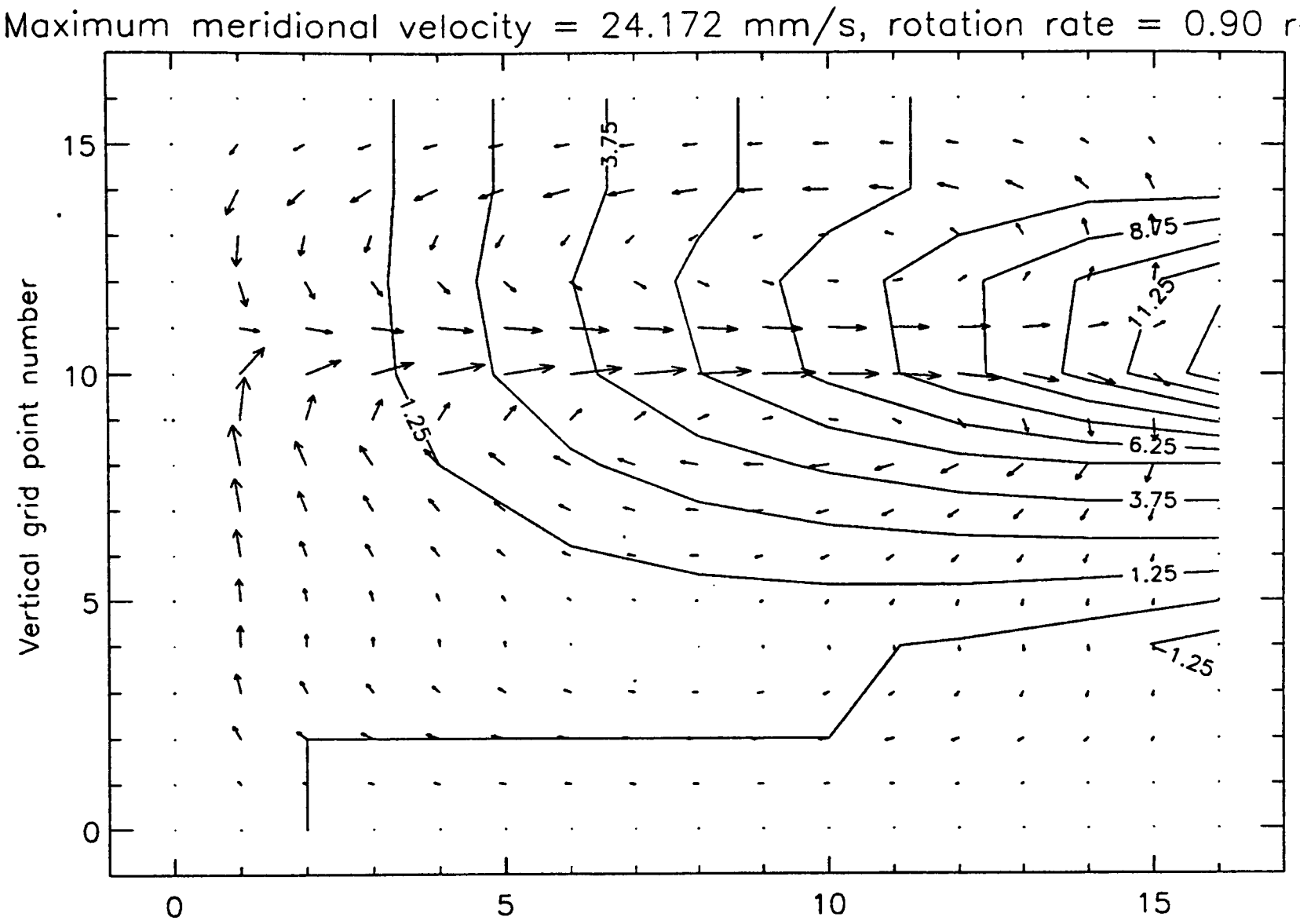


Figure 3.14: The meridional flow shown by the velocity vectors is forced by the pressure field shown by the contour lines. The run parameters are the same here as in the previous figure.

the continuity condition.

In Chapter 6 we discuss the reasons for SARA 's failure and consider the scope for possible improvements and refinements.

Chapter 4

Quasigeostrophic instability theory.

In Chapter 1 we discussed the idea of the stability of a background flow in the annulus or atmosphere to various perturbations. We observed that if the flow is unstable to a particular disturbance it will begin to grow, and in the limit of the validity of our linearization (*i.e.* while the disturbance is 'small'), will dominate the observed state of the system. In this chapter we outline the mathematical description of this linearized normal mode perturbation analysis, and show how we can use the quasigeostrophic approximation to pose the stability analysis as a numerical eigenvalue problem.

Arguably the most important driving mechanism at work for the systems we are considering is sloping convection, or baroclinic instability. As well as briefly discussing the qualitative nature of this mechanism, we describe the Eady problem, which acts as a simple example of baroclinic instability and also provides a useful benchmark for assessing our analysis methods. Towards the end of the chapter, there is a description of INSTAB , a numerical scheme which we use to analyse the stability of various flows and configurations. Most of the techniques used in coding INSTAB were standard; the more interesting elements of the algorithm are summarised below.

4.1 The quasigeostrophic approximation.

In order to examine the stability of axisymmetric flows in the rotating annulus to various perturbations, we begin by simplifying the equation set by making some additional approximations†. These will ultimately enable us to describe the whole system by one equation. The following description is intended only as an outline, and for further details the reader is referred, for example, to Pedlosky (1987) or Gill (1982).

We can write the Navier-Stokes equation as

$$\mathbf{u}_t + \mathbf{u}\cdot\nabla\mathbf{u} = -\frac{\nabla P}{\bar{\rho}} - 2\boldsymbol{\Omega}\wedge\mathbf{u} + \nu\nabla^2\mathbf{u} + \mathbf{g}\alpha(T-\bar{T}), \tag{4.1}$$

where $\mathbf{u}$ is the velocity vector and the other terms are as defined in Chapters 1 and 3. If we take the *curl* of the Navier-Stokes equation, we have

$$\omega_t + \mathbf{u}\cdot\nabla\omega - \omega\cdot\nabla\mathbf{u} = 2(\boldsymbol{\Omega}\cdot\nabla)\mathbf{u} + \nu\nabla^2\omega - \alpha\mathbf{g}\wedge\nabla(T-\bar{T}), \tag{4.2}$$

where ω is the relative vorticity. Next we assume that the main body of the fluid is approximately in geostrophic balance; that is, that the Coriolis force is approximately balanced by the pressure gradients. If we say that the velocity $\mathbf{u}$ is equal to the sum of a geostrophic part, $\mathbf{u}_g$ and an ageostrophic part, then we can write the geostrophic balance condition as

$$\mathbf{u}_g = \hat{\mathbf{k}}\wedge\nabla\psi. \tag{4.3}$$

Here, ψ is the geostrophic stream function, $\psi = P/2\Omega\bar{\rho}$, and $\hat{\mathbf{k}}$ is a unit vector in the vertical. It then follows that we can substitute the horizontal Laplacian of ψ for ω in Eq. 4.2, giving

$$(\nabla_h^2\psi)_t\hat{\mathbf{k}} + (\mathbf{u}_g\cdot\nabla)\nabla_h^2\psi\hat{\mathbf{k}} - [(2\Omega+\omega)\cdot\nabla]\mathbf{u} = -\alpha\mathbf{g}\wedge\nabla(T-\bar{T}) + \nu\nabla^4\psi. \tag{4.4}$$

The next simplification is to assume that the Rossby number, Ro is small ($i.e \ll 1$). If we write the Rossby number as $Ro = U/2\Omega L$, where U and L are characteristic velocity and length scales for the annulus, we then find $Ro \sim 10^{-1}$ and so the assumption is

†The description of the system equations given in Section 3.2 of this thesis lists the standard approximations which are used in describing annulus flows.

justified. This means that the background rotation rate, Ω, is much larger than the relative vorticity, and so ω can be omitted form the third term on the left hand side, giving

$$(\nabla_h^2\psi)_t\hat{\mathbf{k}} + (\mathbf{u}_g \cdot \nabla)\nabla_h^2\psi\hat{\mathbf{k}} - 2[\Omega \cdot \nabla]\mathbf{u} = -\alpha\mathbf{g} \wedge \nabla(T - \bar{T}) + \nu\nabla^4\psi. \tag{4.5}$$

The geostrophic velocity, $\mathbf{u}_g$, only has components in the horizontal, and so any vertical motion must be purely ageostrophic. Hereafter, we shall use w to mean the ageostrophic vertical part of the motion. The advection/diffusion equation for the temperature can then be written

$$T_t + \mathbf{u}_g \cdot \nabla T + wT_z = \kappa\nabla^2 T. \tag{4.6}$$

We can replace the vertical temperature gradient in this equation by $-N^2/g\alpha$, where $N^2(z)$ is the Brunt-Väisälä frequency. This then gives us

$$T_t + \mathbf{u}_g \cdot \nabla T - \frac{wN^2}{g\alpha} = \kappa\nabla^2 T, \tag{4.7}$$

or using the Jacobian, J, and writing Eq. 4.7 as an expression for w

$$w = \frac{g\alpha}{N^2}(T_t + J(\psi, T) - \kappa\nabla^2 T). \tag{4.8}$$

We can form the vertical derivative of this equation, and substitute from the hydrostatic relation, $\psi_z = -\frac{g\alpha}{2\Omega}(T - \bar{T})$, for the temperature, to give

$$\frac{\partial w}{\partial z} = -\frac{2\Omega}{N^2}(\psi_{zzt} + J(\psi, \psi_{zz}) - \kappa\nabla^2\psi_{zz}) + \frac{2\Omega N_z^2}{N^2 N^2}(\psi_{zt} + J(\psi, \psi_z) - \kappa\nabla^2\psi_z). \tag{4.9}$$

We can consider the three components of Eq. 4.5 separately. The radial and zonal equations are

$$2\Omega\frac{\partial u}{\partial z} = -g\alpha\frac{\partial T}{\partial y}, \tag{4.10}$$

and

$$2\Omega\frac{\partial v}{\partial z} = g\alpha\frac{\partial T}{\partial x}, \tag{4.11}$$

respectively (for clarity we have used a local rectangular Cartesian coordinate system with x being the outward radial (or equatorward) direction and y being the zonal direction). These are the thermal wind equations and describe the dependence of the vertical variation

of the radial and zonal velocity on the horizontal variation of the temperature. The vertical component of Eq. 4.5 is

$$(\nabla_h^2\psi)_t + J(\psi, \nabla_h^2\psi) - 2\Omega\frac{\partial w}{\partial z} = \nu\nabla^4\psi. \tag{4.12}$$

If we substitute for $\frac{\partial w}{\partial z}$ from Eq. 4.9 we can write

$$\frac{\partial q}{\partial t} + J(\psi, q) = \nu\nabla^4\psi + 4\Omega^2\kappa\frac{\partial}{\partial z}\left(\frac{\nabla^2\psi_z}{N^2}\right), \tag{4.13}$$

where q, the *quasigeostrophic potential vorticity*, is defined as

$$q = \nabla_h^2\psi + 4\Omega^2\frac{\partial}{\partial z}\left(\frac{\psi_z}{N^2}\right). \tag{4.14}$$

∇_h^2 is the horizontal part of the Laplacian.

q is related to the Rossby-Ertel potential vorticity, Π_E, (Ertel 1942, Rossby 1940), but is not simply the quasigeostrophic approximation to it. Rather, whereas Π_E is conserved following the total flow, we can see from Eq. 4.13 that q is conserved following the geostrophic part, $\mathbf{u}_g$, of an inviscid, adiabatic flow ($\nu = 0, \kappa = 0$), if the quasigeostrophic approximation is valid.

4.2 Normal mode analysis.

We aim to use the quasigeostrophic potential vorticity equation to study the spatiotemporal stability of various flows.

Unfortunately, the nonlinearity of the instability problem makes it too difficult to solve in the form given above. In order to overcome this limitation, we constrain our studies to small amplitude perturbations about a prescribed background state, and then linearise the quasigeostrophic potential vorticity equation (or *qpve*) about this state. This means that any growing unstable states which should emerge from the system are only valid solutions whilst they remain small. The stream function and potential vorticity can be written as

$$\psi = \bar{\psi} + \psi' \text{ and } q = \bar{q} + q', \tag{4.15}$$

respectively, where the over-bar denotes the steady state, axisymmetric solution to the *qpve*, Eq. 4.13, and the prime denotes a small perturbation super-imposed on the background state. We also have that

$$\bar{\psi}_{\theta,t} = 0, \quad \bar{q}_{\theta,t} = 0, \tag{4.16}$$

and that

$$\psi', q' \ll \bar{\psi}, \bar{q}. \tag{4.17}$$

If we substitute for ψ and q from Eq. 4.15 into the *qpve*, we get

$$\begin{aligned} q'_t + \bar{q}_t + J(\bar{\psi}, \bar{q}) + J(\bar{\psi}, q') + J(\psi', \bar{q}) + J(\psi', q') = \\ \nu\nabla^4\bar{\psi} + \nu\nabla^4\psi' + 4\Omega^2\kappa\frac{\partial}{\partial z}\left(\frac{\nabla^2\bar{\psi}_z}{N^2}\right) + 4\Omega^2\kappa\frac{\partial}{\partial z}\left(\frac{\nabla^2\psi'_z}{N^2}\right) \end{aligned} \tag{4.18}$$

This equation can be considerably simplified, however, expanding the Jacobians at this stage, and ruling out terms in the equation with azimuthal or temporal derivatives of $\bar{\psi}$ and $\bar{q}$, since these are, *ab initio*, zero. We then have that

$$q'_t + \frac{\bar{\psi}_r q'_\theta}{r} - \frac{\psi'_\theta \bar{q}_r}{r} = \nu\nabla^4\bar{\psi} + \nu\nabla^4\psi' + 4\Omega^2\kappa\frac{\partial}{\partial z}\left(\frac{\nabla^2\bar{\psi}_z}{N^2}\right) + 4\Omega^2\kappa\frac{\partial}{\partial z}\left(\frac{\nabla^2\psi'_z}{N^2}\right). \tag{4.19}$$

where the terms which are the product of two primed variables have been dropped because the primed variables are themselves small. This step is essentially the linearization of the equation.

Since $\bar{\psi}$ and $\bar{q}$ are themselves solutions to the *qpve*, we can set the perturbation terms to zero in Eq. 4.19 to get

$$\nu\nabla^4\bar{\psi} + 4\Omega^2\kappa\frac{\partial}{\partial z}\left(\frac{\nabla^2\bar{\psi}_z}{N^2}\right) = 0. \tag{4.20}$$

Subtracting this equation from Eq. 4.19 we have

$$q'_t + \frac{\bar{\psi}_r q'_\theta}{r} - \frac{\psi'_\theta \bar{q}_r}{r} = \nu\nabla^4\psi' + 4\Omega^2\kappa\frac{\partial}{\partial z}\left(\frac{\nabla^2\psi'_z}{N^2}\right). \tag{4.21}$$

This is the linearized quasigeostrophic potential vorticity equation. It describes the stability of the background state to small amplitude perturbations. If these are growing perturbations then they will become finite in size, and not only will the assumptions used

in the linearization become invalid, but also the background state will no longer be a good representation of the system. It is then, rather fortunate that there is a definite commonality between the structure of the growing solutions which the linear problem provides us with, and the dominant states observed in atmospheric and oceanic flows (Eady, 1949), as well as in the annulus experiments (Fowlis and Hide, 1965). The nonlinear, finite amplitude problem is not discussed here; this topic is covered, for example, in Pedlosky (1987), his Section 7.16.

We can solve Eq. 4.21 by expressing ψ' and q' in terms of Fourier modes in azimuth and time. This can be written as

$$\psi' = \Re\left(\psi(r,z)e^{i(m\theta-\omega t)}\right), \tag{4.22}$$

and

$$q' = \Re\left(q(r,z)e^{i(m\theta-\omega t)}\right) \tag{4.23}$$

where ψ and q have now been used to represent the non-periodic variation of the stream function and quasigeostrophic potential vorticity, and $\Re()$ means the real part of the amplitude function. The growth rate of any mode is $\Im(\omega)$ where ω is a complex number, and $\Im()$ means the imaginary part. We can categorise the possible solutions into three groups,

(a) $\Im(\omega) < 0$, in which case the amplitude of the mode decays exponentially in time. Such modes will not be observed in a physical system.

(b) $\Im(\omega) = 0$, in which case the mode is purely oscillatory, and its amplitude neither grows nor decays. Such modes are known as neutral modes.

(c) $\Im(\omega) > 0$, in which case the amplitude of the mode grows exponentially in time. Such modes will typically be the dominant observed instability in a physical system.

In order to proceed in this way, it is necessary to assume that the azimuthal and temporal parts of the solution are separable from the meridional variation. This enforced separability limits the range of solutions which we might find, but can be justified for the following reasons:

- *Computational expediency.* The normal mode approach is an effective way to solve the linear problem. It may mean that we are not solving Eq. 4.21 in the *most* general way, but at least we are able to find a solution.

- *Agreement with observations.* This method generates dominant modes which often agree well with those observed in the systems which we are trying to model.

- *Maintained generality.* There is no loss in generality of the zonal structure since any number of Fourier modes can be combined to represent an arbitrary state.

- *Numerical convenience.* By using a Fourier representation of the temporal variation, we are able to calculate the growth rate of the normal modes directly.

If we substitute the Fourier forms given above for ψ' and q' into the linear *qpve*, it can then be written

$$-i\omega q + im\frac{\bar{\psi}_r}{r}q - im\frac{\psi}{r}\bar{q}_r = \nu\nabla^4\psi + 4\Omega^2\kappa\frac{\partial}{\partial z}\left(\frac{\nabla^2\psi_z}{N^2}\right), \tag{4.24}$$

or more simply,

$$-\omega\frac{q}{m} + \frac{\bar{\psi}_r}{r}q - \frac{\psi}{r}\bar{q}_r = -i\frac{\nu}{m}\nabla^4\psi - 4i\Omega^2\frac{\kappa}{m}\frac{\partial}{\partial z}\left(\frac{\nabla^2\psi_z}{N^2}\right), \tag{4.25}$$

where the Laplacian can now be expanded as

$$\nabla^2\phi = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\phi}{\partial r}\right) - \frac{m^2\phi}{r^2} + \frac{\partial^2\phi}{\partial z^2}. \tag{4.26}$$

4.2.1 The basic mechanism for instability.

Before going on to discuss Eady's model, it is useful to consider a rather more physical description of baroclinic instability (or sloping convection). If we consider part of a rotating fluid which is subject to thermal forcing, it could be represented by the system shown in Figure 4.1. In this figure, gravity acts down the z-axis (which is also the rotation axis) and the horizontal line shows a level of constant gravitational potential. The continuous sloping lines are isothermal surfaces. Since the fluid is virtually incompressible over the small temperature range in the annulus we can use isothermal surfaces instead of

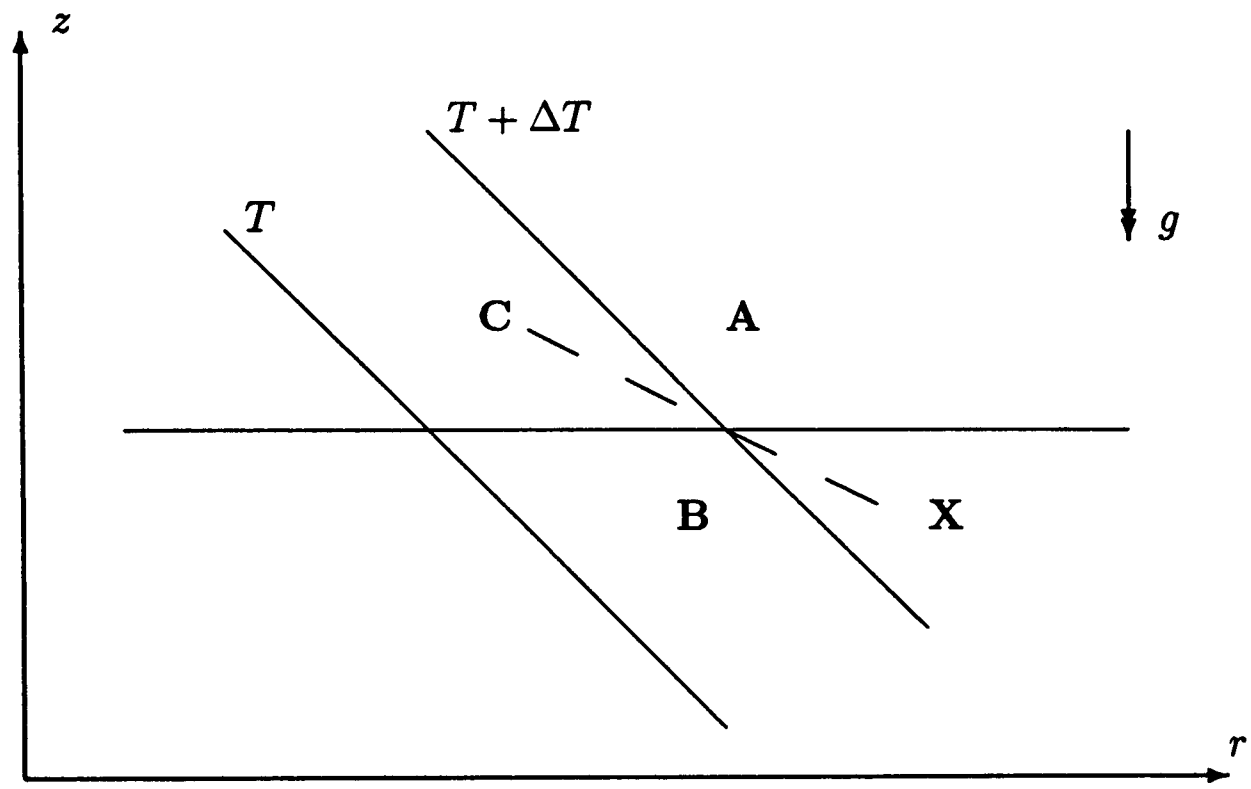


Figure 4.1: A schematic representation of baroclinic instability or sloping convection. There is a release of potential energy associated with the exchange of fluid parcels at points **X** and **C**, but not with those at **X** and **A** or **X** and **B**.

isentropic surfaces without any loss of accuracy. The isothermal surfaces are sloping (they are shown with an exaggerated gradient for the sake of clarity) because of the Coriolis force balancing out the pressure gradients in the fluid. In the absence of rotation, the isotherms would quickly level out to release the potential energy stored in the system.

We can associate notional fluid parcels with the points labelled **X**, **A**, **B** and **C**, and consider the energetics of exchanging various pairs of parcels. Although parcel **A** is higher than parcel **X**, they are both at the same temperature and so have the same density. Exchanging these parcels would not alter the energetic balance of the system. A similar argument applies to parcels **B** and **X**, except that these are at the same geopotential level, but their temperatures and therefore their densities are different. There will therefore, be no energy release in exchanging these fluid parcels. However, if we now consider the exchange of parcels **X** and **C** we find that potential energy is released. The fluid at **C** is cooler and therefore more dense than that at **X**. It is also at a higher geopotential level and so it is energetically favourable for this exchange to occur. Once these fluid parcels begin to move towards each other, the resulting forces will be such as to accelerate them in that direction and away from their original positions. The same will be true of any pair of

fluid parcels which lie within the so called 'wedge of instability' between the geopotentials and isotherms. This is the basic mechanism behind the baroclinic instability.

In the above discussion, it is clear that the stronger the rotation of the system, the more the isotherms will slope and the wider the wedge of instability will become. For very low rotation rates, any motion associated with the baroclinic instability will necessarily be almost horizontal. It should also be noted that the horizontal gradient of the isotherms implies a vertical shear in the zonal flow, as described by the thermal wind relations earlier in this chapter. This is the dominant flow observed both in the body of the fluid in the rotating annulus experiments, and in the atmospheric circulation[†], and this suggests that we might expect the baroclinic instability to play a major rôle which, indeed, it does.

4.2.2 The Eady problem.

In his elegant and simple model, Eady (1949) captured the essence of the baroclinic instability and was able to obtain an analytic solution. The main simplifications which Eady used have already been given in Section 1.2.4 of this thesis. The key point is that by choosing a suitable zonal flow, the horizontal gradient of the quasigeostrophic potential vorticity can be made to vanish. Eady's model used a local rectangular coordinate system and so, for a zonal flow which is proportional to height only,

$$-\frac{\partial \bar{\psi}_\in}{\partial y} = \bar{u}_\in = a_\in z \tag{4.27}$$

where y is the poleward direction, az is the zonal flow and the subscript $\in$ refers to the Eady model. From this equation, it follows that

$$\bar{\psi}_\in \sim -a_\in yz + f(z), \tag{4.28}$$

and that, using x as the zonal direction,

$$\bar{q}_\in = \frac{\partial^2 \bar{\psi}_\in}{\partial y^2} + 4\frac{\Omega^2}{N^2}\frac{\partial^2 \bar{\psi}_\in}{\partial z^2} = \frac{\partial^2 f}{\partial z^2}, \tag{4.29}$$

[†]For the oceans, although the only place that we can sensibly consider a global zonal flow is in the Antarctic Circumpolar Current (ACC), this too has a vertically sheared structure.

so the horizontal (poleward) gradient of $\bar{q}_\epsilon$ is indeed zero. This enabled Eady to simplify the linearized *qpve* by setting this term equal to zero, to get

$$\frac{\partial q'}{\partial t} - \frac{\partial \bar{\psi}}{\partial y}\frac{\partial q'}{\partial x} = 0, \tag{4.30}$$

or using the normal mode form

$$(\omega + m\frac{\partial \bar{\psi}}{\partial y})q' = 0. \tag{4.31}$$

The dissipative terms which were on the left hand side of the linearized *qpve* are assumed to be small for the Eady problem and so are set to zero.

The non-trivial solutions to this equation are found by solving

$$q' = -m^2\psi' + \frac{\partial^2\psi'}{\partial y^2} + 4\frac{\Omega^2}{N^2}\frac{\partial^2\psi'}{\partial z^2} = 0 \tag{4.32}$$

subject to the boundary conditions $\psi' = 0$ at the side walls (this follows from the no-slip boundary condition); and

$$(m\bar{u}_\epsilon - \omega)\frac{\partial\psi'}{\partial z} - m\bar{u}_{\epsilon_z}\psi' = 0 \tag{4.33}$$

at the lower and upper horizontal boundaries.† By assuming a Fourier type solution in latitude for ψ' (this is consistent with the boundary conditions), Eq. 4.32 can be written as an ordinary differential equation,

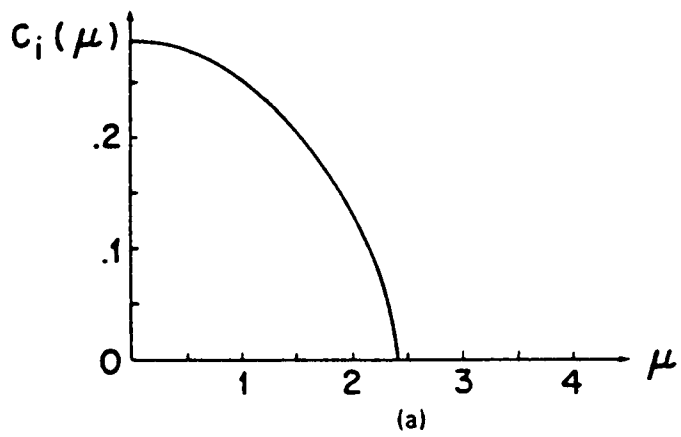
$$\frac{\mathrm{d}^2F}{\mathrm{d}z^2} - \mu^2F = 0, \tag{4.34}$$

where F is the vertical dependence of ψ', l is the latitudinal wave number and $\mu^2 = \frac{1}{4}\frac{N^2}{\Omega^2}(m^2 + l^2)$. The general solution of this equation is

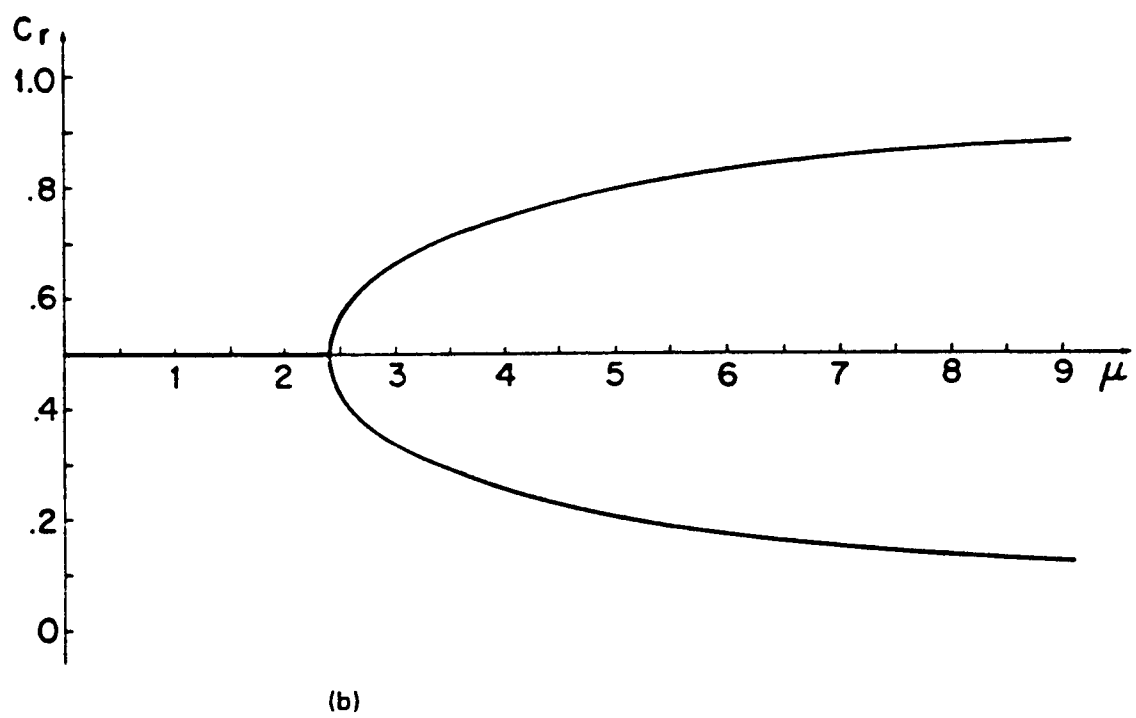
$$F(z) = a\cosh\mu z + b\sinh\mu z; \tag{4.35}$$

by applying the vertical boundary condition at the bottom (where $\bar{u}_\epsilon = 0$) and at the top, and specifying a non-trivial solution, we can find an expression for the growth rate, ω, in

†The horizontal boundary conditions are obtained by using the thermodynamic equation to express the perturbation vertical velocity in terms of ψ'. In the classical Eady problem it is then customary to set this perturbation velocity equal to zero as we have done here. Later in this chapter, we will show how Ekman pumping velocities can be used to allow for the effect of viscus boundary layers.



(a)



(b)

Figure 4.2: The real and imaginary parts of the phase speed, ω/m as a function of μ for the classical Eady problem (after Pedlosky, 1987).

terms of μ. This relationship is discussed later in the context of cylindrical coordinates, however the same crucial point holds: that there is a critical wave number ($\mu_c = 2.3994$) above which the growth rate is purely real and the solutions are pairs of neutral modes, and below which the growth rate is complex and the solution forms pairs of decaying and growing modes. The real and imaginary parts of the phase speed (ω/m) are illustrated in Figure 4.2.

A useful non-dimensional number which plays a very important part in determining the behaviour of this system is the Burger number, B, which is defined as

$$B = \frac{1}{4}\frac{N^2}{\Omega^2}\frac{D^2}{L^2}. \tag{4.36}$$

D and L are typical vertical and horizontal length scales. The Burger number plays two roles in the Eady problem: firstly there is a critical Burger number for any flow which must

be exceeded for regular waves to be present (see the discussion of the regime diagram for the annulus experiments given in Chapter 1), and secondly it is used to scale the combined zonal/lateral wave number of the flow to give μ. Before expecting the presence of growing modes, the Burger number must be of an appropriate value to give regular waves, and this value must then be used in the calculation of the expected zonal wave number of the instability. For example, a typical value for B for the annulus is ~ 0.1; if we then consider the most unstable lateral mode, $l = 1$, (see Pedlosky, 1987) then the maximum zonal wave number is

$$m = \sqrt{\frac{\mu_c^2}{B} - l^2} \sim 7. \tag{4.37}$$

If the rotation rate is now halved, the value of B is quadrupled, and the maximum zonal wave number for a growing disturbance would be 3, assuming that this flow is still in an unstable part of parameter space.

We return to the Eady problem later in Chapter 5, but find a solution for an annular region in cylindrical coordinates. The problem is mathematically very similar to the standard Eady problem, except for the use of Bessel functions in the radial direction. It does, however, offer a system rather closer to the laboratory apparatus, especially when it is extended to include viscous end wall boundary layers.

4.2.3 Formulating the matrix problem.

The linearized *qpve* with normal mode forms assumed for the perturbation terms, Eq. 4.25, relates the perturbation stream function, ψ, to the perturbation quasigeostrophic potential vorticity, q. We already know the relationship between these two functions, and so given suitable boundary conditions, we can express either one in terms of the other. We seek to write Eq. 4.25 in terms of one function only, and choose the perturbation stream functions because it transpires that the boundary conditions are easier to apply to it, than to q.

Using the normal mode structure in the zonal direction and using cylindrical polar

coordinates, we have that

$$q = \nabla_h^2 \psi + 4\Omega^2 \frac{\partial}{\partial z}\left(\frac{\psi_z}{N^2}\right) = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial \psi}{\partial r}\right) - m^2\frac{\psi}{r^2} + 4\Omega^2 \frac{\partial}{\partial z}\left(\frac{\psi_z}{N^2}\right). \tag{4.38}$$

This is equivalent to a Laplacian with a stretched vertical coordinate acting on ψ. If the Brunt-Väisälä frequency is independent of height, then the stretched coordinate is proportional to the normal height and the relationship is particularly straightforward. If we write

$$q = \nabla\!\!\!\!\nabla^2 \psi \tag{4.39}$$

where $\nabla\!\!\!\!\nabla$ is the Laplacian with a stretched vertical coordinate, it is then a natural progression to write

$$\psi = \nabla\!\!\!\!\nabla^{-2} q. \tag{4.40}$$

The boundary conditions on ψ must be applied when calculating the quasigeostrophic potential vorticity, and are an integral part of defining the operator $\nabla\!\!\!\!\nabla$. At the sidewall boundaries there is no radial flow and so we know that

$$u' = -\frac{\partial \psi'}{\partial \theta} = -im\psi = 0 \tag{4.41}$$

and so either $m = 0$ and the flow is purely axisymmetric (a trivial solution), or ψ is zero. The latter gives the appropriate boundary condition for the vertical boundaries, but on the horizontal boundaries, the problem is more complex. By analogy with the radial component of the flow we should (at least in the first instance) seek to apply a zero vertical velocity constraint. We can use the thermodynamic equation, Eq. 4.7, to write the vertical velocity in terms of the perturbation stream function:

$$w = 2i\frac{\Omega}{N^2}\left[\left(\omega - m\frac{\bar{\psi}_r}{r}\right)\psi_z + m\frac{\bar{\psi}_{rz}}{r}\psi\right]. \tag{4.42}$$

If we say that this velocity is zero then this boundary condition simplifies to a similar form to that used for the Eady problem (it differs only in the change of coordinate system). However, at the top and bottom boundaries Ekman layers are formed (see, for example, Gill, 1982) where the main body of the (quasi)geostrophic flow matches to the viscosity driven flow at the boundary. Associated with the formation of these Ekman layers is a vertical Ekman pumping flow; this obviously gives a more realistic boundary flow to

use in determining the boundary conditions for ∇ than no vertical flow at all. Antar and Fowlis (1980) used Ekman layers in an analytical study of the baroclinic instability of flow in a β-plane channel, and found that the critical Burger number for an unstable mode increased as the Ekman number (see footnote) decreased.

From Ekman theory we have that the pumping velocity is

$$w_{EP} = \mp \frac{E_v^{\frac{1}{2}}}{2} \nabla_h^2 \psi, \tag{4.43}$$

where E_v is the Ekman number†, and the + and − signs refer to the lower and upper boundaries respectively. Matching the Ekman pumping velocity to the form for the vertical velocity given in Eq. 4.42 we have that

$$2i\frac{\Omega}{N^2}\left[\left(\omega - m\frac{\bar{\psi}_r}{r}\right)\psi_z + m\frac{\bar{\psi}_{rz}}{r}\psi\right] = \mp\frac{E_v^{\frac{1}{2}}}{2}\nabla_h^2\psi, \tag{4.44}$$

or

$$\left(\frac{\bar{\psi}_r}{r}\frac{\partial}{\partial z} - \frac{\bar{\psi}_{rz}}{r} \pm \frac{i}{m}\Xi\nabla_h^2\right)\psi = \lambda\frac{\partial\psi}{\partial z} \tag{4.45}$$

where $\Xi = \frac{N^2}{2\Omega}\frac{E_v^{\frac{1}{2}}}{2}$, and $\lambda = \omega/m$. This is the form of the boundary condition which is applied at the lid and base of the annulus. It is useful that we can recover the simpler boundary condition of no vertical flow by setting $\Xi = 0$, which is equivalent to ignoring the viscous drag at the boundary.

The expression given by Eq. 4.45 is itself an eigenvalue problem and so cannot be simply applied in order to re-write the main operator matrix, $\hat{\mathcal{Z}}$ (see below), as this makes the problem nonlinear in λ. Rather the boundary condition itself replaces $\hat{\mathcal{Z}}$ for the rows of the matrix which act on the end wall boundary points. We discuss this problem in again in Chapter 5.

With ∇ now properly defined, we can write Eq. 4.25 in terms of the perturbation stream function only:

$$-\frac{\omega}{m}\nabla^2\psi + \frac{\bar{\psi}_r}{r}\nabla^2\psi - \frac{\bar{q}_r}{r}\psi = -i\frac{\nu}{m}\nabla^4\psi - 4i\Omega^2\frac{\kappa}{m}\frac{\partial}{\partial z}\left(\frac{\nabla^2\psi_z}{N^2}\right). \tag{4.46}$$

†The Ekman number is the relative magnitude of viscous diffusion and the Coriolis acceleration, $E_v = \nu/2\Omega H^2$, where H is the vertical length scale.

For a given zonal wave number, m, we seek to solve this equation by finding what forms ψ will take and what the associated growth rate will be. We can represent this by the eigenvalue problem,

$$\hat{\mathcal{Z}}\psi = \frac{\omega}{m}\psi, \tag{4.47}$$

where $\hat{\mathcal{Z}}$ is some operator acting on the stream function, which is described below. The complex eigenvalues of this problem are the complex phase speed of the disturbances and the related eigenvectors are the normal modes of the operator and represent the spatial distribution of the stream function for each eigenvalue. If the eigenvalues are complex then the imaginary part gives the growth rate of the normal mode.

At this stage it is convenient to assume that thermal diffusion is only a small term in the equation, so that we may drop the final term on the right hand side of Eq. 4.46. This is equivalent to saying that there is no diabatic heating.

We define three linear operators as:

$$\begin{aligned}
\hat{\mathcal{E}} &= \nabla^2 \quad \ldots \quad \hat{\mathcal{E}}\psi = q, \\
\hat{\mathcal{F}} &= \nabla^{-2} \quad \ldots \quad \hat{\mathcal{F}}q = \psi \text{ and} \\
\hat{\mathcal{G}} &= \nabla^2 \quad \ldots \quad \hat{\mathcal{G}}\hat{\mathcal{G}}\psi = \nabla^4\psi.
\end{aligned}$$

Eq. 4.46 then becomes

$$-\frac{\omega}{m}\hat{\mathcal{E}}\psi + \frac{\bar{\psi}_r}{r}\hat{\mathcal{E}}\psi - \frac{\bar{q}_r}{r}\psi = -i\frac{\nu}{m}\hat{\mathcal{G}}\hat{\mathcal{G}}\psi, \tag{4.48}$$

the thermal dissipation term now having been dropped. If we multiply this equation through by $\hat{\mathcal{F}}$ and recall that $\hat{\mathcal{E}}\hat{\mathcal{F}} = \hat{\mathcal{I}}$, the identity operator, we then have

$$-\frac{\omega}{m}\psi + \hat{\mathcal{F}}\frac{\bar{\psi}_r}{r}\hat{\mathcal{E}}\psi - \hat{\mathcal{F}}\frac{\bar{q}_r}{r}\psi = -i\frac{\nu}{m}\hat{\mathcal{F}}\hat{\mathcal{G}}\hat{\mathcal{G}}\psi, \tag{4.49}$$

or more simply

$$\hat{\mathcal{Z}}\psi = \frac{\omega}{m}\psi, \tag{4.50}$$

as we required, where $\hat{\mathcal{Z}}$, the total operator, is given by

$$\hat{\mathcal{Z}} = \hat{\mathcal{F}}\left(\frac{\bar{\psi}_r}{r}\hat{\mathcal{E}} - \frac{\bar{q}_r}{r} + i\frac{\nu}{m}\hat{\mathcal{G}}\hat{\mathcal{G}}\right). \tag{4.51}$$

The operators $\hat{\mathcal{E}}$ and $\hat{\mathcal{G}}$ are subject to the boundary conditions we have discussed above, *i.e.* $\psi = 0$ at the two side walls and

$$i\hat{\mathcal{Q}}\psi_z + i\hat{\mathcal{R}}\psi = \mp\Xi\hat{\mathcal{H}}\psi \tag{4.52}$$

at the upper and lower boundaries; the operators $\hat{\mathcal{H}}$, $\hat{\mathcal{Q}}$ and $\hat{\mathcal{R}}$ used here are given by

$$\hat{\mathcal{H}}\psi = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\psi}{\partial r}\right) - \frac{m^2}{r^2}\psi, \tag{4.53}$$

$$\hat{\mathcal{Q}}\psi_z = 2\frac{\Omega}{N^2}\left(\omega - m\frac{\bar{\psi}_r}{r}\right)\psi_z \text{ and} \tag{4.54}$$

$$\hat{\mathcal{R}}\psi = 2\frac{\Omega m}{N^2}\frac{\bar{\psi}_{rz}}{r}\psi. \tag{4.55}$$

Problem specification.

The operator in Eq. 4.50 defines the problem we wish to solve, and its structure is dependent upon the background flow we are analysing, as well as the rotation rate, the kinematic viscosity, and the physical dimensions of the annulus. The simplest way to specify the problem, is to define the background state on which we wish to perform the stability analysis in terms of the meridional distributions of the zonal mean stream function, and the Brunt-Väisälä frequency. From this information, we can construct an operator which can be used to act on the zonal mean stream function to give the corresponding quasi-geostrophic potential vorticity distribution. This operator is very similar to ∇^2, except in the top and bottom boundary conditions, which require no vertical flow.

It is also possible to define the problem by expressing the zonal flow in terms of its zonal velocity and temperature fields, such as might be generated by some numerical model (*e.g.* SARA , the annulus model described in Chapters 2 and 3 of this thesis), or measured in a laboratory experiment. The zonal velocity is equal to the radial gradient of the background stream function, and since it is this, and not the stream function itself, which is required in constructing the final operator matrix, $\hat{\mathcal{Z}}$, then no numerical integration needs to be carried out. The meridional distribution of the Brunt-Väisälä

frequency can be calculated from the vertical gradient of the temperature field. This then needs to be averaged for each horizontal level, so that N^2 is a function of height only. $N^2(z)$ can then be used as described above to calculate the quasigeostrophic potential vorticity distribution.

4.2.4 Invertibility.

The instability problem is posed here in terms of the perturbation stream function, ψ. We chose to formulate the equations in this way because it made the boundary conditions much easier to apply than they would be if the problem were specified in terms of q, the perturbation quasigeostrophic potential vorticity. Once the operators $\hat{\mathcal{E}}$ and $\hat{\mathcal{F}}$ have been defined, there is no reason why we could not rewrite the problem in terms of q using the mapping $\psi = \hat{\mathcal{F}}q$. This is possible because we have formed a one-to-one relationship between ψ and q.

It is the existence of exactly this kind of mapping which leads to the celebrated *invertibility principle*, which is described at some length by Hoskins, McIntyre and Robertson (1985). Essentially, invertibility is concerned with using q, a scalar variable, to determine the velocity, temperature and pressure distributions. It is only possible to extract this much information from one unknown if: (i) some balance condition, such as the quasigeostrophic approximation, is assumed to hold and (ii) we know all the boundary conditions. If this is the case, then q can always be inverted to give the stream function. For the instability problem as it is posed here, there is never any need to invert the final solution since it is already in terms of the stream function. We can calculate the perturbation velocities, temperature and pressure associated with a given stream function distribution by using the following equations.

$$u = im\frac{\psi}{r}, \tag{4.56}$$

$$v = -\frac{\partial \psi}{\partial r}, \tag{4.57}$$

$$P = \bar{\rho}\psi, \tag{4.58}$$

$$T = \bar{T} - 2\frac{\Omega}{g\alpha}\frac{\partial \psi}{\partial z} \text{ and} \tag{4.59}$$

$$w = 2\Omega\frac{g\alpha}{N^2}\left(-i\omega T + im\frac{\psi_r T}{r} - im\frac{\psi T_r}{r}\right). \tag{4.60}$$

These equations follow from the description of the quasigeostrophic approximation given earlier in this chapter and the normal mode form of solution.

4.3 INSTAB: A linear instability code.

We seek to solve the eigenvalue problem posed by Eq. 4.50 for any specified zonal mean state and for a specific zonal wave number. For each wave number the system can be decomposed into as many normal modes as there are degrees of freedom. We then wish to know if any of these modes are linearly unstable, and if so, what the growth rate and structure of the most unstable (fastest growing) mode are.

Except for a few special cases such as the Eady problem, or the sinusoidal internal jets studied for example by Bell (1989), it is not possible to solve Eq. 4.50 analytically. With this in mind we have developed a computer code, INSTAB , to solve for the normal modes using numerical methods. The code represents the various fields on a regularly spaced grid of points in the meridional plane. The operators are then compiled as matrices using standard second order finite differencing methods, with the two dimensional grid of points in the meridional plane being replaced by a vector of the same points, listed sequentially. This same idea is used in representing the pressure in SARA and is discussed briefly in Appendix A. Each of the operators $\hat{\mathcal{E}}$, $\hat{\mathcal{F}}$, $\hat{\mathcal{G}}$, $\hat{\mathcal{H}}$, $\hat{\mathcal{Q}}$, $\hat{\mathcal{R}}$ and $\hat{\mathcal{Z}}$ are represented by matrices, the complex operators being represented by two separate matrices, one for the real part and one for the imaginary part.

In Figure 4.3 the left hand diagram shows a schematic representation of the matrix corresponding to the forward operator, $\hat{\mathcal{E}}$. The right hand diagram shows a similar representation of the backward operator $\hat{\mathcal{F}}$ which is the inverse of $\hat{\mathcal{E}}$. In both cases the matrices are shown flipped vertically, so that the leading diagonal runs from the lower left to the upper right corner. These matrices are from a run emulating the classical Eady model,

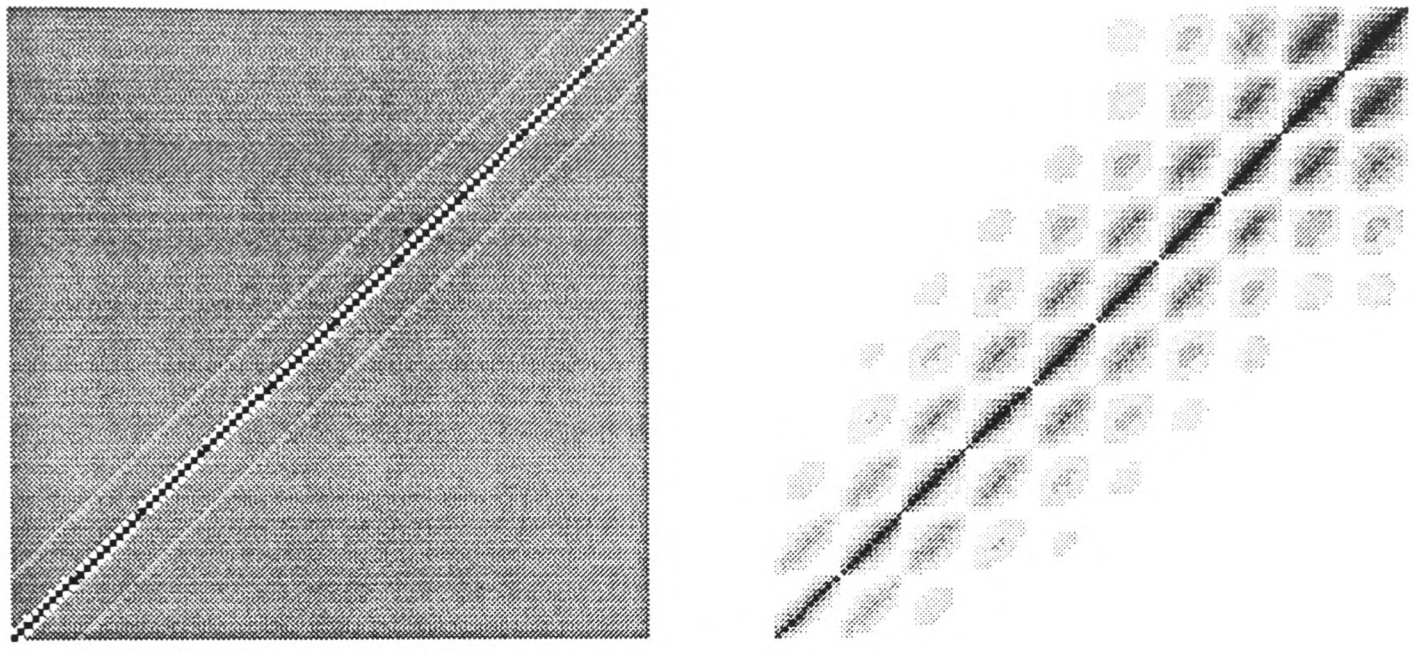


Figure 4.3: An illustration of the typical structure of $\hat{\mathcal{E}}$, the forward operator (left hand side), and $\hat{\mathcal{F}}$, the backward operator (right hand side) in their matrix forms. The matrices are shown flipped vertically so that the leading diagonal runs from the lower left to the upper right corner. The grey levels for each diagram are scaled to range from black to white across the complete range of values in the matrix. Only the real parts are shown, since in this case both $\hat{\mathcal{E}}$ and $\hat{\mathcal{F}}$ had no imaginary part.

and the top and bottom boundary conditions imposed no vertical flow. For each matrix, the range of values has been mapped onto a grey scale using the linear relation

$$v_g = \text{INT}\left[\left(\frac{(v - \min v)}{(\max v - \min v)}\right) \times 255\right], \tag{4.61}$$

where v is the value of the matrix element and v_g is its associated grey scale value ($0 \leq v_g \leq 255$). This results in relatively low values being dark and relatively high values being light. We can see that $\hat{\mathcal{E}}$ has the characteristic quindiagonal structure of the finite difference form of the Laplacian in two dimensions. Most of the matrix is zero (the mid grey areas), illustrating the effectively local nature of $\hat{\mathcal{E}}$ in INSTAB. The ratio of the magnitudes of the two non zero elements on one side of the diagonal on any row is equal to the Burger number, B, discussed above.

For $\hat{\mathcal{F}}$, the zero values are white, and it is clearly much less sparse than $\hat{\mathcal{E}}$. If we consider one particular row (which refers to a particular point in the meridional plane), then the darkness of each element shows how much influence the associated grid point has when applying the operator. The darkest values are on the diagonal so the effect at any grid point is most strongly determined by that point. The neighbouring points, however,

can also be seen to play a significant role, which decreases gradually as they get further away. Therefore $\hat{\mathcal{F}}$, which is the linear inversion operator, acting on q to give ψ, acts in a much more global sense, and we can see pictorially the truth of the assertion that "... a purely local knowledge of P (here q) cannot determine the local absolute vorticity (or ψ) ..." (Hoskins, McIntyre and Robertson, 1985, their section 1d).

The use of regular spacing in radius and height for the grid points in INSTAB is by necessity rather than choice. Ideally, one would construct the grid in a similar way to the one used in SARA (see Chapter 3), which had a finer resolution toward the boundaries where the typical length scale of the motion becomes smaller. The problem with such a scheme is that the eigenvalue solver requires the data to be regularly spaced, and so we would have to interpolate onto a regular grid. A more practical way to increase the resolution is to increase the number of regular grid points. The limiting factors here are:

1. The storage size, which, for a square grid, is proportional to the number of points in one direction to the fourth power, and
2. The run time, which is proportional to the number of points in one direction to the sixth power[†].

This method is not the only way in which growing states can be calculated for this type of system, and in some circumstances a different approach may be more appropriate. Miller and Gall (1983 b) use an axisymmetric, Boussinesq, finite-difference model to integrate the Navier-Stokes equations to a steady state for a baroclinic annulus system. They then use a hydrostatic version of the same model, linearized about the axisymmetric solution, and time march forward after some arbitrary disturbance until the structure and growth rate of the most unstable mode can be found. By developing INSTAB, we have been able to perform the same type of experiments, without having to build a suitable linearized, hydrostatic, finite difference code for this purpose. It would, however, be an interesting exercise to make a linearized, hydrostatic version of SARA (see Chapter 3) and see to what extent the results of Miller and Gall (1983 b) can be reproduced and extended.

[†]According to the NAG documentation. This means changing from a 10×10 to a 20×20 grid would increase the run time by a factor of 64.

4.3.1 Matrix inversion routine.

Central to the eigenvalue problem is the inversion of the forward operator, $\hat{\mathcal{E}}$. The matrix representing $\hat{\mathcal{E}}$ will in generally be complex[‡], and non-symmetric[§]. Normally one would use standard library software for such an operation, but for INSTAB this proved infeasible for two reasons.

Firstly, the appropriate NAG Library routine, F04ADF, could not be made to work with double precision complex non-symmetric matrices larger than 12×12. Since the lowest resolution runs of INSTAB resulted in a matrix $\hat{\mathcal{E}}$, of 100×100, this limitation was very worrying.

Secondly, F04ADF proved to be rather inaccurate when some simple tests were carried out. We specified some non-singular complex matrix $\hat{\mathcal{M}}$, and sought to calculate its inverse. If the exact inverse had been calculated, then we would have, by definition, that

$$\hat{\mathcal{M}}\hat{\mathcal{M}}^{-1} \equiv \hat{\mathcal{I}} \equiv \begin{bmatrix} 1.0+0.0i & 0.0+0.0i & 0.0+0.0i & \cdots \\ 0.0+0.0i & 1.0+0.0i & 0.0+0.0i & \cdots \\ 0.0+0.0i & 0.0+0.0i & 1.0+0.0i & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{4.62}$$

However, if we take the 'inverse' matrix generated by the NAG software (denoted by the tilde), and multiply it by the original matrix instead of the unit operator, we typically get something similar to the following:

$$\hat{\mathcal{M}}\widetilde{\hat{\mathcal{M}}^{-1}} = \begin{bmatrix} 0.8+0.2i & 0.3-0.1i & -0.2-0.3i & \cdots \\ -0.1+0.2i & 0.7-0.1i & 0.2+0.2i & \cdots \\ -0.3+0.1i & 0.2+0.2i & 0.8+0.1i & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{4.63}$$

Using the NAG Library software for the matrix inversions is clearly not satisfactory if we are to be able to obtain any useful insight from the ensuing calculations.

[‡]because of the vertical boundary condition

[§]because of the $1/r$ terms due to the cylindrical coordinate system

Fortunately, there is a simple and commonly used method for inverting complex matrices which is quick and accurate, and not so severely constrained by size. It is relies upon our being able to rewrite the problem in terms of purely real matrices. Let us define four such matrices, $\hat{a}$, $\hat{b}$, $\hat{x}$ and $\hat{y}$ as

$$\hat{\mathcal{M}} = \hat{a} + i\hat{b} \text{ and } \hat{\mathcal{M}}^{-1} = \hat{x} + i\hat{y}. \tag{4.64}$$

If we now substitute these definitions into the identity 4.62, we form the simultaneous equations for $\hat{x}$ and $\hat{y}$,

$$\left.\begin{aligned} \hat{a}\hat{x} - \hat{b}\hat{y} &= \hat{\mathcal{I}} && \text{(4.65-a)} \\ \hat{b}\hat{x} + \hat{a}\hat{y} &= 0 && \text{(4.65-b)} \end{aligned}\right\}$$

By post-multiplying Eq. 4.65-a by $\hat{a}^{-1}\hat{b}$ and subtracting Eq. 4.65-b, $\hat{x}$ and $\hat{y}$ are found to be

$$\hat{x} = \left\{\hat{a} + \hat{b}\hat{a}^{-1}\hat{b}\right\}^{-1} \tag{4.66}$$

$$\hat{y} = -\hat{a}^{-1}\hat{b}\left\{\hat{a} + \hat{b}\hat{a}^{-1}\hat{b}\right\}^{-1} \tag{4.67}$$

We are thereby able to construct the inverse of any complex matrix (which has an inverse), purely by finding the inverse of two real matrices, $\hat{a}^{-1}$ and $\hat{x}^{-1}$, and performing simple additions and multiplications. We used this method to find the inverse of $\hat{\mathcal{M}}$, which, when pre-multiplied by $\hat{\mathcal{M}}$ gave

$$\hat{\mathcal{M}}\widetilde{\hat{\mathcal{M}}^{-1}} = \begin{bmatrix} 0.98 + 0.01i & -0.03 - 0.02i & 0.01 - 0.00i & \cdots \\ -0.02 + 0.03i & 1.03 - 0.04i & 0.03 + 0.01i & \cdots \\ 0.03 + 0.02i & -0.01 + 0.05i & 1.01 + 0.02i & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \tag{4.68}$$

which was an obvious improvement. The only *caveat* concerning this method applies when either $\hat{a}^{-1}$ or $\hat{x}^{-1}$ are singular, as, for example, when the matrix to be inverted is purely imaginary. In this case, Eqs. 4.65-a and 4.65-b can be solved by post-multiplying Eq. 4.65-a by $\hat{b}^{-1}\hat{a}$ and adding Eq. 4.65-b. $\hat{x}$ and $\hat{y}$ are then found to be

$$\hat{y} = -\left\{\hat{a}\hat{b}^{-1}\hat{a} + \hat{b}\right\}^{-1} \tag{4.69}$$

$$\hat{x} = \hat{b}^{-1}\hat{a}\left\{\hat{a}\hat{b}^{-1}\hat{a} + \hat{b}\right\}^{-1} \tag{4.70}$$

The inversion of a real matrix is a straightforward matter. Here we use LU decomposition, and forward/backward substitution, in an algorithm adapted from that of Press *et al.* (1989).

4.3.2 Solution of the eigensystem.

The general eigenvalue problem posed by Eq. 4.50 can be rewritten as

$$\det \left| \hat{\mathcal{Z}} - \frac{\omega}{m} \hat{\mathcal{I}} \right| = 0. \tag{4.71}$$

Solution of the resulting polynomial will give each of the eigenvalues of the system. The problem with this type of approach is that it is very numerically intensive, and prone to inaccuracy.

For the purposes of INSTAB we take the advice offered by Press *et al.* (1989) and use the NAG Library software designed for this problem. Unfortunately, none of the more efficient algorithms can be used because they rely on various attributes of the matrix such as symmetry or being purely real. The subroutine provided for complex non-symmetric matrices where all the eigenvalues and eigenvectors are required, F02AKF, uses the QR algorithm, the details of which are not relevant here, but can be found in any standard text on matrix methods, for example Horn and Johnson (1985).

A simplified formulation.

Our formulation of the matrix problem presented here can be simplified somewhat by making use of the generalised NAG routine F02GJF. This routine allows the eigenvalue problem to posed as

$$\hat{\mathcal{Z}}\psi = \lambda \hat{\mathcal{Y}}\psi, \tag{4.72}$$

where $\hat{\mathcal{Z}}$ and $\hat{\mathcal{Y}}$ can be any square complex non-symmetric matrix. The details of the algorithm (which are summarised in the NAG documentation) are not important here, since it used as a "black-box" routine. Its use removes the need to calculate complex matrix

inverses, and generally simplifies the problem. The formulation of the linear instability problem in these terms is discussed in the first section of Chapter 5.

4.4 Thermal dissipation.

In the preceding description of how the eigenvalue problem is derived, we have assumed that the thermal dissipation terms are small and can be ignored. The relative importance of the dissipation can be estimated from the Nusselt number. This gives the ratio of the total heat transport to the heat transport by diffusion. For the annulus system this can be measured empirically and is typically of the order of 10 (Rayer, 1992), which indicates that thermal dissipation is indeed a small effect, and that most of the heat is transported by the baroclinic waves. Although we are then justified in our simplification, there is no reason why such a term could not be included in the problem in the future. We would simply need to redefine $\hat{\mathcal{Z}}$ as

$$\hat{\mathcal{Z}} = \hat{\mathcal{F}}\left(\frac{\bar{\psi}_r}{r}\hat{\mathcal{E}} - \frac{\bar{q}_r}{r} + i\frac{\nu}{m}\hat{\mathcal{G}}\hat{\mathcal{G}} + \left[4i\frac{\Omega^2\kappa}{m}\frac{\partial}{\partial z}\left(\frac{\hat{\mathcal{G}}}{N^2}\frac{\partial}{\partial z}\right)\right]\right), \tag{4.73}$$

where the term in square brackets is the additional effect of the dissipation. The inclusion of such a term would make very little difference to the solution†, and so is not currently included in INSTAB .

†The viscous term is very small, and since the Prandtl ratio, $Pr = (\nu/\kappa) \sim 1$, the dissipative effect will be similarly small.

Chapter 5

Results from stability analyses.

In this chapter we use the numerical stability analysis code, INSTAB , described in the previous chapter, to investigate some dynamical problems of interest and relevance to both geophysical fluid dynamics in general and the annulus system in particular. After a short discussion of the use of a suitable scheme to apply the relevant boundary conditions at the horizontal surfaces, we proceed by reproducing recognised results from the Eady problem, and thereby demonstrating the functionality of the scheme when applied to a typical system subject to baroclinic instability. The effect of changing the resolution of the code is investigated, as well as the result of including a parameterisation of frictional Ekman layers in the system. We then use INSTAB to investigate the stability of various internal baroclinic jets in the annulus, after Bell and White (1988) and Bell (1989). As a final test of the accuracy of the scheme, this chapter is concluded by describing analyses of four sets of results from the existing Met. Office model of the annulus; two of which were from simulations of a system differentially heated at the side walls, and the other two from an internally heated one.

5.1 Numerical application of the horizontal boundary condition.

In Chapter 4, Eq. 4.45, we showed how an Ekman pumping velocity could be used for the horizontal boundary condition on the perturbation flow. By setting Ξ to zero in this equation, we recover the zero vertical flow condition used in the standard Eady problem. We apply the boundary condition, which is itself an eigenvalue problem, by constructing the overall eigenvalue problem, $\hat{\mathcal{Z}}\psi = \lambda\hat{\mathcal{Y}}\psi$, ($\lambda$, the eigenvalue $= \omega/m$), as follows:

- For all the grid points not at the upper or lower boundary,

$$\hat{\mathcal{Z}} = \left(\frac{\bar{\psi}_r}{r}\nabla^2 - \frac{\bar{q}_r}{r} + i\frac{\nu}{m}\nabla^4\right), \tag{5.1}$$

  and

$$\hat{\mathcal{Y}} = \nabla^2. \tag{5.2}$$

- At the lower boundary, the boundary condition itself is used to form the operator, and so

$$\hat{\mathcal{Z}} = \left(\frac{\bar{\psi}_r}{r}\frac{\partial}{\partial z} - \frac{\bar{\psi}_{rz}}{r} - \frac{i}{m}\Xi\nabla_h^2\right), \tag{5.3}$$

  and

$$\hat{\mathcal{Y}} = \frac{\partial}{\partial z}, \tag{5.4}$$

  as before.

- At the upper boundary, we again use the boundary condition itself to form the operator, giving

$$\hat{\mathcal{Z}} = \left(\frac{\bar{\psi}_r}{r}\frac{\partial}{\partial z} - \frac{\bar{\psi}_{rz}}{r} + \frac{i}{m}\Xi\nabla_h^2\right), \tag{5.5}$$

  and again

$$\hat{\mathcal{Y}} = \frac{\partial}{\partial z}. \tag{5.6}$$

As before, we have used $\Xi = \frac{N^2}{2\Omega}\frac{E_v^{\frac{1}{2}}}{2}$, and E_v is the Ekman number (see Chapter 4).

This method ensures that the boundary conditions are always satisfied. We have tried using first and second order accurate representations of the vertical derivative in $\hat{\mathcal{Y}}$, but found there to be no noticeable difference in our results. The following results were all obtained using a first order representation.

5.2 The Eady problem in cylindrical polar coordinates.

The first system which we investigate using INSTAB is the classical Eady problem. This is a well understood system and we can use the analytical solution as a check both that INSTAB is correctly coded and that it is being correctly applied. Before we can make a direct comparison, it is necessary to modify the analytical solution slightly to take account of the cylindrical polar coordinate system used in INSTAB . We discuss how this is done in the following section.

5.2.1 Changing from local rectangular to cylindrical polar coordinates.

The adaption of Eady's original method to a cylindrical geometry is essentially a matter of adding the appropriate metric terms to the spatial operators. The only other important difference is in the form of the radial structure which is used for the perturbation stream function. In rectangular coordinates, the side wall boundary conditions ($\psi = 0$, see Chapter 4) lead us to assume a sinusoidal variation with radius. For cylindrical geometry, however, it is natural to use Bessel functions. It is also necessary to use an adapted form for the zonal mean stream function,

$$\bar{\psi}(r, z) = \Upsilon r^2 z, \tag{5.7}$$

so that the constraint $\bar{q}_r = 0$, which is part of the specification of the Eady problem, is satisfied (see Section 4.2.2 of this thesis). Υ is a scaling parameter, used to define the

velocity scale of the zonal mean flow. Drazin (1978) concluded from a lengthy analytical study of the effect of the curvature imposed by an annular geometry on the Eady problem, that increased curvature tended to destabilize the flow.

In Chapter 4 we showed (Eq. 4.32) that for non-trivial solutions to the Eady problem

$$q' = 0. \tag{5.8}$$

If we now rewrite this equation using an expression for q' in terms of ψ' in cylindrical coordinates, we have

$$q' = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial \psi'}{\partial r}\right) - \frac{m^2\psi'}{r^2} + 4\frac{\Omega^2}{N^2}\frac{\partial^2\psi'}{\partial z^2} = 0, \tag{5.9}$$

where we have used a normal mode structure of wavenumber m in azimuth. Eq. 5.9 can then be separated into an horizontal and a vertical part. For the horizontal part we can say that

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial \psi'}{\partial r}\right) - \frac{m^2\psi'}{r^2} = -\alpha^2\psi'. \tag{5.10}$$

This is a standard form of Bessel's equation, and its solutions are linear combinations of Bessel functions of the first and second kind, $J_m(\alpha r)$ and $Y_m(\alpha r)$. The particular solution is obtained by using the vertical boundary conditions, as we describe below.

Using this horizontal solution, Eq. 5.9 can be simplified to

$$-\alpha^2\psi' + 4\frac{\Omega^2}{N^2}\frac{\partial^2\psi'}{\partial z^2} = 0, \tag{5.11}$$

or defining

$$\frac{\alpha^2 N^2}{4\Omega^2} = \mu^2, \tag{5.12}$$

we have

$$\mu^2\psi' - \frac{\partial^2\psi'}{\partial z^2} = 0. \tag{5.13}$$

This is the usual form of the equation for the vertical structure in Eady's problem, differing only in our definition of μ. It is now possible to follow the remainder of Eady's method in the usual way, solving Eq. 5.13 subject to suitable horizontal boundary conditions (see Chapter 4), and obtaining a quadratic equation for ω

$$\frac{\mu^2}{m^2}\omega^2 - 2\Upsilon\frac{\mu^2}{m}z_T\omega + 4\Upsilon^2\mu z_T\coth(\mu z_T) - 4\Upsilon^2 = 0, \tag{5.14}$$

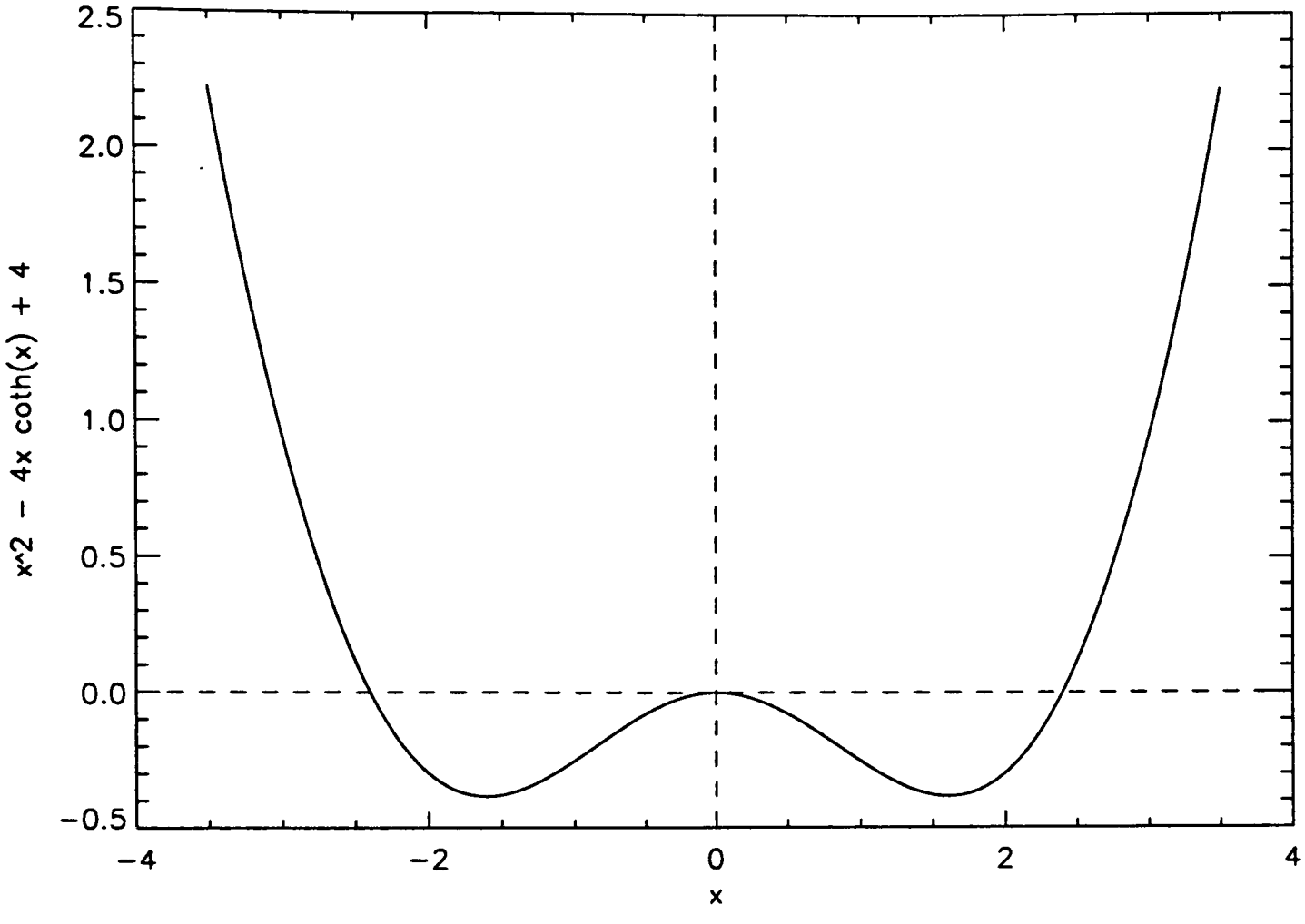


Figure 5.1: The graph of $y = x^2 - 4x\coth x + 4$. Here x is used for μz_T in the text. Where $y \leq 0$ μ has complex values and there will be a growing mode.

which is virtually identical[†] to that given in Pedlosky (1987) (his Eq. 7.7.15). In order to find growing solutions, we seek complex roots to the quadratic equation, since m is necessarily real. This gives the inequality:

$$(\mu z_T)^2 - 4\mu z_T \coth \mu z_T + 4 \leq 0 \tag{5.15}$$

a graph of which is shown in Figure 5.1. On solving this inequality numerically the critical value of μz_T is found to be 2.39935. For values smaller than this there will be complex solutions to the aforementioned quadratic equation in conjugate pairs, and dominant growing states will occur. A typical value for z_T is 0.14 m, and so this gives us a critical value for μ_c of 17.1393 m^{-1}. We can now use Eq. 5.12 to calculate a critical value for α,

$$\alpha_c^2 = \frac{\Omega^2}{N^2} 4 \times 17.1393^2 \ \mathrm{m}^{-2}, \tag{5.16}$$

[†]Where he has used μ we have μz_T, where z_T is the dimensional height of the top boundary. This difference occurs because the reference uses a nondimensionalized height.

which, given a typical value for Ω of 1 rad s^{-1} and for N^2 of 0.07 s^{-2}, gives $\alpha_c = 129.561$ m^{-1}.

The radial boundary conditions on the perturbation stream function are

$$\cdot\psi' = 0 \text{ at } r = r_a \text{ and } r = r_b. \tag{5.17}$$

We have already shown that if the zonal wavenumber is m, then the radial part of ψ can be written as

$$\psi'(r) = Y_m(\alpha r) + aJ_m(\alpha r). \tag{5.18}$$

By substituting this form for ψ into the boundary conditions, and solving the resulting equations simultaneously for a and α we have that

$$Y_m(\alpha r_b) - \frac{Y_m(\alpha r_a)}{J_m(\alpha r_a)} J_m(\alpha r_b) = 0 \text{ and} \tag{5.19}$$

$$a = -\frac{Y_m(\alpha r_a)}{J_m(\alpha r_a)}. \tag{5.20}$$

Eq. 5.19 has an infinite number of solutions, each successive solution corresponding to one more turning point in the radial structure of ψ than the last. As α becomes large the solutions become more and more regularly spaced in α. Figure 5.2 shows the radial structure for the two most unstable solutions to Eq. 5.18 for a flow with zonal wavenumber, $m = 1$. The solid curve is for $\alpha = 59.73$ and $a = 0.7483$ and the dashed curve for $\alpha = 115.82$ and $a = -0.7805$, these values having been obtained numerically. The annular gap used for this example was $r_a = 0.025$m and $r_b = 0.080$m, in accordance with the laboratory apparatus. If a regular zonal perturbation is to be present, then for Eq. 5.12 to be satisfied, modes with a less "wiggly" radial structure will be preferred, as for the standard Eady problem. For this reason we would expect the dominant modes generally to have the radial variation shown by the solid line. The two modes illustrated here both have $\alpha < \alpha_c$, and for a zonal wavenumber 1 flow are the only two which satisfy this criterion.

In Chapter 4 (Eq. 4.35) we stated that the general solution for the vertical structure in the Eady problem was of the form

$$\psi'(z) = a\cosh\mu z + b\sinh\mu z, \tag{5.21}$$

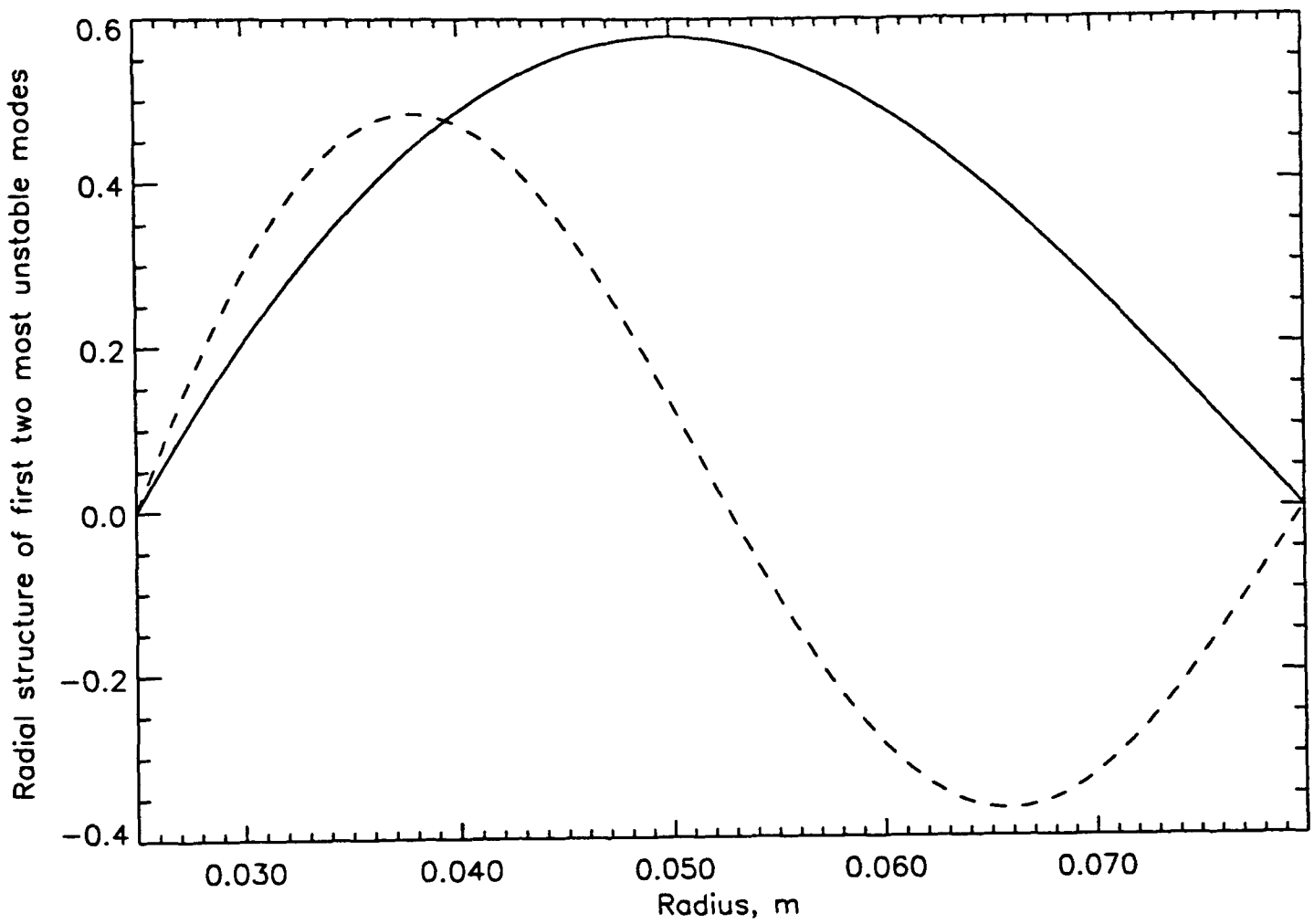


Figure 5.2: The radial structure of the two most unstable modes for the Eady problem in cylindrical coordinates. The solid line represents the most unstable mode, and the dashed line the second most unstable mode. These curves are similar to the equivalent structure obtained using rectangular coordinates, which vary sinusoidally with radius (latitude).

and so we can write the vertical derivative as

$$\psi'_z(z) = a\mu \sinh \mu z + b\mu \cosh \mu z. \tag{5.22}$$

Applying a no-slip boundary condition (Eq. 4.33), and substituting for the mean zonal velocity using Eq. 5.7, we have

$$\left(2\Upsilon z - \frac{\omega}{m}\right)\psi'_z - 2\Upsilon\psi' = 0. \tag{5.23}$$

At the lower boundary $z = 0$, and so we have that

$$c\psi'_z + 2\Upsilon\psi' = 0, \tag{5.24}$$

where c is the complex phase speed, ω/m. Substituting from Eqs. 5.21 and 5.22, and recalling that $\sinh(0) = 0$ and $\cosh(0) = 1$, we have that

$$cb\mu + 2\Upsilon a = 0, \tag{5.25}$$

or

$$b = -\frac{2\Upsilon a}{\mu c}. \tag{5.26}$$

Finally, substituting from Eq. 5.26 into Eq. 5.21, we can express the vertical variation of the perturbation stream function as

$$\psi'(z) \propto \cosh(\mu z) - \frac{2\Upsilon}{\mu c}\sinh(\mu z). \tag{5.27}$$

The proportionality results from our inability to determine the overall scale of the solution. Since ψ' is a complex function, it is helpful to consider its magnitude, given by

$$|\psi'(z)| \propto \left[\left(\cosh(\mu z) - \frac{2\Upsilon c_r \sinh(\mu z)}{\mu|c|^2}\right)^2 + \frac{4\Upsilon^2 c_i^2 \sinh^2(\mu z)}{\mu^2|c|^4}\right]^{\frac{1}{2}}. \tag{5.28}$$

This equation differs from the form common in the literature (Pedlosky (1987) his Eq. 7.7.29, for example) only in the inclusion of Υ, the velocity scale of the zonal mean flow (Eq. 5.7). Its presence in Eq. 5.28 results from our use of dimensional variables.

The overall meridional structure given by Eqs. 5.28 and 5.18 is shown in Figure 5.3 for two most unstable modes. Both modes are symmetric about mid-height, as in the standard Eady problem. The evident lateral asymmetry results from the cylindrical geometry of the problem requiring the use of Bessel functions to form the solution. The m^2/r^2 term in the Laplacian also contributes to the asymmetry at higher wave numbers.

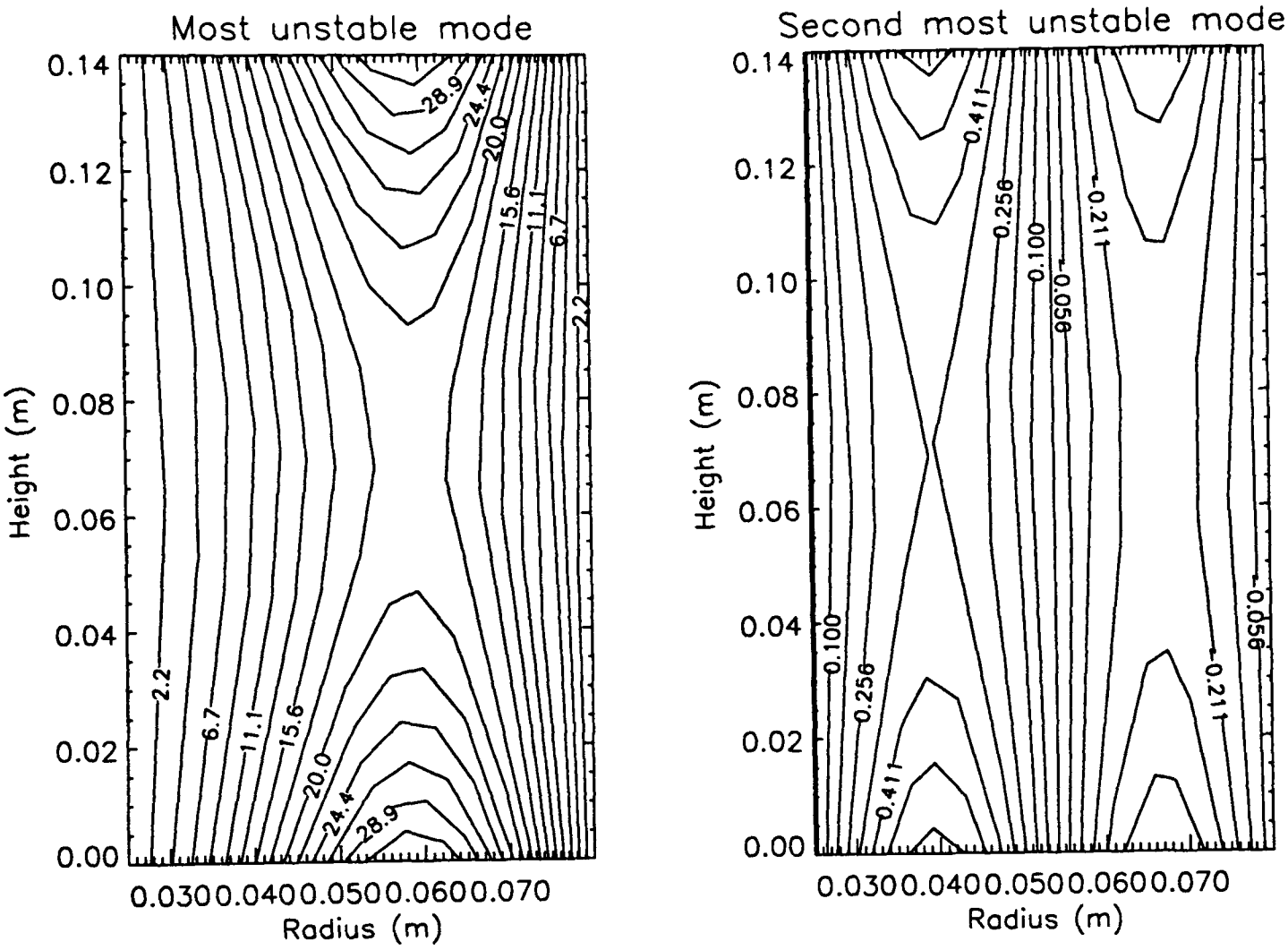


Figure 5.3: The meridional variation of the structure of the two most unstable modes for the Eady problem in cylindrical coordinates. The figure on the left corresponds to the most unstable mode and that on the right to the second most unstable mode. The values of α, and hence μ are as for Figure 5.2. The modes are for flows with zonal wavenumbers of $m = 5$ and $m = 3$ respectively, for ease of comparison with the most unstable modes calculated numerically.

Growth rate curve in cylindrical polars.

Equation 5.14 can be used to write an expression for the complex phase speed, c, of the unstable modes as a function of μ, thus

$$c = \Upsilon z_T \pm \left[\Upsilon^2 z_T^2 - 4\frac{\Upsilon^2}{\mu}\left(z_T \coth \mu z_T - \frac{1}{\mu}\right)\right]^{\frac{1}{2}}. \tag{5.29}$$

Since μ is a function of α for a given rotation rate and Brunt-Väisälä frequency (Eq. 5.12), this expression could equally well have been written in terms of α. The most useful however, is an expression for phase speed in terms of zonal wavenumber, m, but this relies upon our being able to find an expression for α in terms of m. In local Cartesian coordinates this is a trivial exercise, however the use of a cylindrical geometry means that the procedure is a little more complex. Since we are considering the most unstable mode for each wavenumber, the solution is given by the first zero of Eq. 5.19 for a particular choice of radii, r_a and r_b. Using typical values of 0.025 m and 0.080 m for these radii as before, the values in Table 5.2.1 were obtained numerically.

Using this relationship and Eq. 5.29 we can determine the value of the complex phase speed, cm, for a range of discrete zonal wavenumbers. Figure 5.4 shows the variation of the growth rate with wavenumber. From the location of the peak in the curve, we can see that the largest growth rate is close to wavenumber 4, and we would therefore expect disturbances of this scale to dominate the flow. The value of the peak growth rate is 0.0035 s^{-1}, which gives an e-folding time of $\sim$ 40 rotation periods. The curve has a clear short-wave cut-off, which has been determined numerically to be at a wavenumber of 6.65.

5.2.2 Results of stability analysis.

We now present some results obtained by using INSTAB to perform a stability analysis on the system defined by the Eady problem in cylindrical coordinates.

A particularly useful diagnostic for this type of study was the dependence of the growth rate of the fastest growing mode for a given flow on azimuthal wavenumber of that mode. That wavenumber which has the fastest growth rate will prevail in the observed flow (at

Table 5.1: Variation of α and μ with zonal wavenumber m for dominant unstable modes. Values used for parameters were, $r_a = 0.025$, $r_b = 0.080$, $\Omega = 1.0$ and $N^2 = 0.07$.

Zonal wavenumber m	α	μ
1	59.73	7.9015
2	68.95	9.1212
3	81.47	10.7775
4	95.42	12.6229
5	109.81	14.5265
6	124.25	16.4367
7	138.59	18.3338
8	152.82	20.2162
9	166.93	22.0828
10	180.94	23.9316

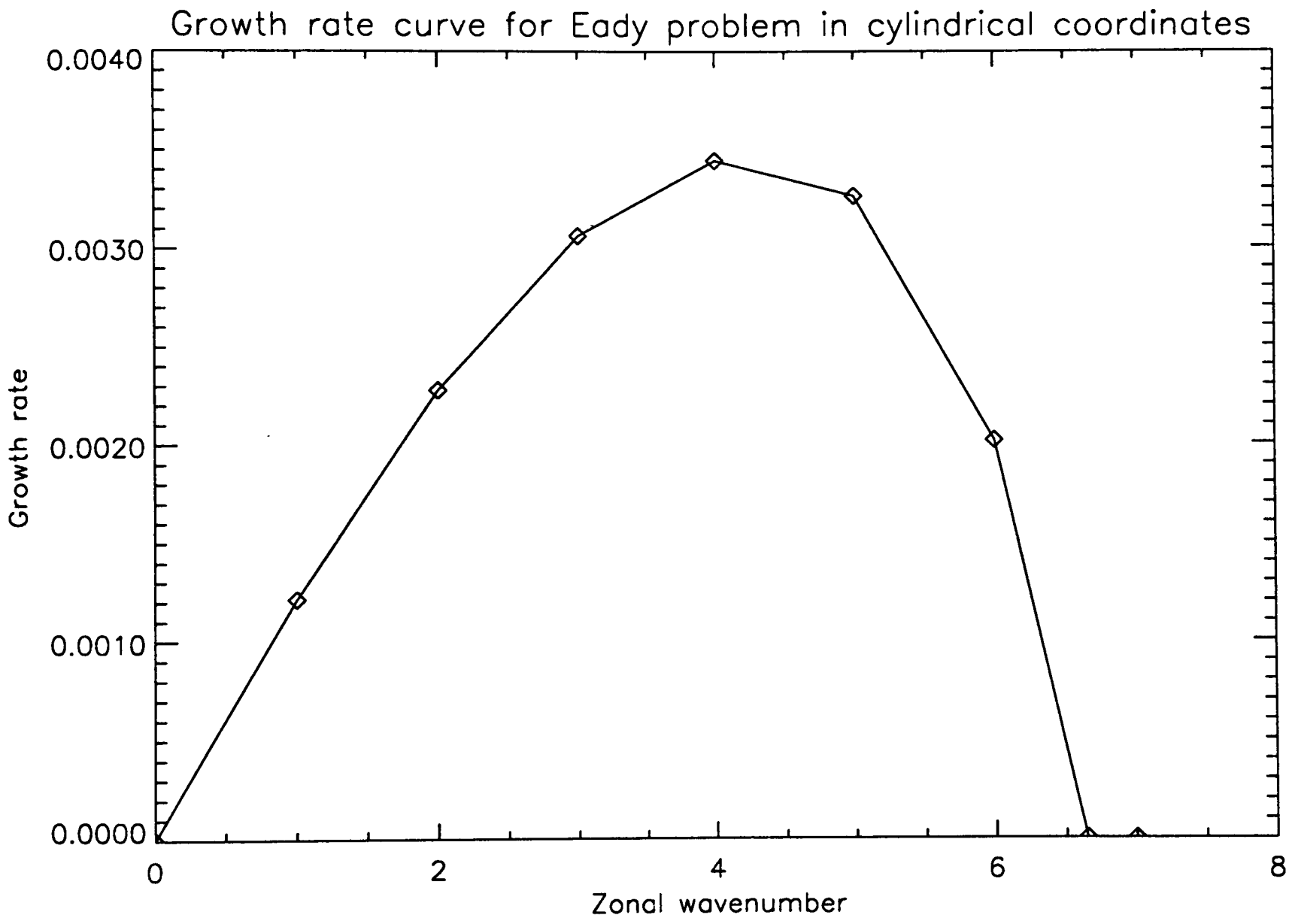


Figure 5.4: The variation of growth rate (s^{-1}) with zonal wavenumber for dominant, unstable modes associated with the Eady problem in cylindrical polar coordinates. The curve applies to a system with inner and outer radii of 0.025 and 0.080 m respectively, a rotation rate of 1.0 rad/sec and a Brunt-Väisälä frequency of 0.07 s^{-2}.

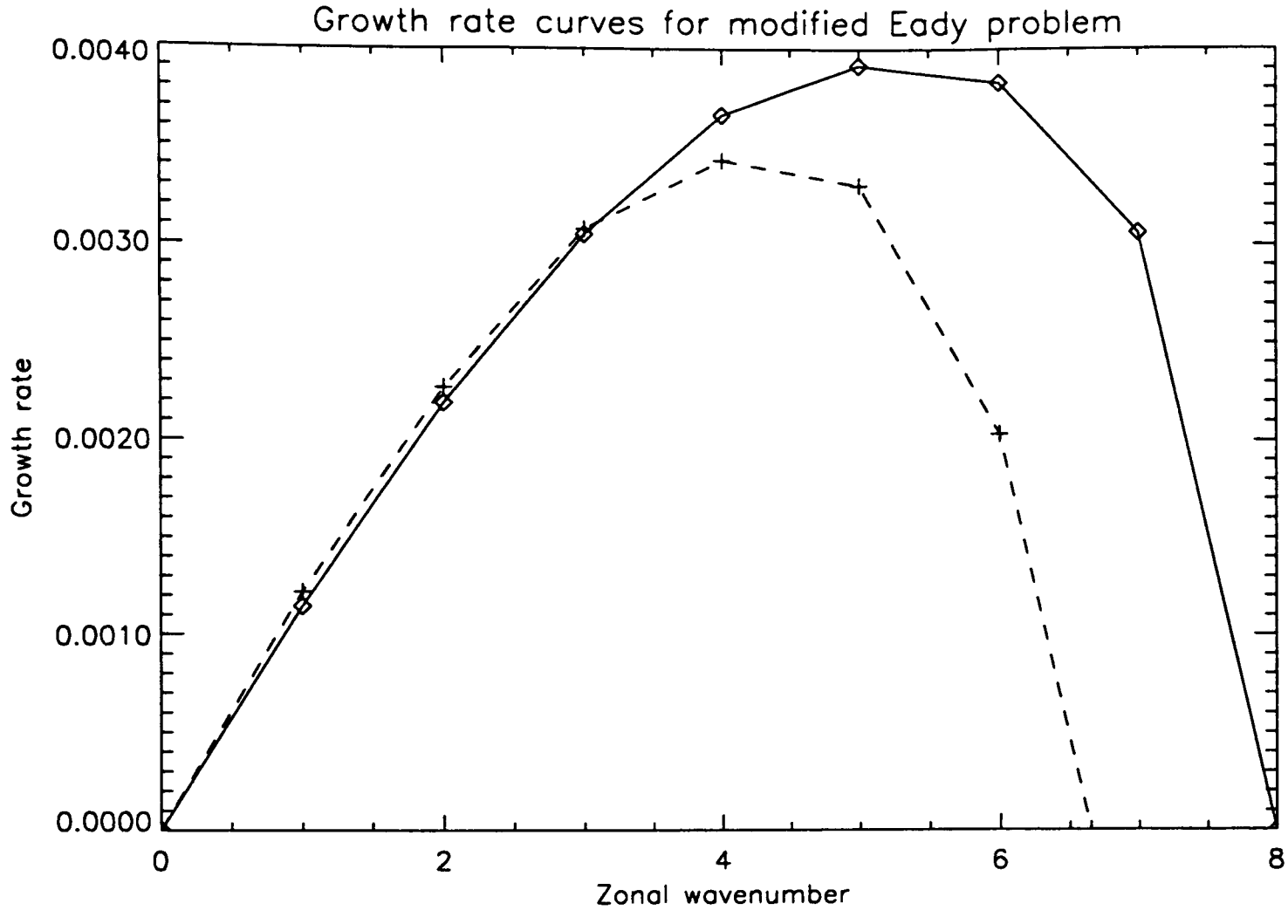


Figure 5.5: Results from analysis of the Eady problem for the annulus using INSTAB . The plot shows the growth rate of the most unstable mode for each zonal wavenumber, m, against m. The solid and dashed lines are the numerical and analytical solutions respectively. The fastest growing disturbance calculated numerically was at wavenumber 5, which is in reasonable agreement with the analytical solution.

least in the limit of the linear analysis used here), and so gives an indication of the type of flow we would expect to see initially, if the zonal mean state was properly representative of the real system.

In order to compile such a diagnostic, we had to run the code several times, each time specifying a different wavenumber to be studied. Each time the code was run, there were typically several unstable modes, each with a different growth rate. We used the growth rate of the fastest growing mode for the diagnostic. It is important to notice, however, that rounding errors and the like will often generate several spurious, unphysical modes. These were easily differentiated from the meaningful modes by their relatively low growth rates (typically several orders of magnitude smaller than that for the fastest growing mode), and their irregular meridional structure.

Figure 5.5 shows the dependence of the growth rate of the most unstable mode on the zonal wavenumber for a typical flow for the Eady problem. These results were obtained using the following parameters:

Rotation rate, Ω	$= 1.0$ rad s^{-1}
(Brunt-Väisälä frequency)^2, N^2	$= 0.07$ s^{-2}
Zonal velocity scale, $\bar{u}$	$= 0.005$ m s^{-1}
Kinematic viscosity, ν	$= 0.0$ m^{-2}s^{-1}
Numerical resolution	$= 20 \times 20$
Ekman pumping parameter, Ξ	$= 0.0$ s^{-1}

The annulus modelled had inner and outer radii of 0.025 m and 0.08 m respectively, and a height of 0.14 m, as in the analytical solution.

The background stream function was chosen to be of the form

$$\bar{\psi}(r, z) \propto r^2 z \tag{5.30}$$

so that the radial gradient of the associated background potential vorticity field would be zero, as required for the Eady problem.

It is clear from the diagram that there was a peak growth rate for wavenumber 5 disturbances for such a flow. This peak growth rate was ~ 0.004 s^{-1}, which gives an e-folding time of 250 s or a few tens of rotation periods. The form of the graph agrees well with the analytical solution for the growth rate (see Eq. 5.29), indicated by the dashed line. It is clear, however, that for wavenumbers greater than 3 the numerical solution for the growth rate is too large, resulting in the displacement of the peak growth rate and SWC to higher wavenumbers. This effect is related to the resolution used in the numerical discretisation, and is discussed below.

A further test of the performance of the code is how well the meridional structure of the growing modes which we have calculated agrees with Eady theory. For each complex eigenvalue, there is an associated complex eigenvector which contains the meridional

structure of the mode. If we write the perturbation stream function as

$$\psi' = \Re(\psi_R + i\psi_I)e^{i(m\theta - \omega t)}, \tag{5.31}$$

then the eigenvector calculated by INSTAB, $\varphi = (\varphi_R + i\varphi_I)$, is the representation of $(\psi_R + i\psi_I)$. This enables us to express the modulus and phase of the perturbation stream function as

$$|\psi'| \simeq \sqrt{\varphi_R^2 + \varphi_I^2} \tag{5.32}$$

and

$$\text{phase}(\psi') \simeq \tan^{-1}\left(\frac{\varphi_I}{\varphi_R}\right). \tag{5.33}$$

We have described above the analytical form of the meridional structure of the two fastest growing modes (see Eqs. 5.28 and 5.18 and Figure 5.3). The structure can only be determined up to an arbitrary scale (see Pedlosky, 1987), and so quantitatively, the numerical results may be different although qualitatively they should be the same.

The following figures show the meridional sections of the modulus and phase structure of the first two fastest growing modes from a typical run for the Eady problem. The parameters used for this run were identical to those above, pertaining to the growth rate curve, Figure 5.5. Figure 5.6 shows the structure of the fastest growing mode. Its structure agrees exactly with the theory (Figure 5.3, left hand figure), having only one turning point in the radial direction, slightly off centre. The perturbation stream function is vertically symmetric, and shows two waves trapped toward the lower and upper boundaries in the form of hyperbolic cosine variation in height. The overall scale of $\psi(r, z)$ is determined by the normalization procedure adopted by the NAG software.

It can be shown (Pedlosky, 1987) that the phase of ψ should vary almost linearly with height from 0 to $\pi/2$ and be independent of radius (latitude). This implies that the disturbances near the bottom of the system lead those at the top, and the mode can be thought of as leaning in a retrograde sense, or 'westwards', with height against the background zonal flow, and extracting available potential energy of the mean flow. In Figure 5.6 the phase of the unstable mode can be seen to vary in just such a way over the meridional section.

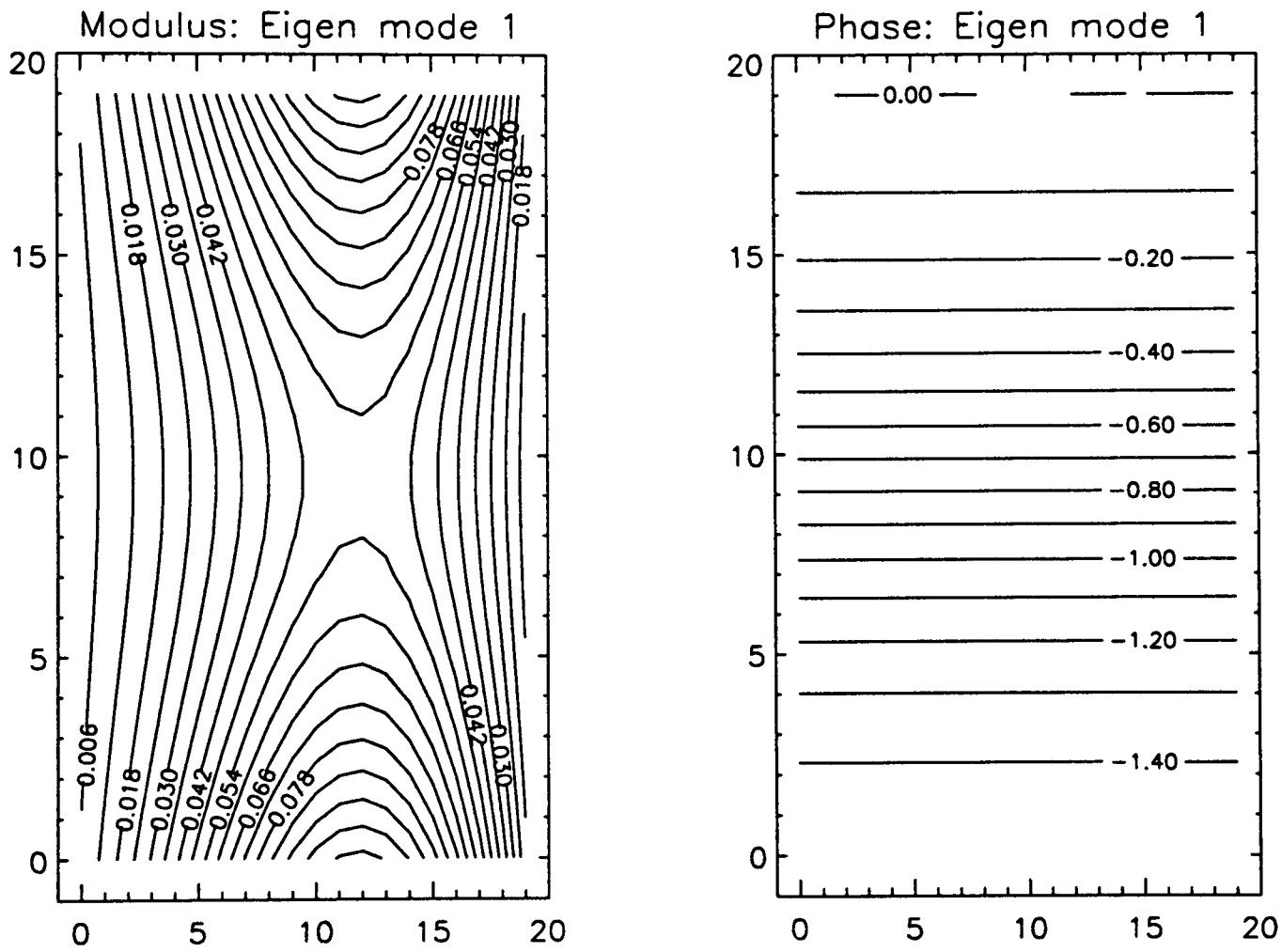


Figure 5.6: Results from analysis of the Eady problem for the annulus using INSTAB . In both of the diagrams the vertical and horizontal scales are the grid point indices in the vertical and radial directions respectively. Eigenmode 1 was the dominant eigenmode for wavenumber 5 perturbations to this flow, and its modulus and phase structure agree well with the theory.

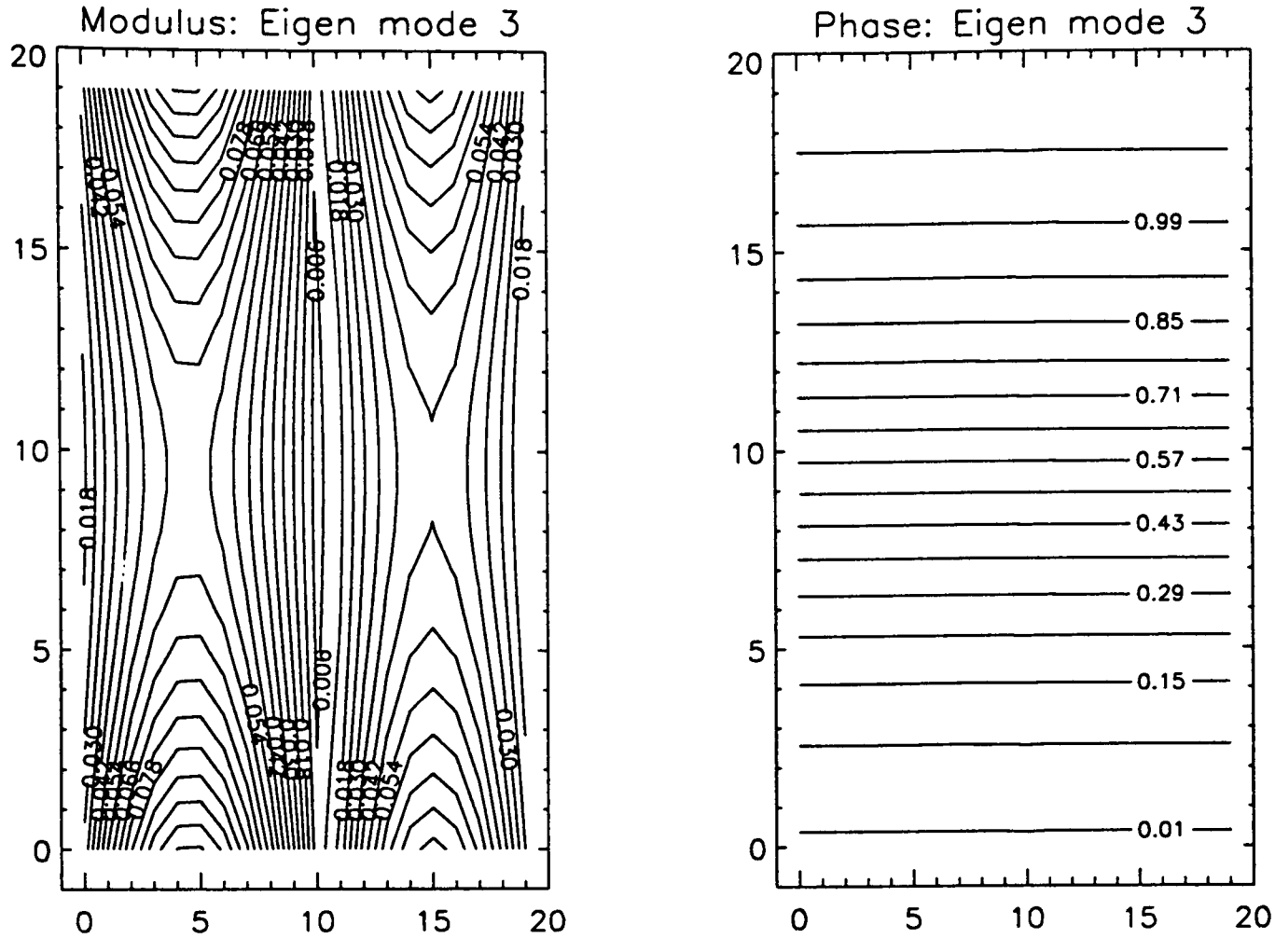


Figure 5.7: Results from analysis of the Eady problem for the annulus using INSTAB . The vertical and horizontal scales are as for the previous figure. Eigenmode 3 was the second fastest growing eigenmode at wavenumber 3 for this flow, and its modulus and phase structure also agree well with theory.

Figure 5.7 is similar to the previous figure, except that it shows the structure of the second fastest growing mode for wavenumber 3[†] perturbations to the same flow. The structure is again in exact agreement with the analytical solution for this mode (Figure 5.3, right hand figure), and the phase angle again varies only with height, as required.

5.2.3 The effect of grid resolution.

We have shown then that INSTAB can be used to reproduce the analytical solution to the modified Eady problem quite closely. The modulus and phase structure of the first and second most unstable Eady modes were in close agreement with the analytical solution. The growth rate curve (Figure 5.5) had the correct general form, however, the SWC and

[†]Wavenumber 3 had the strongest second most unstable mode.

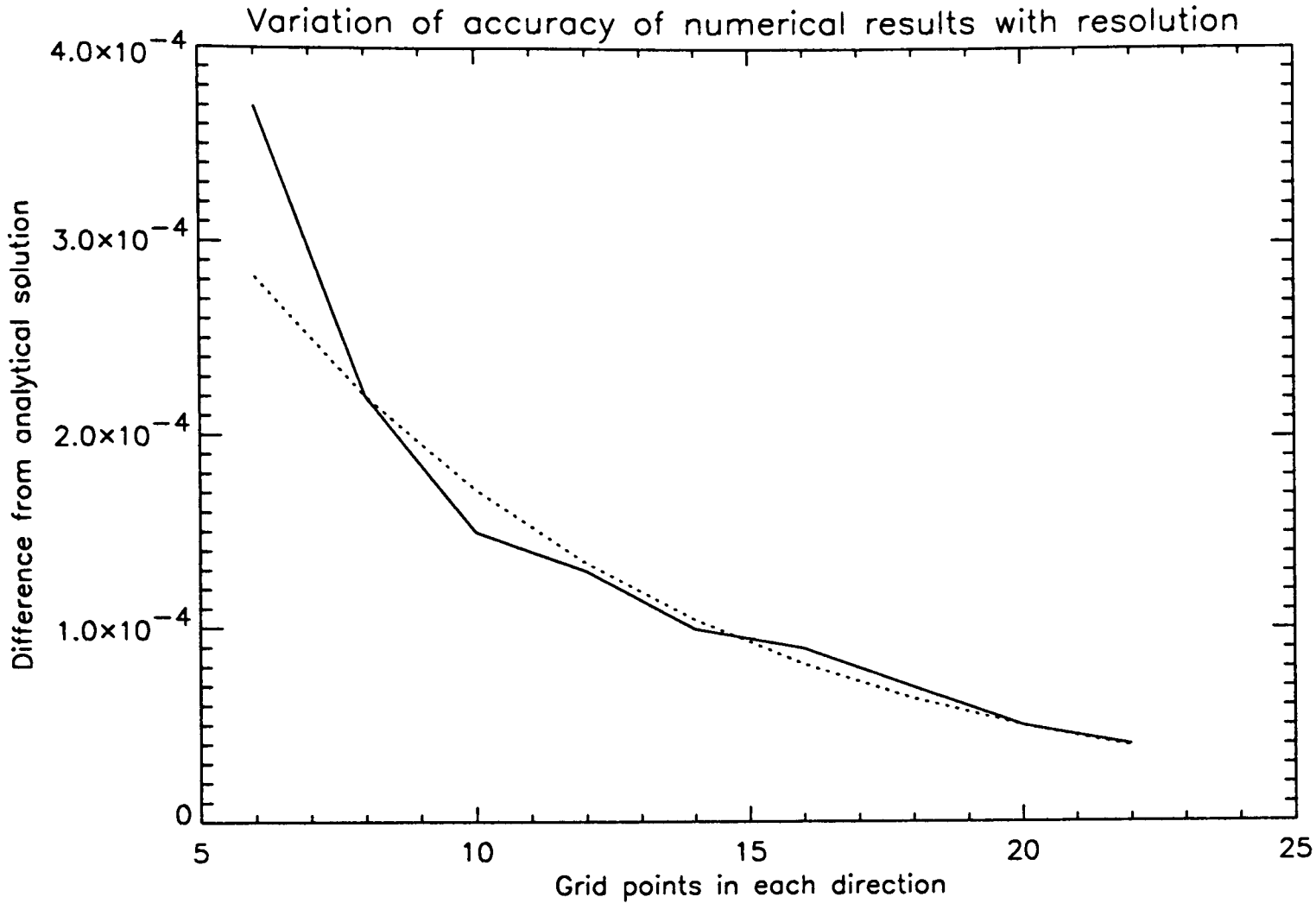


Figure 5.8: Variation of the accuracy of numerically calculated growth rates for a range of grid resolutions (solid line). All grids used have the same number of points in the radial and vertical directions. The dotted line shows an exponentially decaying function fitted to the data using a least squares fit (see main text).

peak growth rate occured at slightly larger wavenumbers than the analytical solution gave us. When we attempted to obtain the growth rate curve at a lower resolution, the general form of the curve remained the same, but the whole curve became further stretched and amplified. This transformation was such that the SWC moved out to a larger zonal wavenumber and the peak growth rate became larger and also moved to a higher wavenumber. Since it was apparent that increased resolution led to better agreement with the theory, we carried out a set of runs, at various resolutions, to investigate this effect. For each run the physical parameters were as for the previous sub-section, and wavenumber 1 disturbances were analysed. The results are shown in Figure 5.8, where the difference between the numerical and analytical values of the growth rate of wavenumber 1 disturbances is plotted as a function of resolution (solid line). The dotted line shows a least squares fitted exponential, $\Delta = 0.000639 * e^{-0.127m}$, where Δ is the error, and m is the zonal wavenumber. The fit is quite good, indicating that the difference between

the exact result and the numerical solution tends to decay exponentially as the resolution is increased. Because of computationally imposed constraints on memory, the highest resolution at which INSTAB can be used at the time of writing was 22×22. This could be improved by a change of hardware and/or operating system, although it is evident from Figure 5.8 that a large increase in resolution would be required to produce a significant increase in accuracy. The run-time may also become a constraint at very high resolutions, being approximately proportional to the cube of the number of meridional grid points.

5.2.4 The effects of frictional Ekman layers.

As we have discussed in the previous chapter and in the first section of this chapter, viscous Ekman layers at the lid and base of the annulus provide a reasonably realistic boundary condition which can be applied to the vertical velocity equation, and so to the formulation of the eigen value problem. If we use a similar method to that which gave us Eq. 5.29, but include Ekman layers, we obtain the following expression for the complex phase speed,

$$c = \left(\Upsilon z_T - 2i\mu\frac{\Xi}{m}\Omega\coth\mu z_T\right) \pm \left[\Upsilon^2 z_T^2 + 4\frac{\Upsilon^2}{\mu}\left(\frac{1}{\mu} - z_T\coth\mu z_T\right) + 4\mu^2\frac{\Xi^2}{m^2}\Omega^2\left(1-\coth^2\mu z_T\right)\right]^{\frac{1}{2}}. \quad (5.34)$$

Here, as before, we have used $\Xi = \pm\frac{N^2}{2\Omega}\frac{E_v^{\frac{1}{2}}}{2}$, where E_v is the Ekman number, $\sqrt{\frac{\nu}{2\Omega z_T^2}}$†. Eq. 5.34 can be used to determine the growth rate for waves in the modified Eady problem, for a given value of Ξ. The dashed line in Figure 5.9 shows the analytical form of the growth rate curve when $\Xi = 2.0\times10^{-5}\mathrm{s}^{-1}$. It is similar to that obtained from the solution in the absence of Ekman layers (Figure 5.5, dashed line), except that the growth rates are consistently lower. This is due to some energy being dissipated at the viscous boundaries, and so not contributing to the growth of the baroclinic waves.

We used INSTAB to obtain the solid line in Figure 5.9. The run parameters were as

†For values characteristic of the annulus system ($N^2 = 0.07$, $\Omega = 1.0$, $\nu = 1.0\times10^{-6}$, $z_T = 0.14$), the Ekman pumping parameter, $\Xi \sim 1\times10^{-5}$.

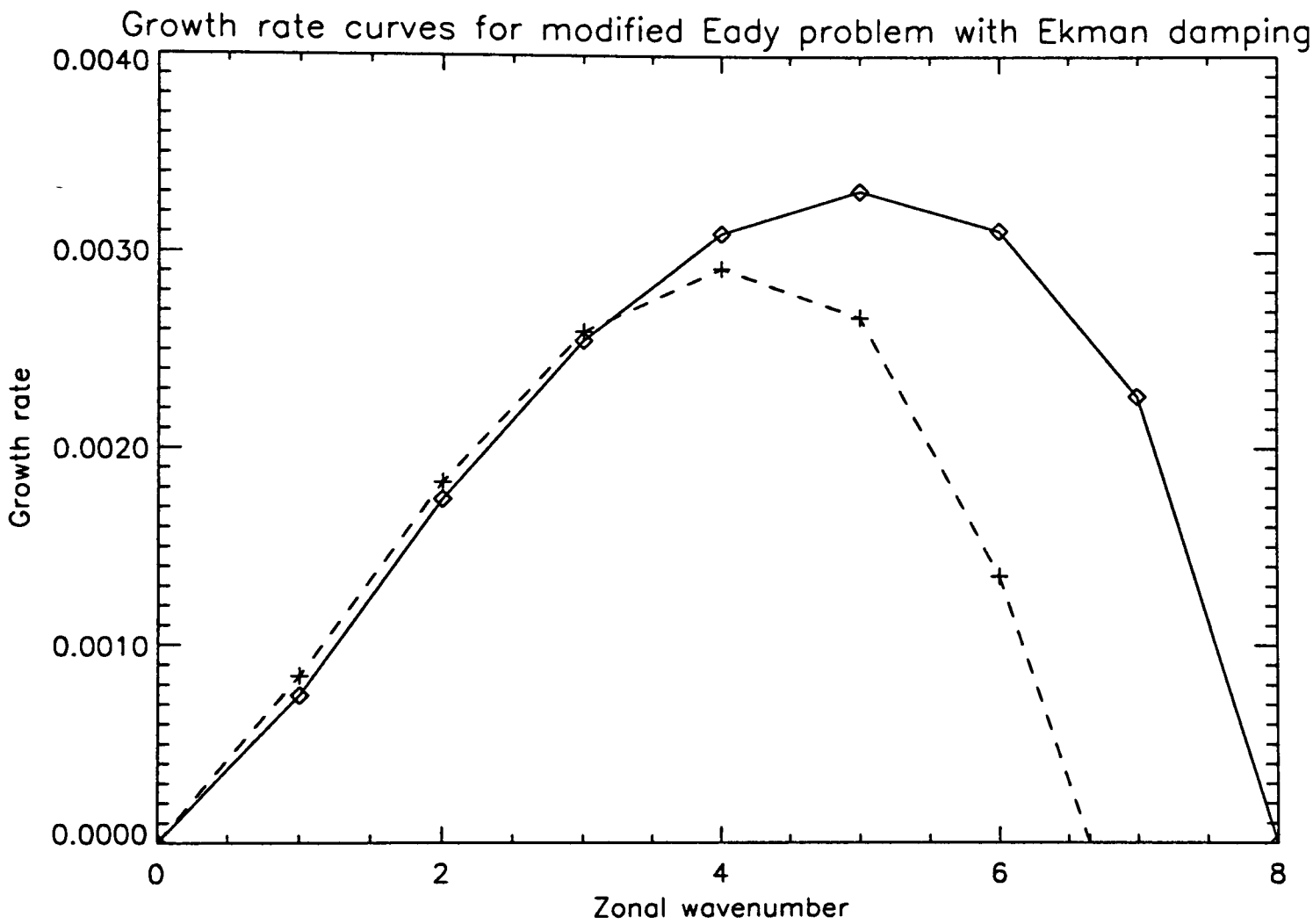


Figure 5.9: Results from analysis of the Eady problem with frictional Ekman layers for the annulus, using INSTAB . The plot shows the growth rate of the most unstable mode for each zonal wavenumber, m, against m. The solid and dashed lines are the numerical and analytical solutions respectively. The curves are similar to those for the case without Ekman damping, the growth rates being somewhat lower in this case, due to the dissipative energy loss.

follows:

Rotation rate, Ω	$= 1.0$ rad s^{-1}
(Brunt-Väisälä frequency)^2, N^2	$= 0.07$ s^{-2}
Zonal velocity scale, $\bar{u}$	$= 0.005$ m s^{-1}
Kinematic viscosity, ν	$= 1.01 \times 10^{-6}$m^{-2}s^{-1}
Numerical resolution	$= 20 \times 20$
Ekman pumping parameter, Ξ	$= 2.0 \times 10^{-5}$s^{-1}

The annulus modelled had inner and outer radii of 0.025 m and 0.08 m respectively, and a height of 0.14 m, as in the analytical solution. Again, the curve resembles that determined without Ekman layers (Figure 5.5, solid line), except for the decreased growth rates. The numerical result is close to the analytical solution, but not identical. The difference is a result of the finite grid resolution used, and could be reduced if higher resolution calculations were possible.

If the wave number is fixed for a particular system, Eq. 5.34 can be used to evaluate the dependence of the growth rate on the Ekman pumping parameter, Ξ. In Figure 5.10, the dotted line shows the growth rate the most unstable wave, $m = 4$, for a range values of the Ekman pumping parameter. For very small values ($< 1 \times 10^{-6}$s^{-1}), the Ekman damping has very little dissipative effect, and the wave grows at much the same rate as it would in the absence of viscous boundary layers. As Ξ is increased, more of the energy is lost to dissipation, and the growth decreases, until above some critical value ($\sim 1.3 \times 10^{-4}$s^{-1}) the wave is no longer unstable.

Also plotted in Figure 5.10 (solid line) is the growth rate of the most unstable wave, $m = 5$, from our numerical analysis. The curve is clearly very similar to the analytical solution. The difference is again due to finite resolution effects, which cause the dominance of different wavenumber flows in the analytical and numerical solutions.

We have demonstrated that INSTAB can produce reasonable numerical solutions to the Eady problem in cylindrical polar coordinates, with and without Ekman layers at the top and bottom boundaries. We have also shown that these solutions tend to closer

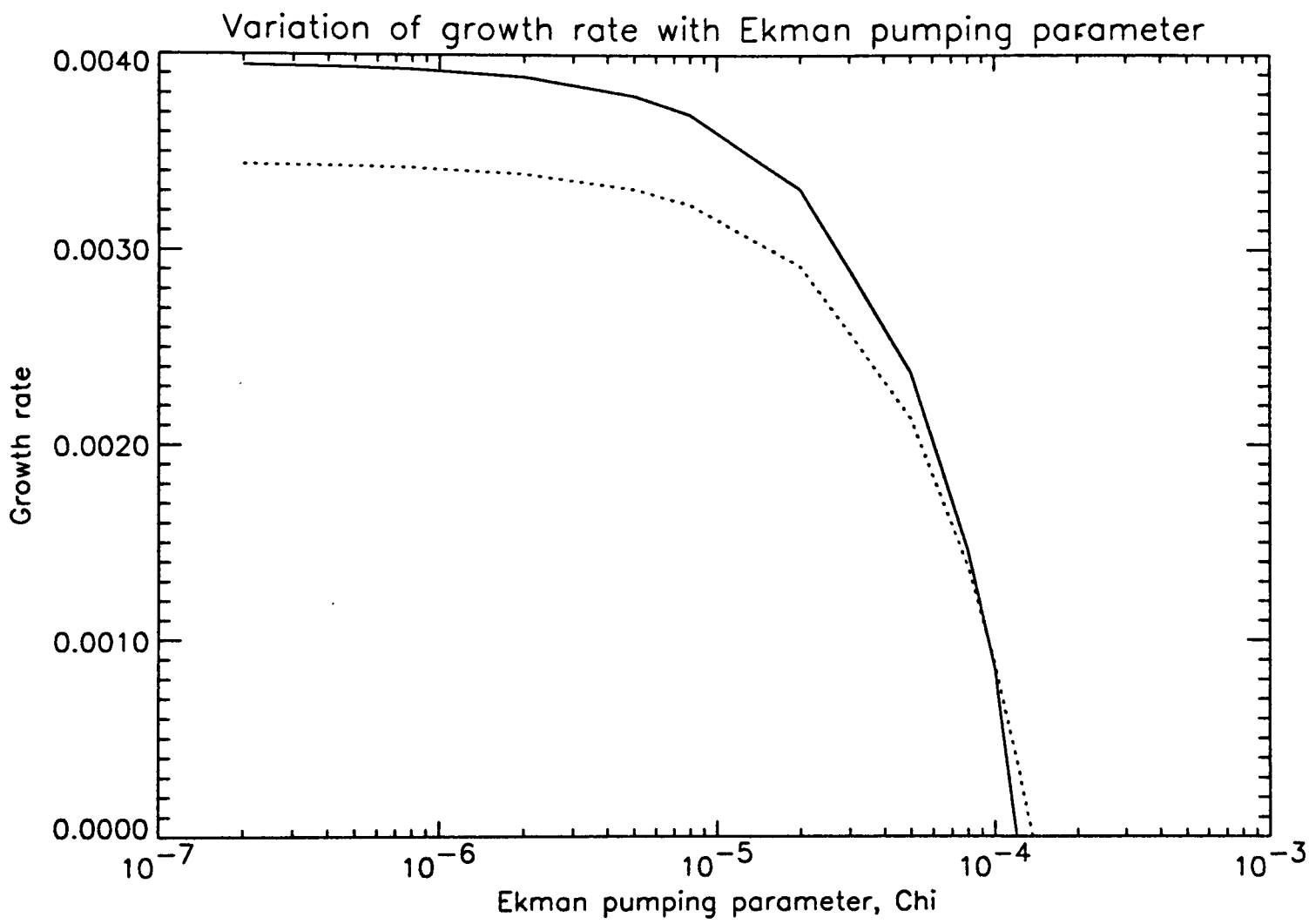


Figure 5.10: Numerical (solid line) and analytical (dotted line) curves, showing the dependence of the growth rate of the dominant wave (s^{-1}) on the magnitude of the Ekman damping (s^{-1}). Beyond a critical value, the wave does not grow at all.

agreement with the analytical solutions as the resolution of the numerical grid is increased. In the following section, we discuss the use of INSTAB to investigate the linear stability of various internal baroclinic jets (Bell and White, 1988 and Bell, 1989). These flows, whilst exhibiting a SWC, are subject to a lateral shear. This makes the lateral gradient of the quasigeostrophic potential vorticity in general non-zero, and so constitutes a more representative system for analysis.

5.3 Experiments with internal baroclinic jets.

In this section we examine the baroclinic stability of internal sinusoidal jets in the annulus system. Bell (1989) discusses three sets of such jets in his study of the stability of various flows, each of which can be related to observed flows in the laboratory system. A first set have the form

$$\bar{u} \propto -\cos(\pi\hat{z})\cos(\eta\pi[\hat{r}-0.5]), \tag{5.35}$$

where $\hat{u}$ and $\hat{z}$ are the non-dimensional radial and height coordinates, ranging in value from 0 to 1 linearly across the domain, and η is a measure of the radial (latitudinal) curvature of the flow. This is a good representation of the type of dominant axisymmetric zonal flow observed in wall heated experiments. All the results from INSTAB presented in this section use this type of zonal flow structure. Bell (1989) also mentions a set of jets given by

$$\bar{u} \propto -\cos(\pi\hat{z})\sin(\eta\pi[\hat{r}-0.5]), \tag{5.36}$$

which "resemble quite closely" axisymmetric flow measured in internally heated experiments (Read, 1985). Figure 5.11 shows the structure of the mean flows described by Eqs. 5.35 and 5.36 in the left and right plots respectively, with a value of η of 1.0. Figure 5.12 shows similar plots for a value of η of 0.5, and the decrease in the degree of radial variation of the flow can be clearly seen. Zonal velocity fields resulting from the integration of the Met. Office model for wall heated and internally heated flows are shown in Figures 5.17 and 5.23 respectively, and a degree of similarity is apparent.

These sets of sinusoidal jets were studied by Bell (1989), (a) because they exhibit a

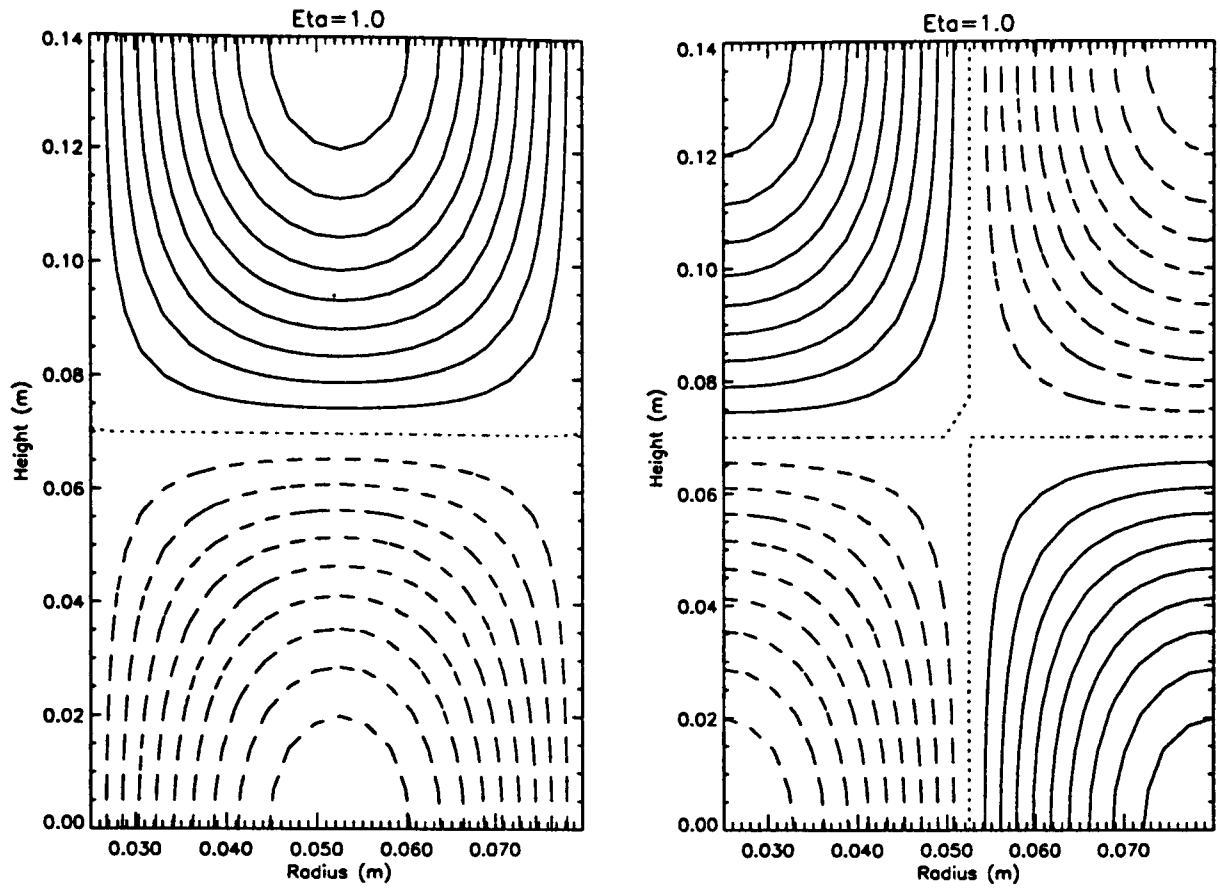


Figure 5.11: The zonal mean zonal velocity as described by the two sets of internal baroclinic jets described in the text. For these flows, the curvature parameter, $\eta = 1.0$. Solid lines show prograde flow and dashed lines retrograde flow. The horizontal and vertical axes are radius and height in metres, respectively.

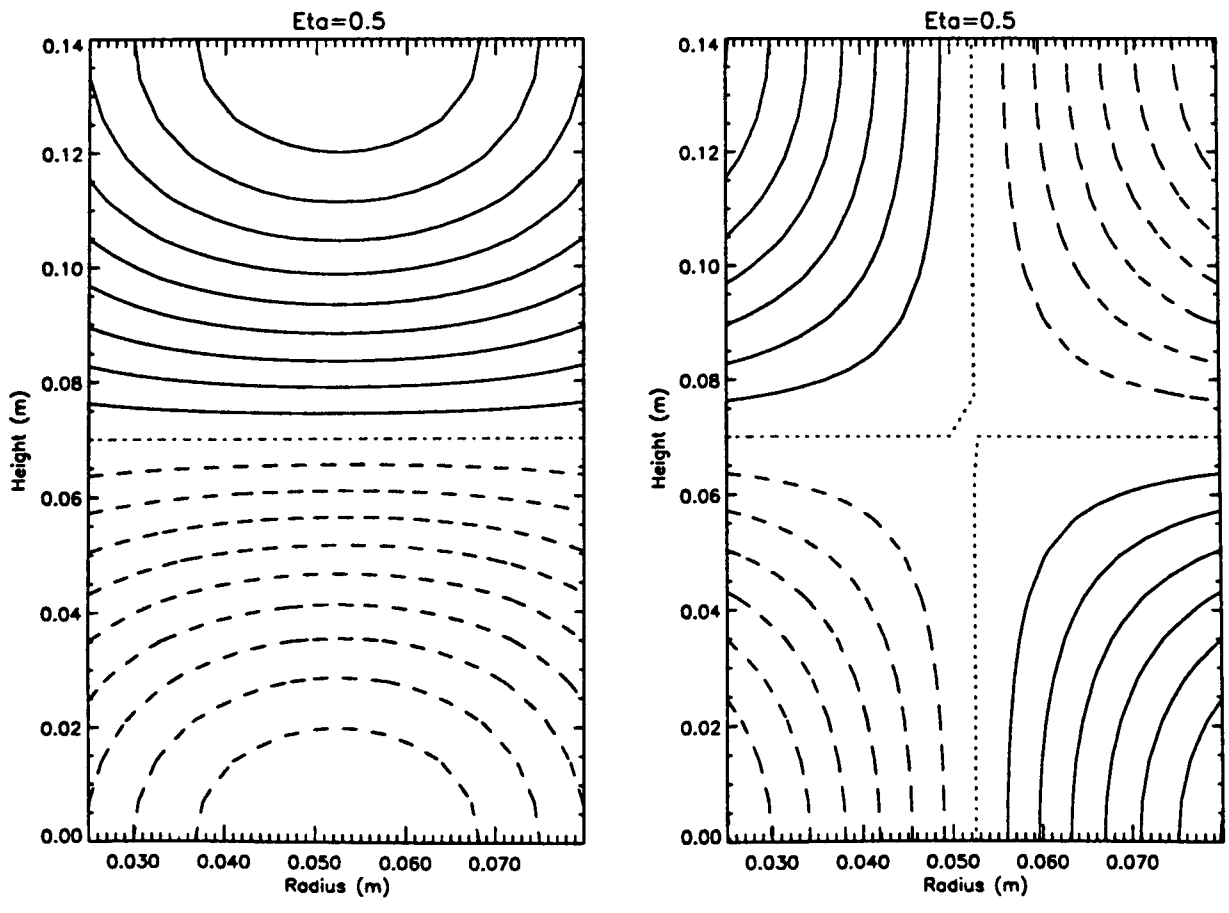


Figure 5.12: The zonal mean zonal velocity profiles for baroclinic jets with a curvature parameter, $\eta = 0.5$. Details as for the previous figure.

short-wave cut-off (see previous Section), and (b) because he was able to obtain analytical expressions for the short wave cut-off wavenumber as a function of Burger number and the flow curvature. Both of the flows described by Eqs. 5.35 and 5.36 satisfy the criteria necessary for baroclinic instability given by Charney and Stern (1962), which require a stratified flow bounded by horizontal surfaces to have a latitudinal potential vorticity gradient which changes sign, and constant (potential) temperature at the boundaries. The form for the flow which we have used here proved to be particularly useful, and we are able to calculate quite well the dominant wavenumber perturbation to flows in the wall heated annulus and the mid-latitude terrestrial atmosphere.

5.3.1 Results for sinusoidal internal jets.

Figure 5.13 shows the meridional structure of the fastest growing mode for a flow with the following parameters:

Zonal wavenumber, m	$= 6$
Rotation rate, Ω	$= 1.0$ rad s^{-1}
(Brunt-Väisälä frequency)^2, N^2	$= 0.07$ s^{-2}
Zonal velocity scale, $\bar{U}$	$= 0.005$ m s^{-1}
Flow curvature, η	$= 0.5$
Kinematic viscosity, ν	$= 1.01\times10^{-6}$ m^{-2}s^{-1}
Numerical resolution	$= 20 \times 20$
Ekman pumping parameter, Ξ	$= 1.0 \times10^{-5}$s^{-1}

The annulus modelled had inner and outer radii of 0.025 m and 0.08 m respectively, and a height of 0.14 m. The value of Ξ was chosen such as to be large enough to have an appreciable effect, but not so large as to damp out the waves altogether (see previous section).

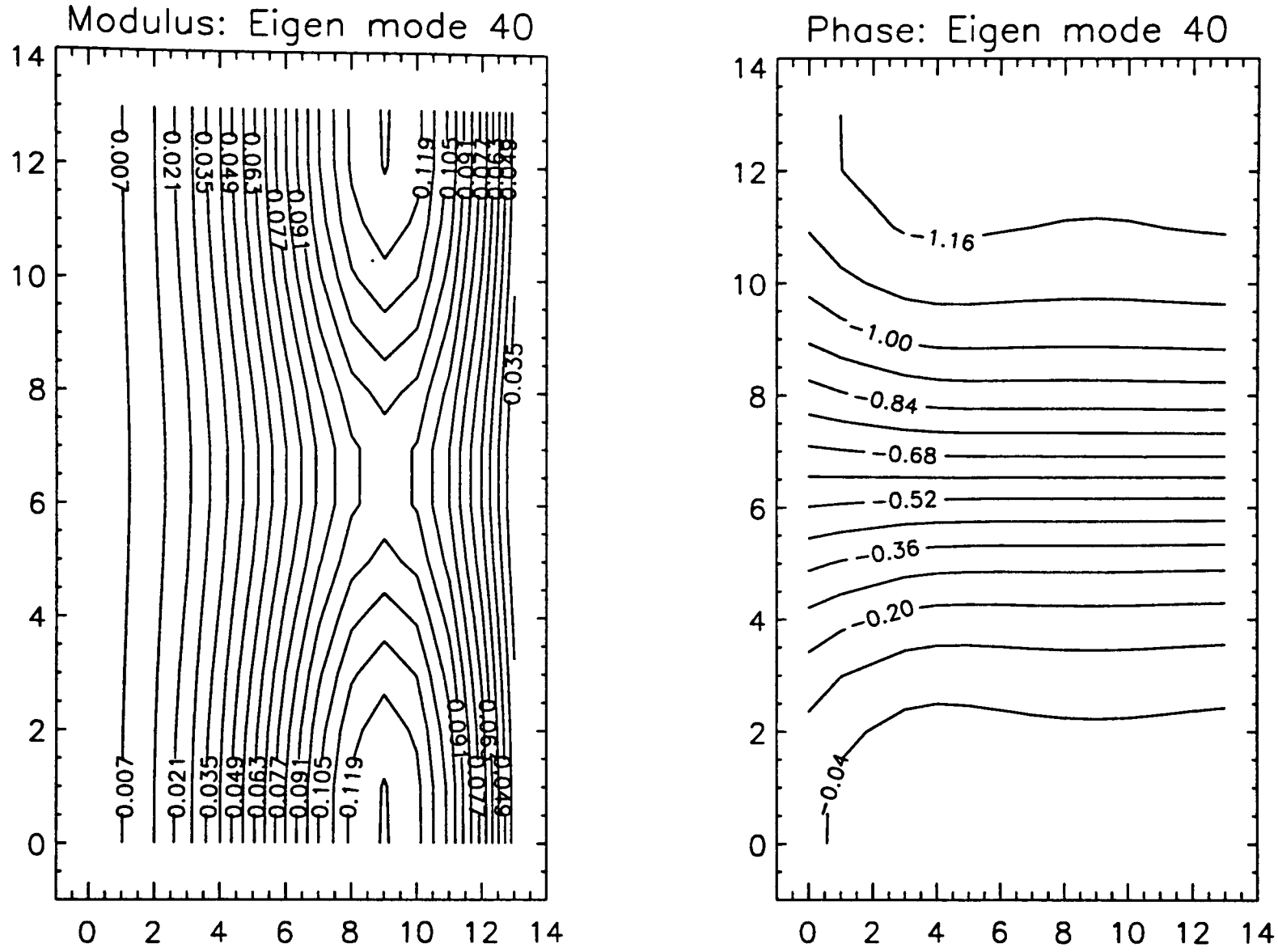


Figure 5.13: Results from analysis of internal jets in the annulus using INSTAB. In both of the diagrams the vertical and horizontal scales are the grid point indices in the vertical and radial directions respectively. Eigenmode 40 was the dominant eigenmode for wavenumber 6 perturbations to this flow, this being the most unstable wavenumber.

The background zonal velocity field used for this and each of the other runs in this section was

$$\bar{u} = -\bar{U}\cos(\pi\hat{z})\cos(\eta\pi[\hat{r} - 0.5]), \tag{5.37}$$

as above, with an intermediate value of 0.5 used for η (see Figure 5.12, left hand plot). The vertical structure of the dominant mode is similar to that which would be expected for the Eady problem. This would be anticipated since as $\eta \to 0$, there is less lateral shear in the background state, and the radial structure of the flow approaches a state similar[†] to that used in Eady's model. The phase angle of the mode varies monotonically

[†]The zonal flow in the Eady problem varies linearly with height, whereas here it varies as the cosine of the height.

with height, and near mid-height the variation is almost linear. This is the region where the background state most closely resembles the uniformly sheared vertical flow used in Eady's model, and so this similarity is a further validation of our numerical scheme. The lack of pronounced radial (latitudinal) variation in the phase of the growing mode shows that it is almost purely baroclinic.

In order to investigate the effect of the lateral shear of the background flow on the development of instabilities, we used INSTAB to numerically determine the growth rate curves for a set of flows with different degrees of shear. The parameters used for these runs were:

(Brunt-Väisälä frequency)^2, N^2	$= 0.07\ \mathrm{s}^{-2}$
Zonal velocity scale, $\bar{U}$	$= 0.005\ \mathrm{m\ s}^{-1}$
Kinematic viscosity, ν	$= 0.0\ \mathrm{m}^{-2}\ \mathrm{s}^{-1}$
Numerical resolution	$= 20 \times 20$
Ekman pumping parameter, Ξ	$= 0.0\ \mathrm{s}^{-1}$

The annulus dimensions were as for the previous description. The growth rate curves are shown in Figure 5.14. Each of the curves appear to jump suddenly to a SWC, but this an artifact of only performing the calculations for integral wavenumber flows. All the curves would in fact turn smoothly (albeit steeply) down to zero growth rate if they could be resolved. The steepness results from these runs having been performed without Ekman damping at the end walls. This acts to limit the growth of the higher wavenumber instabilities, and give a more symmetric growth rate curve. It is clear from the figure that increasing the lateral shear acts to limit the growth of the unstable waves (see Bell 1989). The remainder of our analysis considers jets with $\eta = 0.5$, because at higher values of η, the flow curvature becomes such that the quasigeostrophic approximation is no longer valid (the relative vorticity becomes comparable to the planetary vorticity, 2Ω). Larger values of η also necessitate the use of finer resolution to capture the radial structure of the flow.

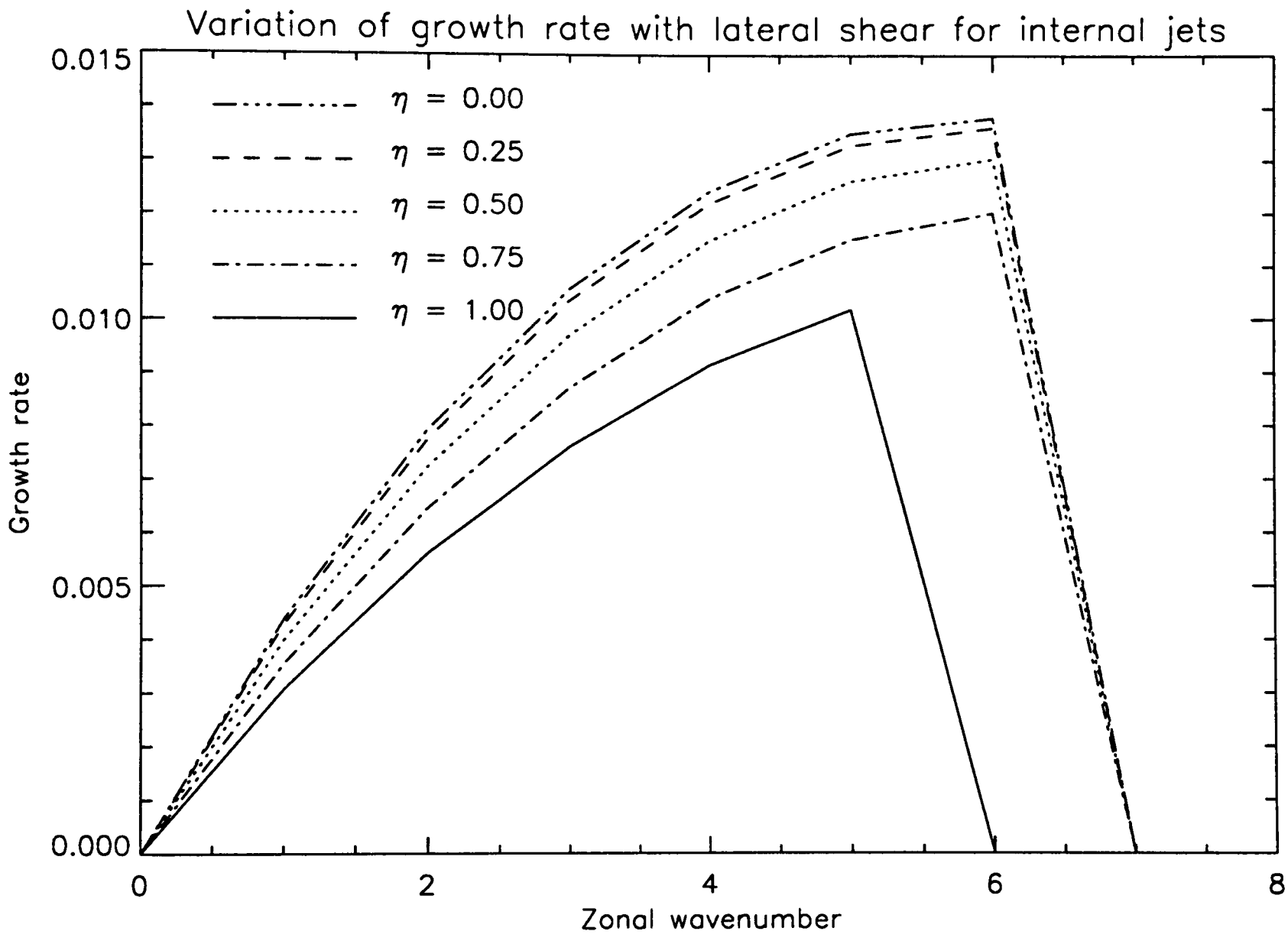


Figure 5.14: Results from analysis of internal baroclinic jets with various degrees of lateral shear, using INSTAB . The curves show the growth rate of the most unstable mode for each zonal wavenumber m against m, for a range of curvatures from 0 (laterally uniform) to 1 (half sine lateral variation).

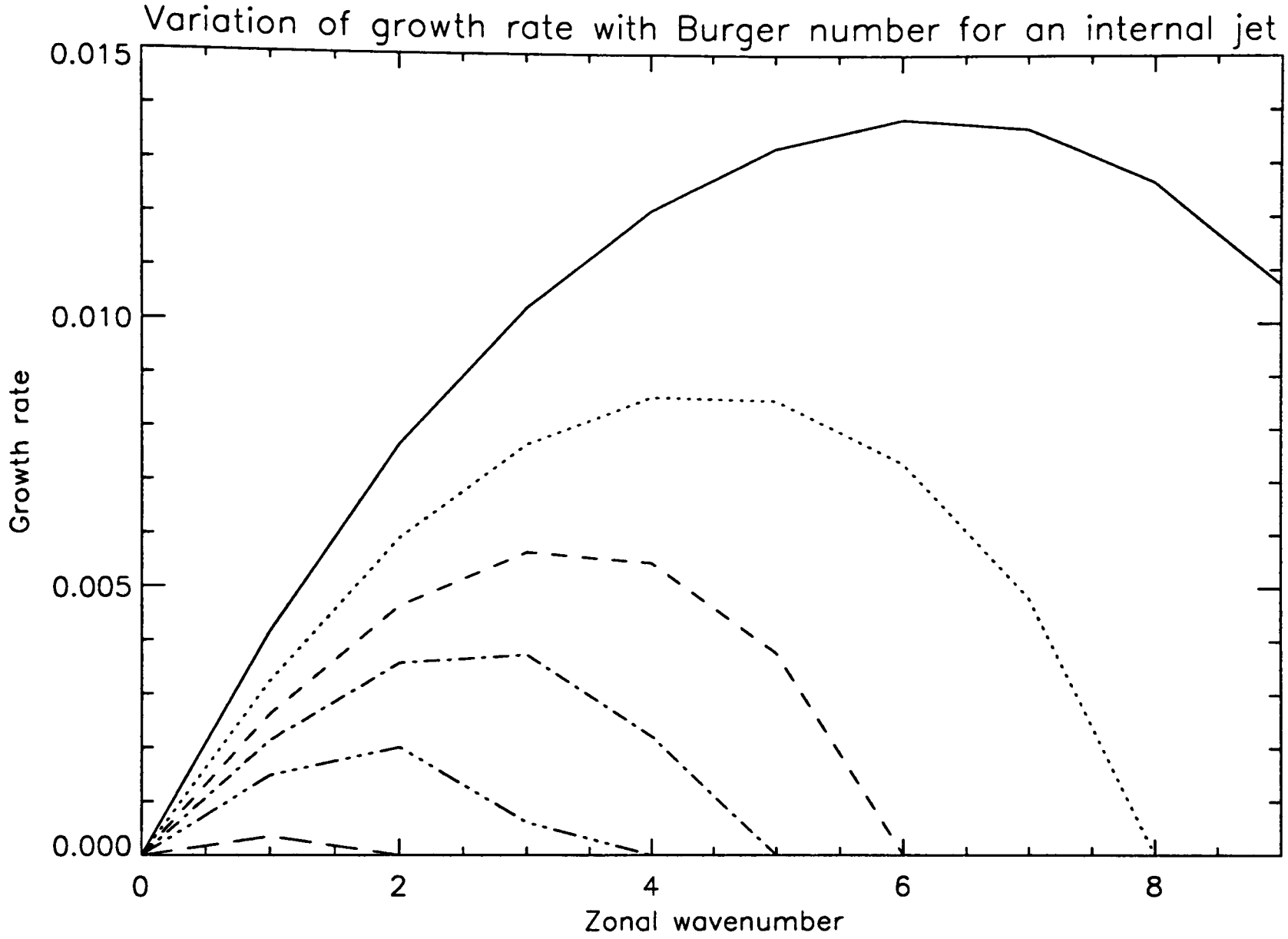


Figure 5.15: Results from analysis of internal baroclinic jets for the annulus using INSTAB . The curves show the growth rate of the most unstable mode for each zonal wavenumber m, against m. The different curves are for flows with Burger number 0.113 (solid), 0.202 (dots), 0.315 (dashes), 0.454 (dash-dot), 0.709 (dash-dot-dot-dot) and 1.260 (long dash).

Figure 5.15 illustrates the variation of the shape of the growth rate curves with the applied Burger number, B. For each of the curves given in the figure, the rotation rate was varied to select a different Burger number, and the other parameters were as follows:

(Brunt-Väisälä frequency)^2, N^2	$= 0.07\ \mathrm{s}^{-2}$
Zonal velocity scale, $\bar{U}$	$= 0.005\ \mathrm{m\ s}^{-1}$
Flow curvature, η	$= 0.5$
Kinematic viscosity, ν	$= 1.01\times10^{-6}\ \mathrm{m}^{-2}\ \mathrm{s}^{-1}$
Numerical resolution	$= 14 \times 14$
Ekman pumping parameter, Ξ	$= 1.0 \times10^{-5}\mathrm{s}^{-1}$

The annulus modelled had inner and outer radii of 0.025 m and 0.08 m respectively, and a height of 0.14 m, as before. These runs were carried out at a relatively coarse resolution because of the large number of runs required to generate the data for the figure.

Despite the fact that truncation errors (and perhaps the internal functioning of the NAG software) may mean that the growth rates shown here are somewhat inaccurate (Simmons and Hoskins, 1976), for a given grid resolution this is a systematic effect and the relative sizes of the calculated growth rates still provide useful information. For each curve there is a SWC, above which there are no unstable perturbations. It can be seen from each of the curves that the growing mode with the fastest growth rate has a wavelength of about twice the SWC. As the (destabilising) imposed rotation rate was made greater (*i.e.* $B \rightarrow 0$), a greater number of wavenumbers became unstable, and the dominant flow tended to have a shorter zonal wavelength.

In a study of the baroclinic instability of the northern hemisphere winter zonal flow, Wyatt (1981) obtained a similar diagram for the variation of theoretical growth rate curves with decreasing latitude. This was despite the use of a different formulation of the eigenvalue problem, which solved the matrix relating the time variation of the system to its overall state. This is relevant here, since Wyatt used a static stability profile which increased monotonically with latitude (when averaged over the lowest three model levels). Since the rotation rate was fixed, this implied an increase in Burger number with latitude, and so the observed variation in growth rate is consistent with our results.

The previous observation leads naturally to the idea of a "critical Burger number", B_S for a given zonal wavenumber m. This is the Burger number for marginal stability of the wave, *i.e.* the largest value of the Burger for which wavenumber m disturbances are unstable. Figure 5.16 shows how B_S determined numerically and analytically varies with zonal wavenumber, m, for a system with the physical parameters given earlier in this section[†]. The analytical solution (Bell, 1989) is,

$$B_S = \frac{1}{1 + (m^2/\pi^2) - \eta^2}. \tag{5.38}$$

[†]These results are, of course, only applicable to flows with a zonal mean structure identical to that used in the calculation, see Eq. 5.37.

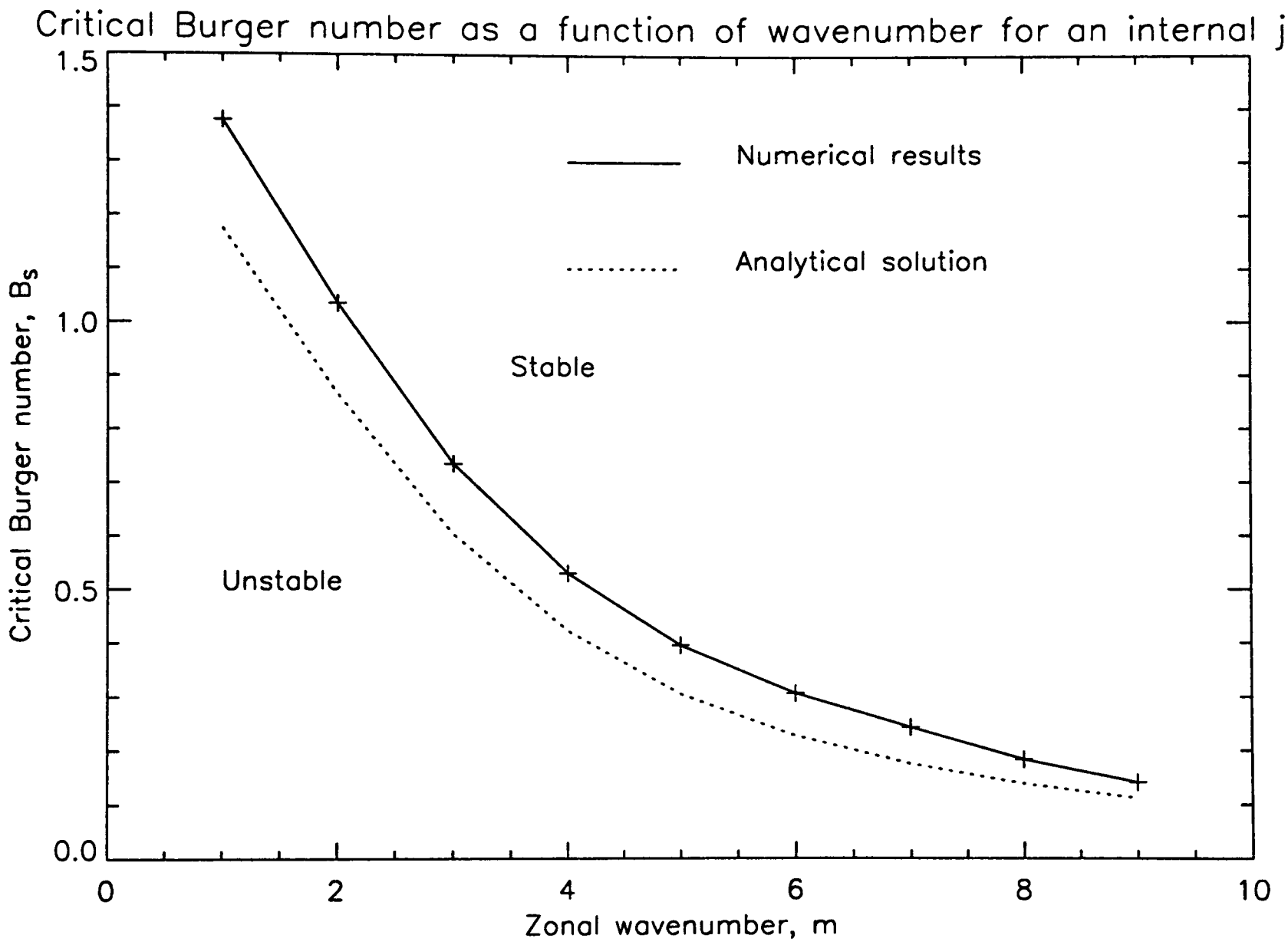


Figure 5.16: Analytical and numerical curves showing the variation of the critical Burger number with wavenumber. The critical Burger number for wavenumber m, is the value above which that wave is no longer unstable.

For values of $\eta < 1$, B_S increases as η increases, and as $\eta^2 \to 1 + m^2/\pi^2$, $B_S \to \infty$. For larger values of η, B_S would be negative, and no baroclinic waves would grow.

Each of the numerical data points in Figure 5.16 were calculated using a "Hunter-Catcher" method (Press *et al.*, 1989) to vary the Burger number systematically for each wavenumber until the critical value was found. The agreement between the two curves is satisfactory, and again would almost certainly improve with increased resolution. A graph of this type does, however, require a very large number of runs to produce, and at a high resolution would be very time consuming to construct.

For a given flow, we can calculate the Burger number and use Figure 5.16 to see which wavenumber disturbance can be expected to prevail. For example, if $B = 0.6$, then the highest wavenumber flow which would be unstable is $m = 3$, since $m = 4$ lies on the stable side of the critical line. In this case we would expect the dominant wave to be either $m = 1$ or 2, since this is half the maximum unstable wavenumber (see Figure 5.15).

We attempted to naively apply this "regime diagram" by using the Burger number of some known flows, assuming that they have the required meridional structure, and determining the associated dominant wavenumber by inspection.

The annulus experiment For a typical annulus flow used above, we have a value for N^2 of 0.07 s^{-2}, a rotation rate $\Omega = 1.0$ rad s^{-1} and an aspect ratio of ~ 3. This gives a Burger number $B = 0.2$ and so from the graph we can expect $m = 8$ to be the maximum observed wavenumber, and $m = 4$ to be the dominant flow for such a system. This agrees well with typical laboratory measurements and results from the Met. Office numerical model described by Hignett *et al.* (1985) (see also Figure 1.6 of this thesis).

The Earth's mid-latitude atmosphere The time scale for buoyancy driven oscillations in the atmosphere is ~ 600 s (Houghton, 1986). From this we can estimate a value for N^2 of $\sim 1 \times 10^{-6}$ s^{-2}. If we use the Earth's rotation rate and a ratio of length scales of 10Km:1000Km (height:width) then the relevant Burger number is $B \sim 0.17$. Using Figure 5.16 we can see that the dominant zonal wavenumber

will be ~ 10 which, at mid-latitude, corresponds to a length scale of 700–2800 Km. This is in good agreement with observed length scales (see, for example, Hoskins *et al.*, 1985).

Despite the success of the preceding analyses, it is again important to consider the *caveat* concerning the internal jet structure, which we have assumed. This was a reasonable approximation, particularly to the differentially heated annulus, and to a lesser extent to the mid-latitude terrestrial atmosphere, and the success of the regime diagram in predicting the observed flows is reassuringly rather than surprising. Before we could perform this kind of analysis in earnest to investigate the linear stability of some other system, it would be necessary to repeat the procedure described here using some other zonal flow more representative of the system we wished to consider.

5.4 Experiments concerning the stability of axisymmetric states from a numerical model.

In this section we use INSTAB to examine the stability of various mean states calculated using the Met. Office numerical annulus model (Farnell and Plumb, 1976, Hignett *et al.*, 1985, White, 1986). This model was run by spinning up an axisymmetric version of the same code (Farnell and Plumb, 1975) with the chosen parameters to a steady state, and then using this state, with some small perturbation added to it, as the initial fields for the full three dimensional model. INSTAB was used with a mean state which was the zonal average of the model state shortly (200 time-steps, or 4 seconds) after the perturbation was added. We were then able to compare the behaviour of the model after some time with the results of our stability analysis. Listed below are the four model runs which we discuss in this section:

Run 408 A differentially heated flow which developed a steady wave three component.

The parameters used for this flow were:

Run time = 4.0 s

Time step	= 0.02 s
Inner radius, r_a	= 0.025 m
Outer radius, r_b	= 0.080 m
Depth	= 0.14 m
Temperature, $T(r = r_a)$	= 18.0 deg C
Temperature, $T(r = r_b)$	= 22.0 deg C
Rotation rate, Ω	= 1.0 rad s^{-1}

where the run time referred to is measured from the introduction of the perturbation to the axisymmetric flow (which was itself run for 1600 seconds to reach a steady state).

Run 805 A differentially heated flow which was stable to the perturbation imposed on it. This run used the same parameters as those given for Run 408 except that the rotation rate in this case was 0.2 rad s^{-1}.

Run 557 An internally heated flow which developed a steady wave four component. The parameters used for this flow were:

Run time	= 4.0 s
Time step	= 0.02 s
Inner radius, r_a	= 0.0406 m
Outer radius, r_b	= 0.0855 m
Depth	= 0.12 m
Temperature, $T(r = r_a)$	= 19.0 deg C
Temperature, $T(r = r_b)$	= 19.0 deg C
Rotation rate, Ω	= 1.22 rad s^{-1}
Internal heating rate	= 100.0 W

where the internal heating is provided by Ohmic dissipation of an alternating electric current applied across the conducting walls of the annulus. For a detailed description of the internally heated laboratory system, see Read (1986a).

Run 755 An internally heated flow which was stable and remained axisymmetric. This run used the same parameters as Run 557, but with the rotation rate decreased to 0.1 rad s^{-1}.

5.4.1 Differentially heated, regular wave flow.

Figure 5.17 shows the mean state from Run 408 which was analyzed with INSTAB . The left hand plot shows the zonal velocity divided by radius (or relative angular velocity), the centre plot shows the radial gradient of the quasigeostrophic potential vorticity, divided by the radius, and the right hand plot shows the horizontally averaged profile of the square of the Brunt-Väisälä frequency, N^2. The data shown in the left hand and centre plots are used in compiling the total operator, $\hat{\mathcal{Z}}$, described in Section 4.2.3 of this thesis. The full-three dimensional integration of this flow went on to develop into a regular wave 3 state. The data plotted are calculated using the zonal mean temperature and zonal velocity, once they have bi-linearly interpolated onto the regular internal grid used for INSTAB . The zonal flow has a structure which quite closely resembles that of the internal jet flows studied in the previous section. In contrast, N^2 varies by a factor of 3 in the lower third of the fluid, whereas, in the internal jet flows it was assumed to be constant throughout.

The zonal mean state adjacent to the side walls was adjusted slightly for each of the runs described below. The procedure involved replacing the values at the edges of the grid nearest the side walls with a value extrapolated from the interior part of the flow. This had the effect of removing the large horizontal shear (in the Stewartson layers, Gill, 1982, for example) which was otherwise present in the mean state; this shear tended to increase the local vorticity such as to make the quasigeostrophic approximation (and therefore our investigations with INSTAB) invalid. All the results described in this section were calculated with $\Xi = 1.0 \times 10^{-5} s^{-1}$, to allow for the Ekman layers which develop in the numerical integration.

We used INSTAB to examine the stability of the mean state from Run 408 to perturbations of each zonal wavenumber in turn as described earlier in this chapter. The results are shown in Figure 5.19. The curve is clearly of a rather different character to those presented earlier in this chapter. Rather than having one maximum and a clearly defined SWC, the curve is somewhat spiky and shows that wavenumbers > 20 still have a significant growth rate. Although the equilibrated flow from the nonlinear integration of Run 408 had higher wavenumber components, these arose from nonlinear interactions,

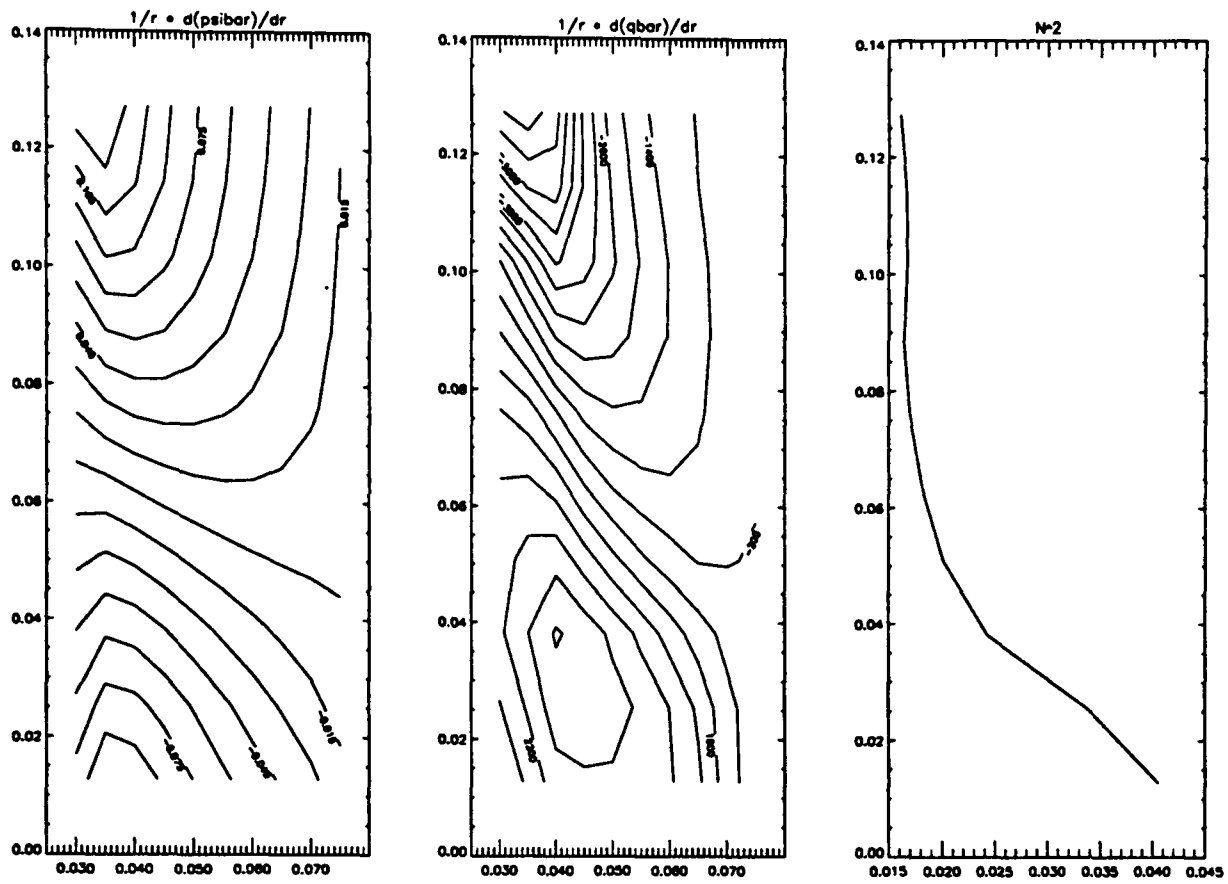


Figure 5.17: Zonal mean fields for Run 408. Left to right the plots show $\bar{u}/r$ (s^{-1}), $\bar{q}_r/r$ ($m^{-2}s^{-1}$) and N^2 (s^{-2}). The horizontal and vertical axes are radius and height, in metres, respectively.

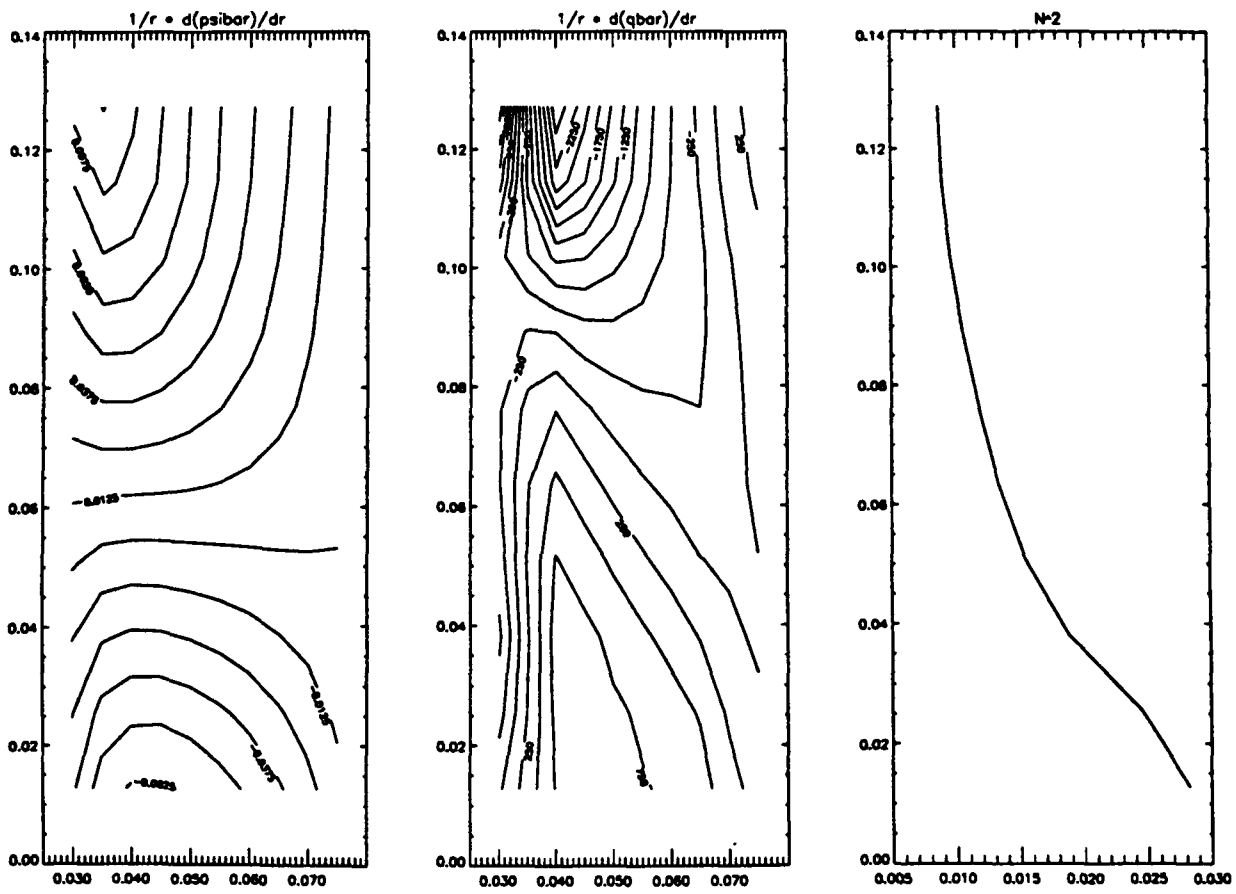


Figure 5.18: Zonal mean fields for Run 805. Details as for the previous figure.

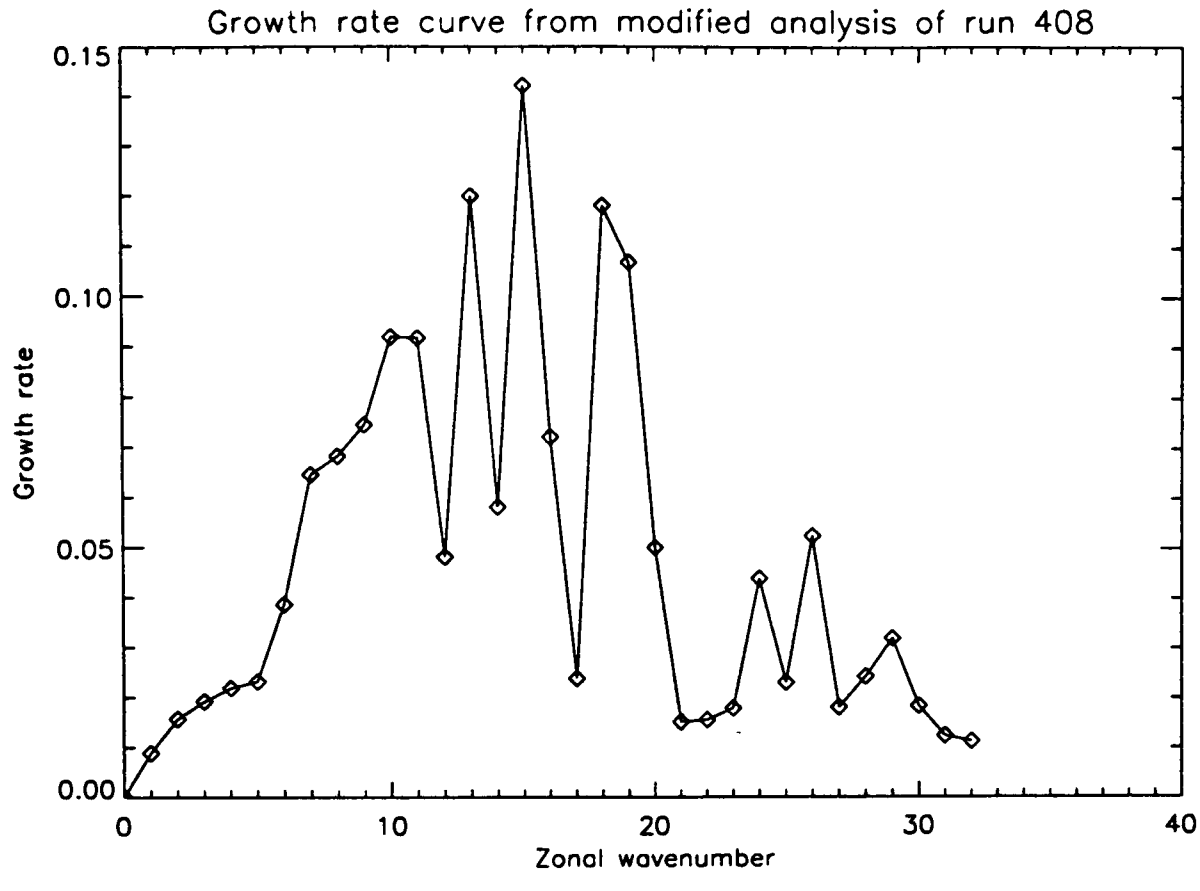


Figure 5.19: Growth rate curve from the stability analysis of the zonal mean state from Run 408. Wavenumber 15 has the largest growth rate.

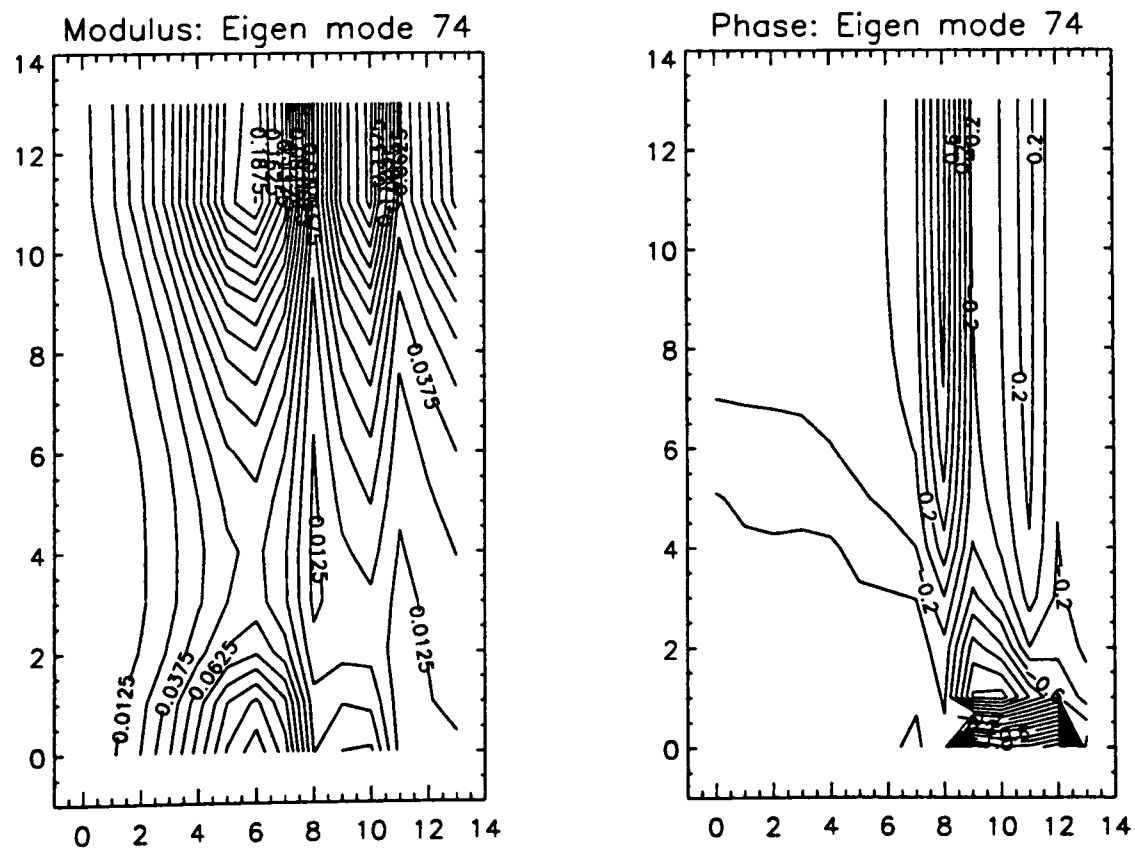


Figure 5.20: Results from analysis of Run 408 using INSTAB. In both of the diagrams the vertical and horizontal scales are vertical and radial grid point index respectively. Eigenmode 74 was the dominant eigenmode for wavenumber 15 perturbations to this flow.

and so their growth would not have been predicted by the linear analysis of INSTAB. The zonal mean flow used in this analysis was, however, highly supercritical (the vertical temperature gradient was small giving a small Burger number), and we would therefore expect to see a range of unstable modes associated with the initial growth, as shown in the figure.

Although wavenumber 3 was not shown as the most unstable mode in this analysis (as it was in the numerical integration), the results are only applicable in the linear limit of infinitesimal growth; the result we had was that wave number 15 perturbations would begin to grow fastest. When this perturbation had grown a finite amount, the linearity no longer held†, and our analysis was invalid.

The meridional structure and phase of this mode are shown in the left and right hand plots of Figure 5.20 respectively. The structure varies coherently throughout the domain; this is in agreement with experimentally observed steady waves (Jonas, 1980, James *et al.*, 1981, Hignett *et al.*, 1985). The structure of the wave is not unlike that of the second lateral Eady mode, calculated analytically (Eq. 5.28 and Figure 5.3) and numerically (Figure 5.7) for the annulus system, earlier in this chapter. The wave is not symmetric about mid-height, but nor was the zonal mean state on which it began to grow (see Figure 5.17).

The phase of the wave is almost uniform throughout the inner half of the annulus, but towards the outer wall the wave becomes more barotropic, especially away from the base. Near the base, in the lower right hand corner, the phase is poorly defined because of the small amplitude of the real and imaginary parts of the eigenmode.

5.4.2 Differentially heated, axisymmetric flow.

Figure 5.18 shows the mean state from Run 805. This was identical to Run 408 except that it was run at a rotation rate of 0.1 rad s^{-1}. This increased the Burger number sufficiently

†Since the perturbation terms are no longer "small" compared to the zonal mean state, it is not a valid approximation to ignore terms which are a product of perturbation terms.

to make the flow stable, so that even after the perturbation was added the flow remained axisymmetric in the numerical integration. Again the flow was similar to that used for the study of the internal jets; N^2, however, can be seen to vary monotonically with height.

We performed a stability analysis on the mean state from Run 805 using INSTAB. The resulting growth rate curve is shown in Figure 5.21. Rather than there being no growing modes as we might have expected, perturbations of wavenumbers 1–12 were all found to be unstable, but with growth rates 1/30 of those for the analysis of Run 408 (see Figure 5.19).

These disturbances are similar to the baroclinic weak waves observed by Hide and Mason (1978), which were found to occur at lower rotation rates (smaller Taylor numbers) than the cut-off for large amplitude wavelike flows. The weak waves typically had a zonal wavenumber of 1 or 2; these are precisely the waves which had the largest growth rates in our analysis.

The structure and phase of the numerically determined dominant mode are shown in Figure 5.22. The wave is clearly stronger nearer the bottom, and is centred slightly off mid-radius. Its vertical structure is similar to that of the Charney wave (see Charney (1947) or Gill (1982) for example). This is in good agreement with Jonas (1980), who observed experimentally that the weak waves tended to be bottom trapped. The phase diagram shows that the wave comprised a mixture of baroclinic and barotropic parts. The baroclinicity was strongest in the region where the radial gradient of the background quasigeostrophic potential vorticity changed sign. This change of sign is one of the necessary conditions for instability given by Charney and Stern (1962).

Hide and Mason (1978) noted that baroclinic weak waves in the annulus had an amplitude (measured in the temperature field) of about 10% or less that of the large scale waves, and that their phase speed (given by the real part of the eigenvalue) was much higher. Jonas (1980) measured the phase speeds of both sets of waves experimentally, and found the weak waves to have a typical phase speed $\sim$ 4 times faster than that of the large amplitude waves. By considering the real part of the eigenvalue for the growing waves, our analysis gave the phase speed of the fastest growing mode for Run 408 was

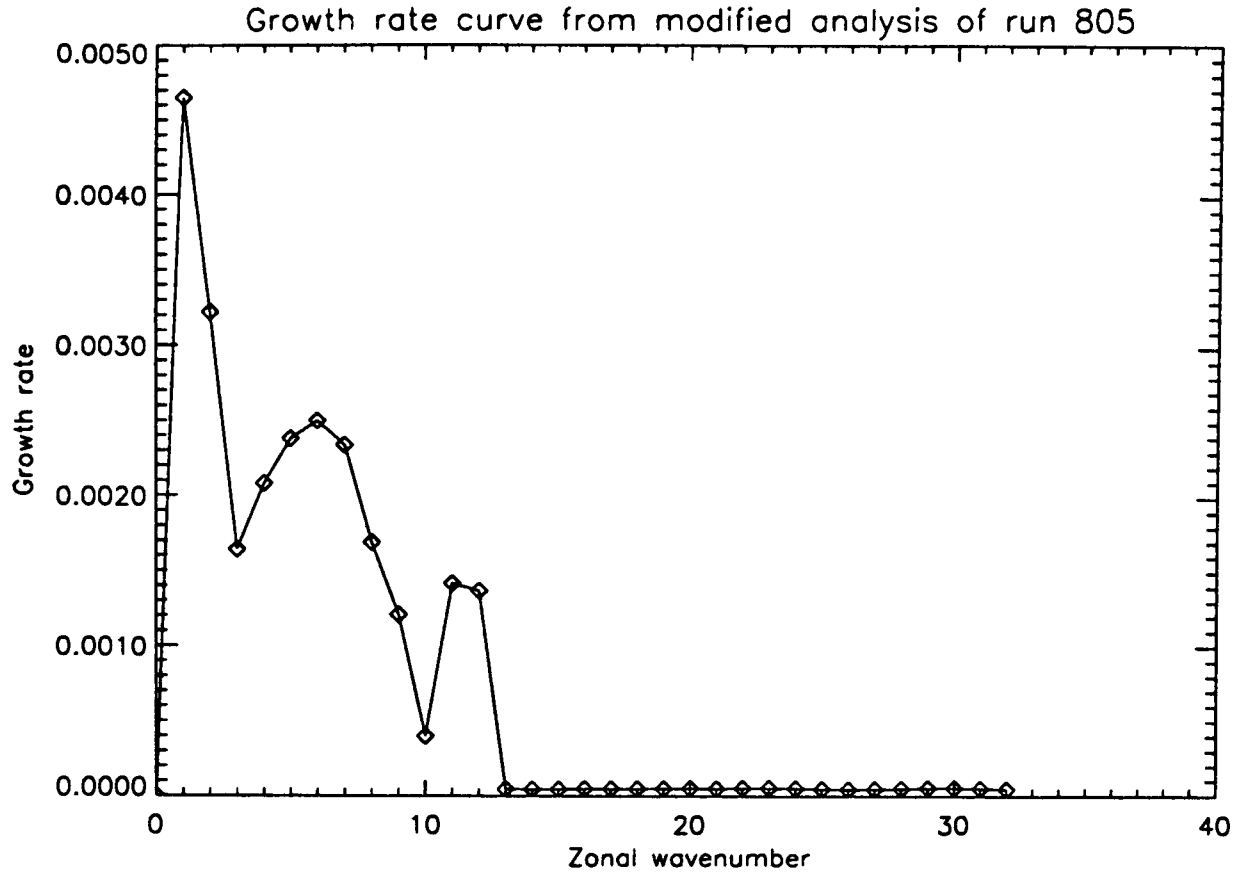


Figure 5.21: Growth rate curve from the stability analysis of the zonal mean state from Run 805. Wavenumber 1 has the largest growth rate.

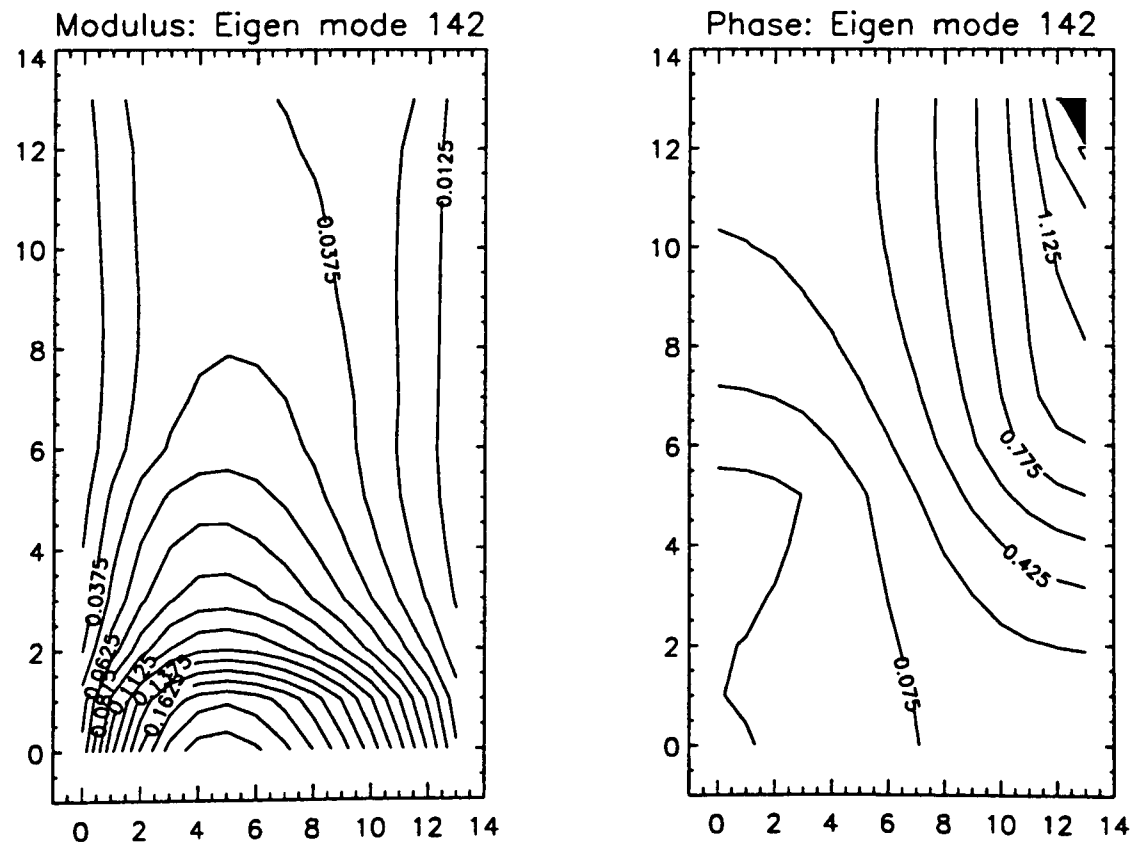


Figure 5.22: Results from analysis of Run 805 using INSTAB. In both of the diagrams the vertical and horizontal scales are vertical and radial grid point index respectively. Eigenmode 142 was the dominant eigenmode for wavenumber 1 perturbations to this flow.

0.012 rad s^{-1}, whilst that for the equivalent mode for Run 805 was 0.032 rad s^{-1}. These speeds are of the same order as those measured by Jonas (1980) for his experiments†, and the weak waves had the higher phase speed, as in Hide and Mason's observations.

The much lower growth rate, larger phase speed and location near the base of the annulus, all suggest that the instability of the mean flow state from Run 805 resulted in the development of baroclinic weak waves.

5.4.3 Internally heated, regular wave flow.

Figure 5.23 shows the mean state from Run 557. The velocity field is quite similar to the analytical sinusoidal internal jet profile

$$\bar{u} \propto -\cos(\pi\hat{z})\sin(\eta\pi[\hat{r}-0.5]), \tag{5.39}$$

used by Bell (1989) to model the zonal mean flow from internally heated experiments. N^2 is a minimum in the main body of the fluid, where the fluid is almost uniformly stratified, and is larger toward the end-walls. After a longer integration, Run 557 equilibrated to a steady wavenumber 4 flow of apparently isolated eddies. This flow was similar to that discussed by Read (1986b), although the model run in that case had a stress-free rigid upper boundary to simulate a system with a free upper surface.

We used INSTAB to examine the stability of the mean state from Run 557 to perturbations of each zonal wavenumber in turn, as we have done for the other runs considered in this chapter. The results are shown in Figure 5.25. It can be seen that, although the full integration went on to be dominated by a wavenumber 4 instability, our analysis showed the flow to be most unstable to wavenumber 24 disturbances. It must be stressed that INSTAB is a linear code, and that this is the probable reason for not predicting the eventual prevalence of a wavenumber 4 disturbance. Almost all the flow in the fully developed system was confined to four isolated eddies; our approximation, that the zonal flow should comprise of some mean state plus a *small* periodic perturbation, was obviously

†Jonas (1980) used an annulus system with different physical parameters, so we would not expect to repeat his results exactly.

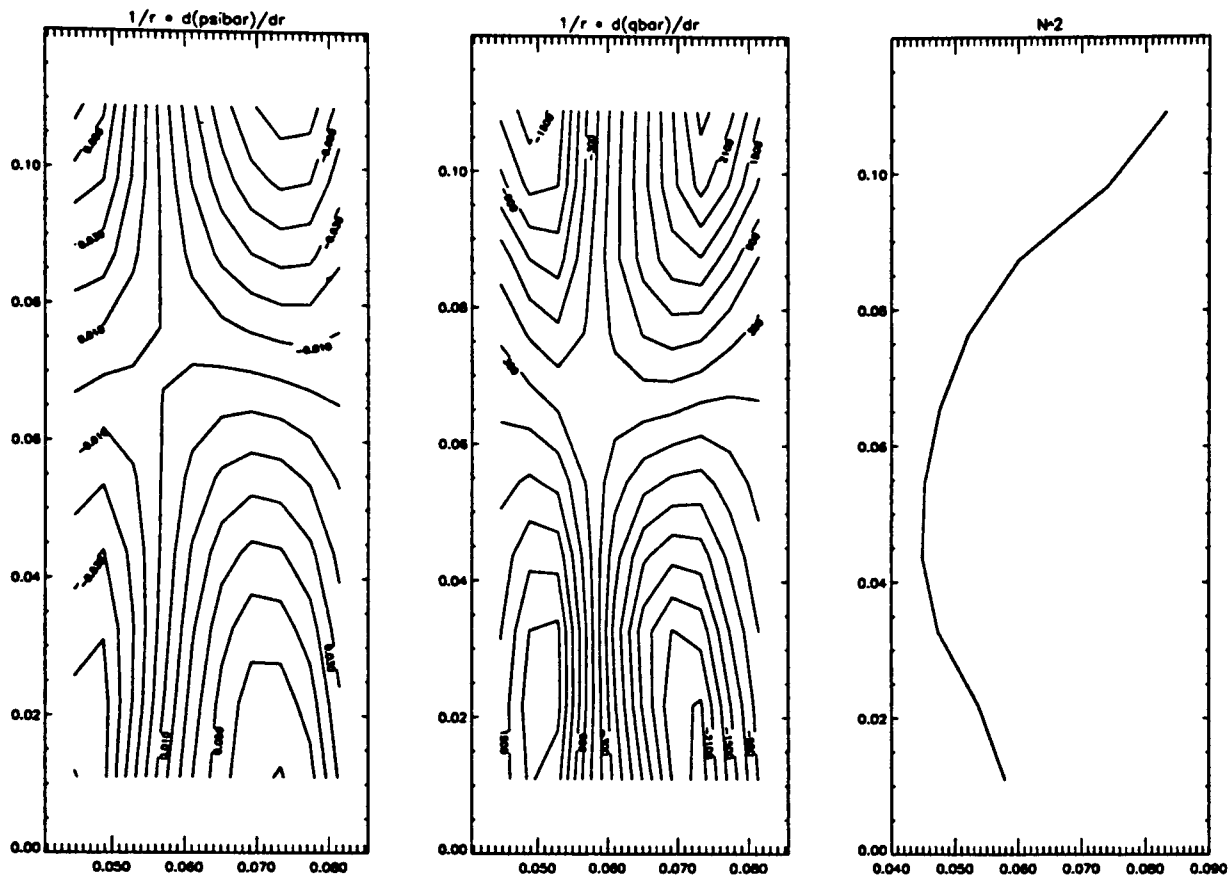


Figure 5.23: Zonal mean fields for Run 557. Left to right the plots show $\bar{u}/r$ (s^{-1}), $\bar{q}_r/r$ ($m^{-2}s^{-1}$) and N^2 (s^{-2}). The horizontal and vertical axes are radius and height, in metres, respectively.

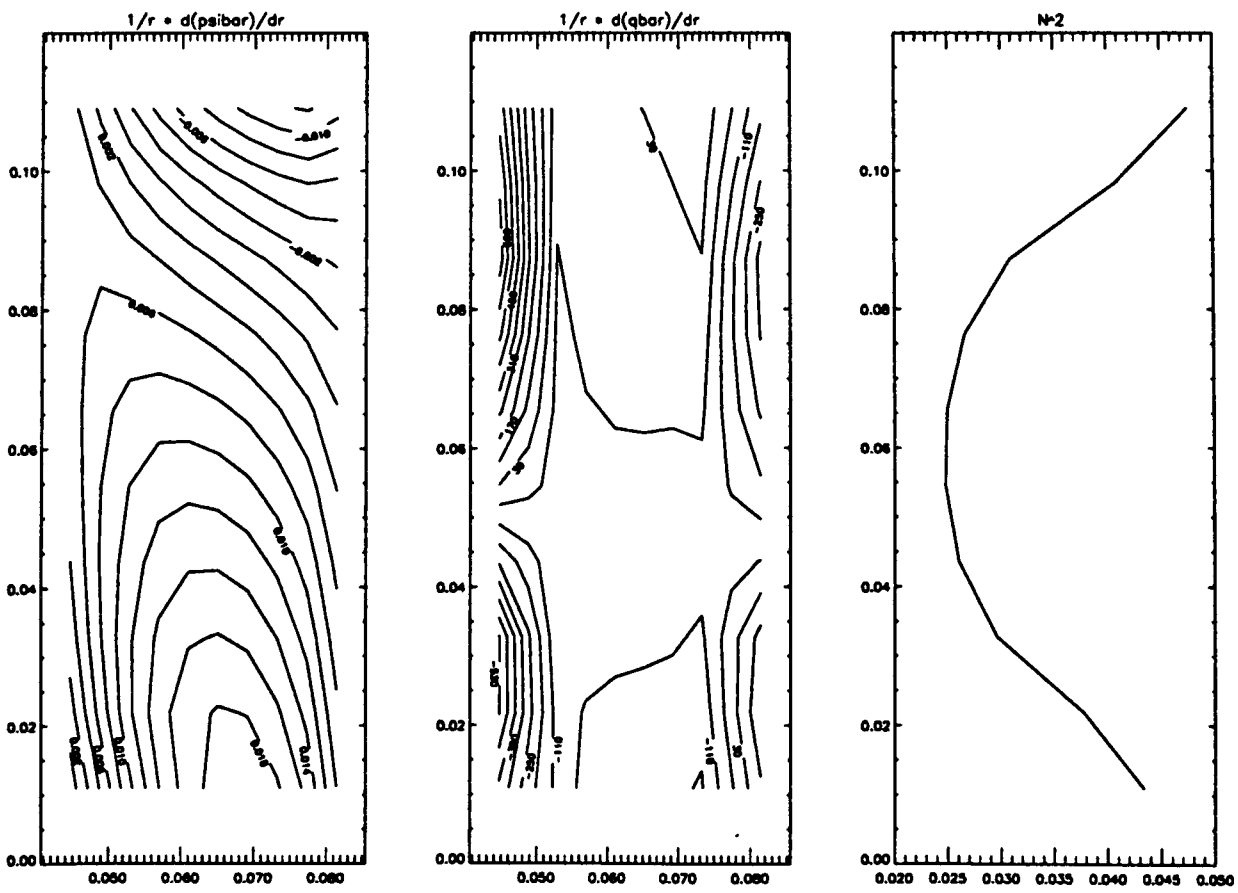


Figure 5.24: Zonal mean state from Run 755. Details as for previous figure.

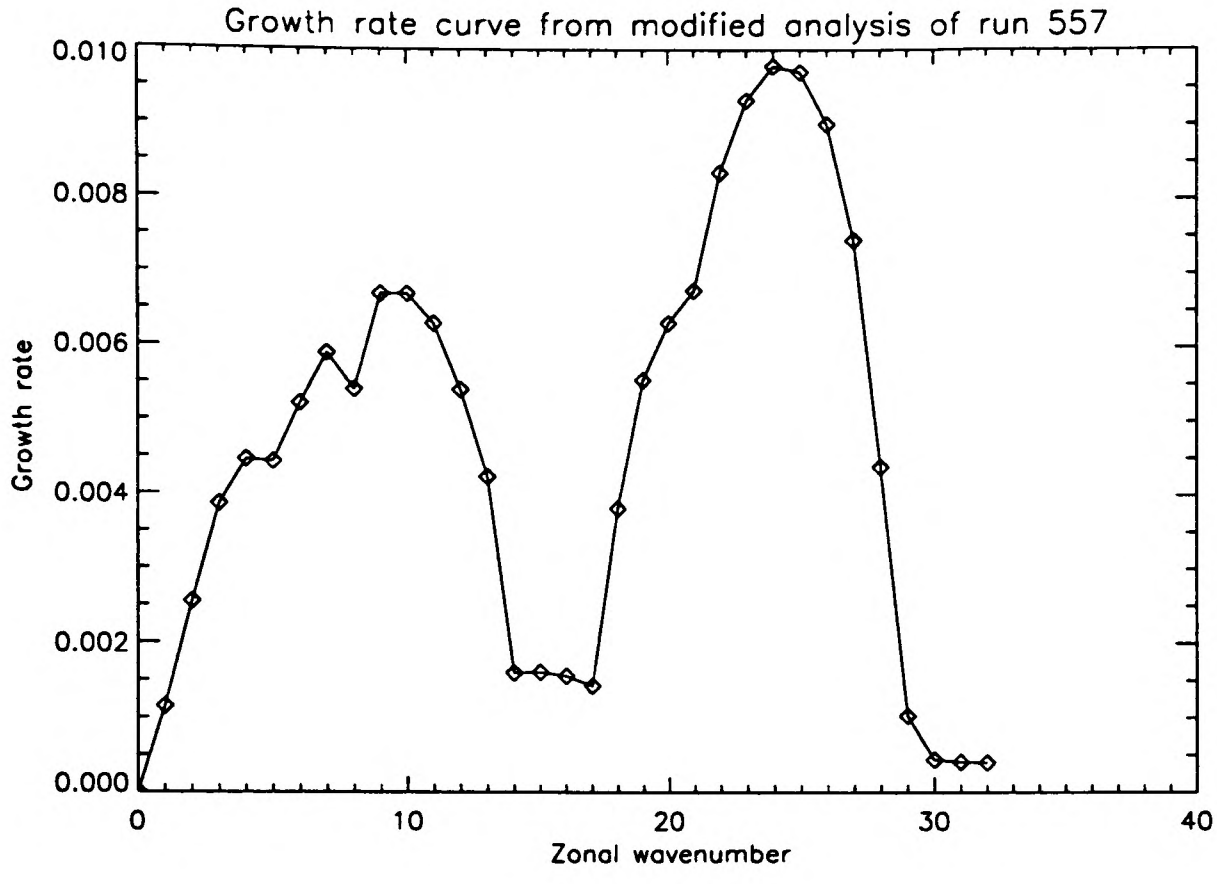


Figure 5.25: Growth rate curve from the stability analysis of the zonal mean state from Run 557. The figure shows that wavenumber 24 had the highest growth rate, although there was a second local maximum in the curve at wavenumber 10.

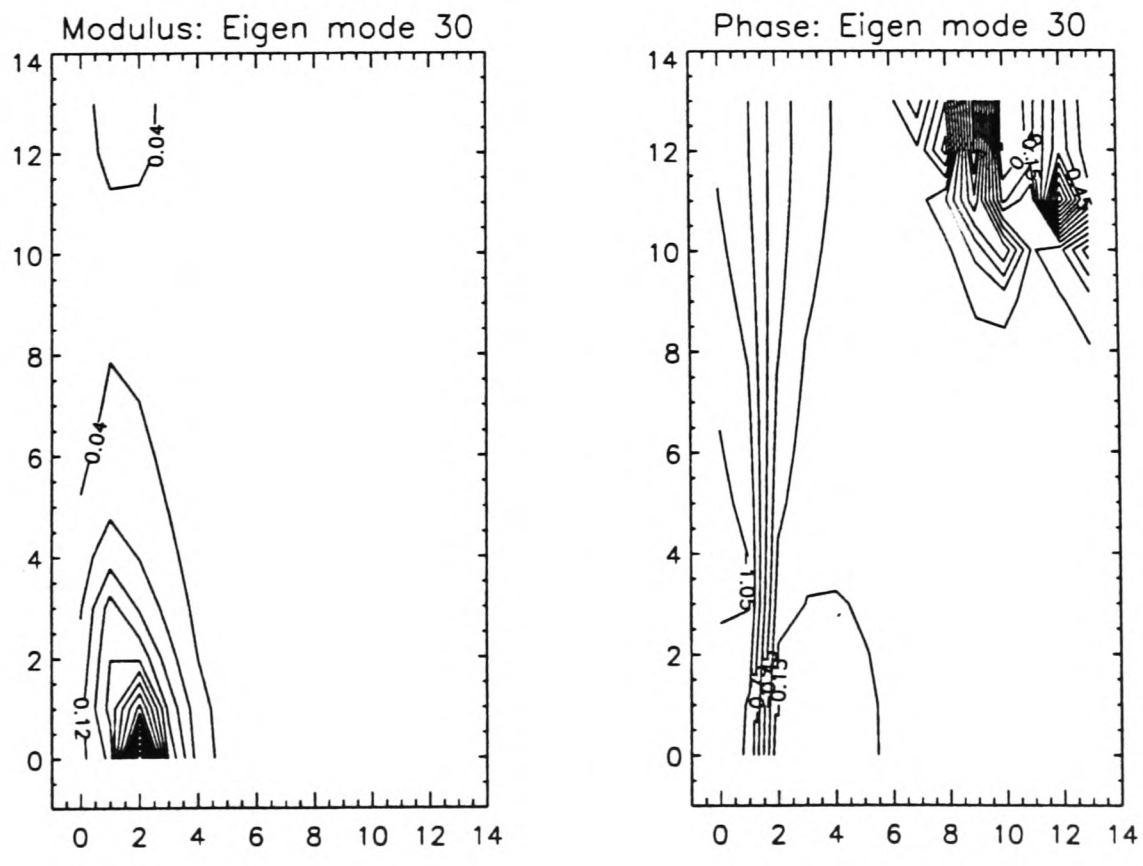


Figure 5.26: Results from analysis of Run 557 using INSTAB . In both of the diagrams the vertical and horizontal scales are vertical and radial grid point index respectively. Eigenmode 30 was the dominant eigenmode for wavenumber 24 perturbations to this flow.

far from valid in this case. Figure 5.25 does, however, show that in the linear limit, all wavenumber perturbations from 1–29 were unstable. Other zonal Fourier modes apart from the dominant mode 4 were present in the eddy diagnostics from the 3D numerical integration of Run 557; the first harmonic was also prominent. It is not then unreasonable, that our analysis of the mean state from Run 557 should show growth rates of the same order, for a range of zonal wavenumbers bordering the ultimately dominant wave.

It is interesting that there are two distinct peaks in the growth rate curve. The first part of the curve ($m < 6$) resembles the form of analytical solutions given earlier, but then begins to grow rapidly again to a local maximum at $m = 10$. This pattern is repeated between wavenumbers $m = 17$ and 21 before the global maximum is reached at $m = 24$. The structure and phase of the dominant mode at wavenumber 10 were found to be very similar to that for wavenumber 24 (see Figure 5.26).

The meridional structure and phase of the dominant mode for wavenumber 24 perturbations to the mean state from Run 557 are shown in Figure 5.26. The perturbation stream function was strongest towards the base of the annulus, and close to the inner wall. The phase plot suggests that the wave is predominantly barotropic in this region. Elsewhere, however, the phase is not well defined due to the absence of any strong structure of the wave. In his analysis of the eddy (total - zonal mean) terms in similar system, Read (1986b) found the dominant wavenumber 4 flow to be centred approximately about mid-radius. Although the dominant mode from our analysis is located rather nearer the inner wall, the vertical structure agrees reasonably well with Read. We would not expect to calculate an identical solution firstly because of the use of a different upper surface boundary condition, and secondly because our analysis is only valid in the linear limit.

5.4.4 Internally heated, axisymmetric flow.

Figure 5.24 shows the mean state from Run 755. This run was identical to Run 557 except that a much lower rotation rate (0.1 rad s^{-1}) was used. The flow shown in the figure is, again, typical of the "very weak rotation" state for axisymmetric, internally heated flows,

described by Read (1986a). The zonal flow was spun up by a meridionally asymmetric, radially overturning cell occupying most of the fluid region, and strongest by the outer wall. The interior flow was determined by a balance between advection and viscous forces[†] (Read, 1986c). The vertical profile of N^2 is typical of numerical simulations of internally heated flows with rigid insulating lid (Read, 1988b), in that it is almost symmetric about mid-height, where it is a minimum.

We used INSTAB to examine the stability of the mean state from Run 755 to perturbations of each zonal wavenumber in turn, as we have done above. The growth rate curve is shown in Figure 5.27. The figure shows that all wavenumbers analysed were determined to be unstable, the growth rates were ~ 0.0006 s^{-1}, or 10% of those of the growing modes from our analysis of Run 557. The fastest growing waves were $m = 2$ and $m = 3$, the growth rate then tending systematically toward zero at higher wavenumbers.

The form of these very weakly unstable modes is typified by that shown in Figure 5.28, for wavenumber 3 perturbations to the mean state. The structure of the wave indicates that it is strongly trapped to the upper boundary of the annulus. It occupies the full width of the system, and appears strongest slightly inward of mid-radius. Where the wave is strongest, the phase plot shows that it is a baroclinic disturbance. The phase speed of the fastest growing mode for Run 557 was 0.0033 rad s^{-1}, whilst that for the $m = 3$ mode here was 0.0061 rad s^{-1}. This increase in phase speed in moving from fully developed waves to the weak waves in an internally heated system is of the same order as that noted earlier for differentially heated flows.

It is interesting to note that although these waves are not observed in either the laboratory experiments, or full numerical integrations, they would seem to have a reasonably coherent structure (not unlike that calculated for regular waves in the internally heated annulus system, A1, described by Read, 1988b), and a growth rate of the same order as that calculated earlier for baroclinic weak waves in the wall heated system. We can

[†]In the Navier-Stokes equation for steady, axisymmetric flow, the three main terms are the Coriolis acceleration, advection and viscous drag. Since the rotation rate was so small for Run 755, the Coriolis force was not important, and the resulting flow was governed by a balance in the remaining terms.

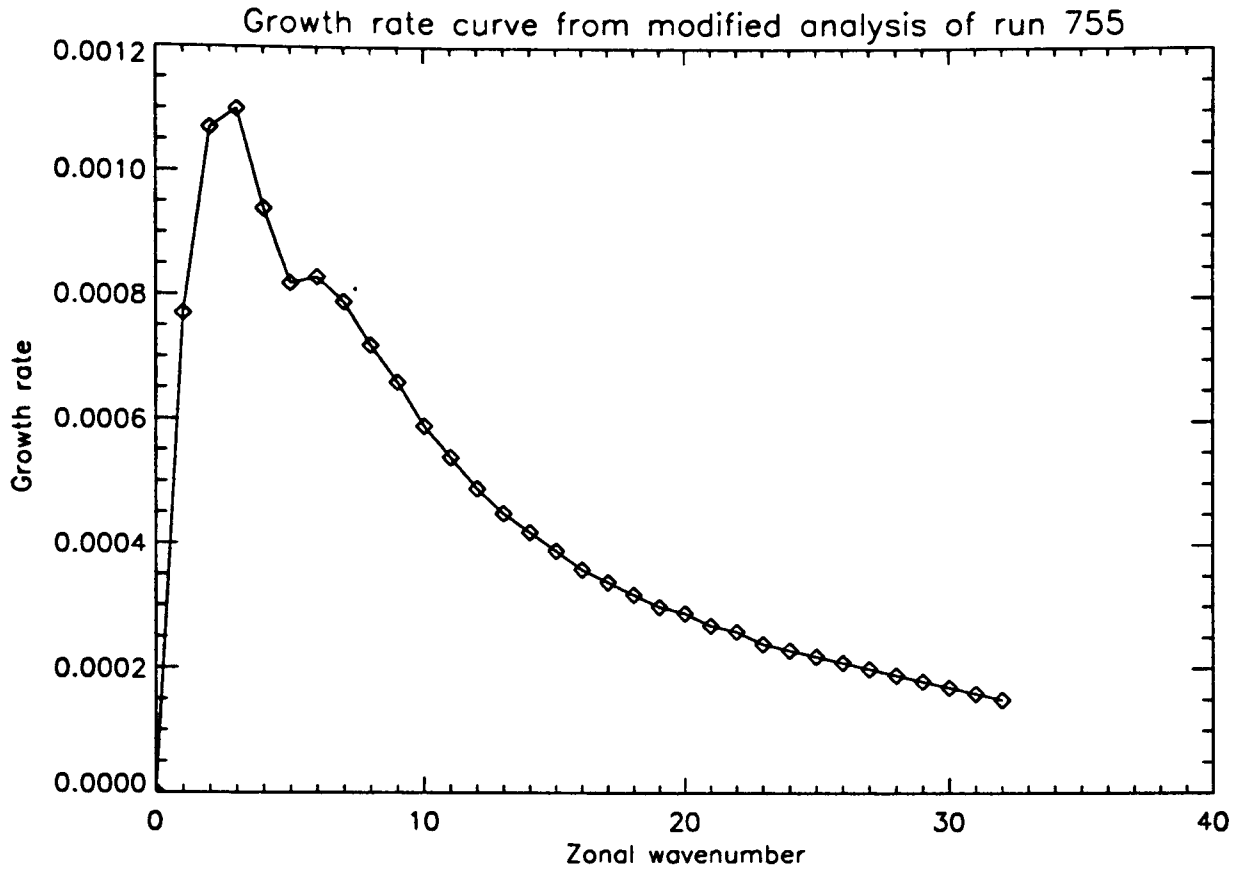


Figure 5.27: Growth rate curve from the stability analysis of the zonal mean state from Run 755. The figure shows that wavenumber 3 had the highest growth rate.

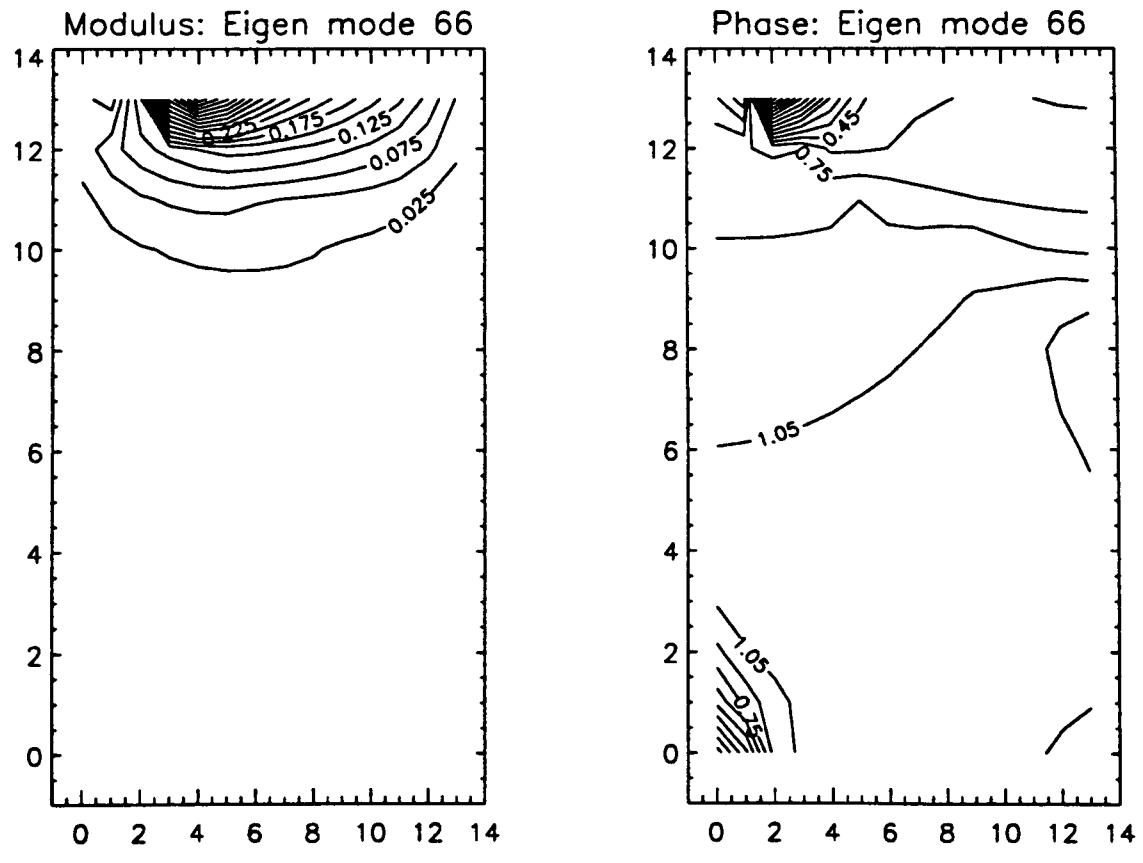


Figure 5.28: Results from analysis of Run 755 using INSTAB. In both of the diagrams the vertical and horizontal scales are vertical and radial grid point index respectively. Eigenmode 66 was the dominant eigenmode for wavenumber 3 perturbations to this flow.

conclude then that our analysis would indicate the existence of baroclinic weak waves with an amplitude $\sim 10\%$ that of a fully developed wave, in an internally heated annulus flow, which proved stable to larger scale disturbances. These waves grew fastest when the zonal wavenumber was 2 or 3.

5.5 Conclusions

We can conclude from the results presented in this chapter, that INSTAB (i) was coded correctly, and (ii) can provide a useful tool for analysing the linear stability of various zonal mean states to Fourier type perturbations in azimuth and time. We have investigated the effect of resolution and the inclusion of viscous Ekman layers on our results, and where analytical comparisons have been possible, these have matched the numerically derived results quite well. We used INSTAB to construct a graph of dominant zonal wavenumber as a function of Burger number, for a particular prescribed internal jet flow. This graph was then interpreted to determine the probable observed length scale of baroclinic disturbances for a laboratory scale experiment and for the mid-latitude terrestrial atmosphere. Further more, our analysis of the zonal mean state of Run 755 indicated the presence of hitherto unobserved baroclinic weak waves in an internally heated annulus system.

In Chapter 6 we discuss possible refinements to the code and describe some experiments which might be undertaken in the future.

Chapter 6

Conclusions.

6.1 A new annulus model.

In Chapters 2 and 3 of this thesis we have described the development of SARA , a new numerical model of the differentially heated, rotating fluid annulus system. The model was developed firstly in order to offer a more flexible code than was previously available, capable of simulating an extended range of experimental configurations, whilst accurately conserving mass and momentum. These considerations led us to adopt a control volume finite element method for the advective/diffusive part of the model; the finite elements offering adaptability to various topographies, and the control volumes ensuring conservation of the required fluxes. A second motivation was to investigate the applicability of some promising novel numerical techniques to modelling in geophysical fluid dynamics. The most important changes from the existing annulus model (Williams 1969, Hignett *et al.* 1985, White 1986) were the use of a control volume finite element method or CVFEM (Baliga and Patankar, 1983), the introduction of a continuous time stepping scheme, and the adoption of a "fast" matrix solver for the pressure equation. It must be stressed that although the pressure solver was the primary source of failure in SARA , this was a result of the formulation of the pressure determining part of the model, and in no way due to the "fast" matrix solver, which worked correctly.

6.1.1 Successful innovations.

A CVFEM was used in SARA to model the advective/diffusive processes. In essence, the CVFEM involved firstly dividing the calculation domain into a set of space-filling regions or finite elements. Bilinear shape functions were then fitted to each of the model variables, across each element in turn (see Section 2.3.2 of this thesis). Then, a control volume was associated with the vertex of each element (or node), in such a way that the control volumes also filled up the calculation domain (see Figure 2.2). This procedure is described in Section 2.4 of this thesis. The system equations were then cast into a flux form (Eqs. 3.16–3.20), and the divergence of the fluxes at each node calculated by applying the divergence theorem and integrating around the outside of the appropriate control volume. Chapter 2 describes the CVFEM further, and outlines some possible enhancements to the particular method currently used in SARA .

The code was validated carefully to ensure that this part of the scheme was working correctly. In Section 3.7.1 of this thesis we have shown that when the fluid was forced to remain at rest throughout the system, SARA reproduced the analytical solution for the diffusion of heat through the fluid very accurately. For this set of runs a stretched grid was used, and the agreement with the analytical solution indicated that not only was the diffusion being correctly reproduced by the CVFEM , but also that the grid stretching had been successfully incorporated into the model. In Section 3.7.2 we discussed the use of SARA to reproduce the radial overturning, characteristic of a differentially heated fluid annulus in the absence of rotation (Gill 1966, McIntyre 1968, Hignett *et al.* 1981 and Read 1986a). The convective boundary layers which developed were of the right length scale, and we were able to infer that the CVFEM parts of the model were functioning as planned.

The time stepping scheme, described in Section 3.5 of this thesis, avoids the need for periodic suppression of the numerical mode (Potter, 1973) by mixing together values from successive time steps continuously (priv. comm., S. D. Mobbs 1992). This scheme differs from normal continuous time stepping schemes in that the finite difference approximation to $\frac{\partial \psi}{\partial t}$, where ψ is some variable, is equal to the actual value of $\frac{\partial \psi}{\partial t}$ plus higher order terms.

A traditional continuous time stepping scheme would have some function of the mixing parameter, $f(\varepsilon) \simeq 1$, multiplying the first order derivative (see Eq. 3.40). If $\varepsilon = 0$, the time stepping scheme is exactly as for the Lax-Wendroff or the two-step leapfrog scheme (Potter, 1973). Setting $\varepsilon = -1/2$ gives the unconditionally unstable forward difference scheme, which we can use at the first time step, when no data from the previous time step are available to mix in. For a typical time step in SARA , $\varepsilon = -0.02$ and only a little smoothing takes place. This ensures that the scheme is quasi-second order accurate in the time step. The CFL condition applies and so the largest time step which can be used in SARA is ~ 0.5 s.

Various "fast" matrix solvers were tested for the pressure equation, and a comparison of their performance for various sizes of matrices is presented in Appendix A. We concluded that for relatively small matrices (*i.e.* less than 10×10), a direct solver was fastest, whilst for larger systems it would be faster to use an iterative solver. Extensive validation of the various matrix solvers has demonstrated their accuracy, and the matrix solver itself plays no part in the failure of SARA to calculate the pressure correctly. A further discussion of the matrix solvers is given in Section 6.4.

6.1.2 Possible modifications to SARA.

Despite the successful implementation of several of the constituent parts, SARA does not run correctly at the time of writing. The primary source of failure has been identified as the calculation of the pressure field, and in particular on the boundary conditions which are used (see Section 2.4.1 of this thesis). There are no independent boundary conditions on the pressure since, for a non-divergent fluid flow, the pressure is uniquely determined by the velocity field and the velocity boundary conditions. Solving a diagnostic Poisson equation for the pressure, as we do in SARA , however, necessitates the imposition of some kind of boundary condition on the pressure for the Laplacian to be properly defined. We have described (Section 2.4.1) the pressure correction method, which enables the pressure and velocity to be calculated simultaneously, so that only one set of boundary conditions is required. The incorporation of a modified form of pressure correction into SARA would

be the next step toward completing a fully operational numerical model. This would constitute a major alteration to the code, however the modular structure which has been adopted for the coding (see Section 2.2 of this thesis) would help somewhat. The scheme would need to be modified slightly, so that the time dependent nature of the flow can be modelled, rather than just the final steady state which pressure correction tends to iterate toward. It may be possible to achieve this modification by iterating the pressure correction scheme only two or three times at each time step, and then seeking a non-divergent velocity field using the edge-weighted cross-over relaxation scheme described in Chapter 3.

Although SARA is fully non-hydrostatic, by assuming that the dynamic pressure is negligible it would be possible to formulate a hydrostatic code. This alleviates the need for either a diagnostic Poisson equation for the pressure, or for a pressure correction scheme to be included, and also simplifies the vertical momentum equation. A hydrostatic code would not be subject to the difficulties described above in formulating the pressure boundary conditions, since the pressure would be, by hypothesis, hydrostatic throughout the fluid. The omission of any pressure equation would also result in a substantially faster code[†], although run time is not currently a major constraint, particularly while the code is axisymmetric, and the number of elements/nodes is small. A hydrostatic code would, of course, be less accurate than a fully operational, non-hydrostatic code. This is particularly relevant to the convective flow near the vertical side walls, which is often under-estimated by hydrostatic models (see, for example Miller and Gall, 1983a).

The edge-weighted cross-over relaxation scheme described in Section 3.6 is used to diagnose the radial velocity from the vertical velocity and the non-divergence condition. The scheme is similar to a simple shooting algorithm (see, for example, Burden and Faires 1985), and was designed to be simple and fast. The weighting of the forward and backward solutions insures that the calculation is most accurate near the boundaries, which is where the main part of the meridional flow is confined, by the boundary layers. The same algorithm can be used to calculate the meridional stream function, which can be

[†]A version of SARA with the pressure equation omitted required $\sim$ 60% of the CPU time per time step of the full code.

a useful diagnostic of the meridional flow. It would be a useful exercise to adopt a different algorithm for determining the radial velocity such as a relaxation method (see Press *et al.*, 1989). We could then compare the performance to that of the edge-weighted cross-over relaxation scheme presently used, to see what scope there would be for improvement.

In producing the results shown in Section 3.7.3, the meridional velocity field was slightly spatially smoothed at each time step. This smoothing was applied with the aim of moderating the small length scale variability in the velocity field, which grew to dominate the flow because of the errant pressure field. Once the main part of each time step had been completed, the radial and vertical velocity components were smoothed independently by using an artificial viscosity term,

$$u^n \mapsto u^n + \varpi \nu \nabla^2 u^{n-1} \tag{6.1}$$

$$w^n \mapsto w^n + \varpi \nu \nabla^2 w^{n-1} \tag{6.2}$$

where ϖ was a parameter used to adjust the degree of extra smoothing applied, and the superscripts refer to the time step index. The smoothing is lagged by one time step in order to avoid the numerical instability which would otherwise arise (see, for example, Potter, 1973). $\varpi = 0.2$ for all the results presented in Section 3.7.3. The effect of this smoothing was minimal, except in the last few time steps before the code crashed. It then had the desired effect of partially tempering the very rapid variation in the velocity, but was insufficient to prevent the model from crashing, as before. The artificial viscosity has been retained in SARA since it would have very little effect were the code working fully, and is computationally very inexpensive.

6.1.3 Future work with SARA.

Once the code has been made fully operational, there are several useful pieces of work which could be undertaken without extensive enhancements being made. The most obvious problem to investigate is that of an annulus system with sloping end walls. This type of system has been used to simulate dynamics on a β-plane, on which the Coriolis parameter is assumed to vary linearly with latitude about some reference value. The

sloping end walls can alter the susceptibility of the flow to the formation of baroclinic waves by constraining the fluid to move on surfaces parallel to the lid and base, as the rotation rate is increased (Mason, 1975). If these surfaces happen to lie within the "wedge of instability", described in Chapter 4, the change in topography can tend to promote baroclinic instability, whilst a different degree of slope can inhibit its development. The effects of sloping bases in annulus experiments are discussed further in Mason (1975) and Rayer (1992). Since the topography would still be axisymmetric, this problem would be ideally suited to analysis using SARA in its 2D form.

By modelling an internally heated system, and setting the inner radius to a suitably small value, it would be possible to simulate the "quasi-open" cylinder experiments, as described by Bastin (1994). In these experiments the inner cylinder was replaced by a wire, so that Ohmic dissipation could still be used to heat the fluid, whilst allowing almost uninterrupted flow throughout the domain. Although the baroclinic waves present in such are system would not be reproduced by an axisymmetric model, it would give a useful insight into the background state which would be present in the absence of zonal variation. The existing Met. Office model can be used to study this type of system and a comparison of the two would help to illuminate any shortcomings in SARA . It would also be instructive to run models of exactly the same system on SARA and the axisymmetric version of the current Met. Office model (Farnell and Plumb, 1975). A comparison of the two could provide useful information about the shortcomings of either one and, by "swapping-in" alternative schemes for the same part of SARA (for example, the diagnosis of the radial velocity), provide a control against which to judge the change. The model would also be ideally suited to generating background states for analysis using INSTAB ; this is discussed in Section 6.3.

Ultimately, the natural extension of SARA would be the development of a full 3D model. Its modular structure means that sections can be replaced individually whilst leaving the main structure of the program intact. The finite elements can be adapted to three dimensions by using tetrahedral elements and tri-linear basis functions. If a matrix solver were required for either a diagnostic Poisson equation for the pressure or a pressure correction algorithm, then the large number of grid points involved would require the use

of an iterative solver (see Section 6.4 and Appendix A of this thesis). Once a model of this type has been written and validated, one would be able investigate flow over more complex, three dimensional orography and make detailed comparisons with laboratory experiments.

In summary, SARA has the potential to become a very useful model once the remaining difficulties with the pressure have been dealt with. A control volume finite element method has been shown to be eminently suitable for this type of numerical modelling, and together with a pressure correction scheme, would prove a powerful combination for the more advanced numerical requirements of a full 3D code.

As a final *caveat* for those concerned with the numerical modelling of geophysical fluid dynamics, consider the words of Richardson (1922):

> Perhaps some day in the dim future it will be possible to advance the computations faster than the weather advances and at a cost less than the saving to mankind due to the information gained. *But that is a dream.*

6.2 Linear stability analysis.

In Chapter 4 we described the formulation of the linear instability problem, and the use of normal mode expansions to study the linear growth of zonal perturbations to an axisymmetric background state. In Section 4.3 we discussed the formulation of INSTAB, a numerical code which, given a zonal mean background state, formulates the instability problem in a complex matrix form, and then solves the eigensystem for the structure and growth rate of the normal modes. In Chapter 5 we presented some results obtained using INSTAB to study the linear stability of zonal mean states characteristic of the Eady problem (Eady, 1949), internal baroclinic jets (Bell and White, 1988 and Bell, 1989) and simulations from the Met. Office model.

The structure and phase of the dominant and second most dominant modes for the analysis of the Eady problem calculated using INSTAB agreed very well with the analytical solution. The problem was solved for a range of different resolutions, and there was a clear

tendency for the accuracy to increase as the resolution was increased (see Figure 5.8). We also demonstrated that the flow could be stabilised by increasing the Ekman damping at the top and bottom boundaries, and that this was again in close agreement with the analytical solution (Figure 5.10).

We investigated the effect of the variation of both lateral shear and Burger number on the linear stability of an internal baroclinic jet. Our results showed that the form of the dependence of the critical Burger number on the zonal wavenumber was shown to be very close to the analytical function (see Figure 5.16), further indicating the reliability of our numerical scheme.

Our analysis of the zonal mean states from the Met. Office model showed a range of zonal wavenumbers to be linearly unstable for both those runs which spun up to be dominated by a steady large scale wave, and those which spun up to axis-symmetric flow. The growth rates of the linearly unstable modes were an order of magnitude larger for those flows which developed large scale waves in the nonlinear integration. The structure of the dominant unstable modes were found to be consistent with existing observations.

6.2.1 Some key results from the stability analysis.

In Section 5.3.1, we showed how the zonal wavenumber of unstable perturbations (and in particular the dominant perturbation) to a zonal mean state of a particular analytical form varied with the Burger number of the flow. The flow used was an internal baroclinic jet (Bell, 1989), and by assuming that its structure was the same as that of the typical background state for the annulus and the terrestrial atmosphere, we were able to calculate the dominant zonal length scale of linear perturbations to these systems. The inferred results agreed well with observations for annulus and the atmospheric flows. This suggested that the structure for the mean flow which we had used was in some way representative of the real system, and that the nonlinear effects, which we had neglected, were not vital in determining the state of the system.

Run 755 was a simulation of an internally heated system with very weak rotation,

which equilibrated to give a purely axisymmetric flow in a nonlinear 3D integration. Our analysis of the mean state from Run 755 indicated that the flow was unstable to perturbations of all zonal wavenumbers from 1–32, although the growth rates of these unstable modes were very small ($\sim$ 0.001 s^{-1}, or 10% of that typical of larger scale waves). The form of these very weakly unstable modes was typified by that shown in Figure 5.28, for wavenumber 3 perturbations to the mean state (Figure 5.24). The waves appeared to have an organised structure which was approximately symmetric about mid-radius, and trapped toward the upper boundary. The low growth rate of these waves was indicative of a probable small wave amplitude in annulus flow. This, together with their coherent structure, would suggest that they are not mere artifacts of the numerics, but rather an indication of a phenomenon akin to the baroclinic weak waves discussed by Jonas (1980), but in an internally heated annulus system. Such waves have yet to be observed experimentally at the time of writing.

6.2.2 Future studies using INSTAB.

We have shown that INSTAB can be a valuable tool in the linear normal mode analysis of the stability of various background zonal mean states. There is no end to the range of zonal mean states which one might consider worthy of this type of study; two such examples are given below.

We have not yet used INSTAB to investigate the stability of the zonal mean state measured in a laboratory experiment. Although it would not be possible to measure enough data to be able to interpolate accurately onto the grid used by INSTAB , it would be interesting to see if the available data were sufficient to predict the zonal variation of the observed flow.

By choosing different prescribed structures for the background flow, such as those described in Section 5.3, it would be possible to construct other graphs for the variation of critical Burger number with zonal wavenumber, similar to Figure 5.16. If the observed variation of dominant zonal wavenumber with Burger number for a

particular system agrees with the form of such a graph, then we might infer that the meridional structure of the flow is related to that which was assumed in calculating the graph.

6.3 Combined analysis using SARA and INSTAB.

In Section 5.4 we successfully used INSTAB to analyse the stability of various zonal mean states, calculated from the output of the Met. Office annulus model. The only pre-processing required was to use bilinear interpolation to project the output field onto the regular grid used by the instability code. This would need to be done in order to import data from any source, such as laboratory measurements or from SARA. For greater accuracy a more complex interpolation technique, such as bi-cubic splines (see Press *et al.*, 1989 for example), might be used. The zonal mean fields which we used to study the internal baroclinic jets (Section 5.3) were calculated at the grid points used by INSTAB directly, so no interpolation was necessary.

Once a fully operational version of SARA has been completed, it will be a straightforward matter to use the output velocity and temperature fields as input for INSTAB, and then to perform a linear stability analysis on the flow. In this way we will be able to use the results from an axisymmetric model to investigate the growth rate and structure of the baroclinic waves which might develop in the annulus. This type of combined study would be very useful for the detailed exploration of state space, and mapping of transitions between flow types necessary for the calculation of regime diagrams (see, for example, Figure 1.3).

6.4 A comparison of matrix solvers.

In Appendix A of this thesis we compared a range of matrix solvers with a view to finding the most suitable one for solving the diagnostic Poisson equation in SARA for the pressure.

A matrix solver can be used to solve any set of linear equations, and so the development of a "fast" one has potential benefits for a variety of applications.

We compared one direct solver, which sought to calculate the exact solution in a finite number of operations, with a range of iterative solvers which successively refine an initial guess toward the actual solution. Each of the solvers were speeded up by using "sky-lining", a technique which takes advantage of the sparsity of the matrices associated with the finite difference representation of the Laplacian. In general, although they are adequate for small problems, direct solvers cannot be used for very large matrices because of numerical overflow. Iterative solvers are typically rather more complex and although they can be used for smaller matrices, a direct solver will often be rather quicker. The aim of our comparison was to find out how large a matrix need be before an iterative solver became preferable, and then which of the iterative solvers was fastest.

We showed that for the solution of a matrix of a size corresponding to the use of a 10×10 internal pressure grid, there was very little difference in performance between the various schemes, and therefore the simpler, direct solver could be used in SARA. For larger systems, the conjugate gradient solver, pre-conditioned using incomplete Choleski factorisation, was four times as fast as the direct solver and twice as fast as the standard conjugate gradient solver. If the matrix to be solved from SARA were to be any larger (as it would be if the model were extended into 3D, for example), then the pre-conditioned conjugate gradient solver would be the most appropriate to use.

The conjugate gradient solver pre-conditioned with incomplete Choleski factorisation has been used successfully for several other applications, both by the author and other numerical modellers (X. Li, 1993 and priv. comm., A. Weaver, 1992).

Appendix A

Matrix solvers for linear equations.

A.1 The pressure equation.

As we discussed in Chapter 3, the pressure perturbation is diagnosed by solving a Poisson equation,

$$\nabla^2\left(\frac{P}{\bar{\rho}}\right) = -\nabla_s \cdot \left[\mathbf{u}\cdot\nabla_s\mathbf{u} + 2\Omega\wedge\mathbf{u} - g\alpha(T-\bar{T})\right] \tag{A.1}$$

which can be written as

$$\nabla^2\left(\frac{P}{\bar{\rho}}\right) = -\nabla_s \cdot \mathbf{F} \tag{A.2}$$

where $\mathbf{F}$ is the vector formed by the terms on the right hand side of Eq. A.1. The divergence of $\mathbf{F}$ is calculated using the control volume finite element technique, CVFEM, on the coarser pressure grid which the mixed interpolation method requires. This need only be done for the internal nodes since the boundary condition specifies the pressure at the walls. We then have that

$$\nabla^2\Pi = b \tag{A.3}$$

where Π is the scaled pressure $P/\bar{\rho}$ and b is the divergence of $\mathbf{F}$. The two dimensional discretized distribution of a scalar such as b can readily be recast as a one dimensional vector with the whole numerical domain ordered in terms of one index. In a similar way (see for example Burden and Faires, 1985) the Laplacian operator can be represented as a matrix, which, if the sequencialization of the grid nodes is done correctly, is quindiagonal

for two dimensions. Eq. A.3 can then be written

$$\mathcal{A}\,\vec{\Pi} = \vec{b} \tag{A.4}$$

where $\mathcal{A}$ is the matrix representation of the Laplacian. It is worth noting that, although the matrix is very regular if the two dimensional grid has a regular spacing in both directions, it is rather more complex when a stretched grid is used, as in SARA . Although the matrix is more computationally intensive to set up, this only has to be done once since the grid is fixed in time. The matrix will then retain its sparsity, and positive definite quindiagonal structure.

A.2 General matrix representation.

The problem as shown in Eq. A.4 is a very general one, since any linear operator can be represented by a matrix and the function it acts on, and the result, can be represented by vectors. As a result, there has been a large effort in the past fifteen years or so to develop efficient and accurate numerical solvers for matrix problems. These can be divided into two kinds: direct and iterative solvers. Direct solvers set out to find an exact solution with a finite amount of calculation using techniques similar to those used to solve simple sets of simultaneous equations by hand. Iterative solvers take an initial guess at the correct solution and then make successive attempts to improve upon this, converging towards the correct solution by some means or other. Both of these techniques have distinct advantages and disadvantages, depending upon how they are implemented and in what situation.

Most matrices encountered in physical problems where differential equations have been discretized share the property that they are positive-definite, that is

$$\vec{x}\,\mathcal{A}\,\vec{x}^{\,t} > 0 \quad \forall \vec{x}. \tag{A.5}$$

If this condition holds, then the matrix can always be factorised into a lower and an upper triangular matrix with a common diagonal; this is known as Choleski factorisation (see Burden and Faires 1985, Horn and Johnson 1990, Laurie 1982, or Press *et al.* 1989). There

is no test which can be applied to a matrix to determine whether or not it is positive-definite before we attempt this factorisation. If the algorithm successfully computes the factors, however, then we know that the matrix must have been positive-definite *a posteriori*. The direct solver to be discussed takes advantage of this factorisation to make a simple and useful method for solving matrix equations of the type we are considering. The factorisation is accomplished using a standard algorithm (Burden and Faires, 1985 or Laurie, 1982) which is fast and reliable. This only has to be done once at the start of the program, when the matrix form of the Laplacian operator has been defined. The next step can be expressed algebraically:

$$\mathcal{A} = \mathcal{L}\,\mathcal{U} \tag{A.6}$$

where $\mathcal{L}$ and $\mathcal{U}$ are the lower and upper triangular factors of $\mathcal{A}$ respectively, and $\mathcal{U} = \mathcal{L}^t$. Then from Eq. A.4 we have

$$\mathcal{L}\,\mathcal{U}\,\vec{\Pi} = \vec{b} \tag{A.7}$$

which, defining $\mathcal{U}\,\vec{\Pi} = \vec{z}$, can be written as

$$\mathcal{L}\vec{z} = \vec{b} \tag{A.8}$$

Because $\mathcal{L}$ is triangular, forward substitution can be used to solve A.8 for $\vec{z}$ very simply, since $\vec{b}$ is known. Once we know $\vec{z}$, backward substitution can be used to solve

$$\mathcal{U}\,\vec{\Pi} = \vec{z} \tag{A.9}$$

for $\vec{\Pi}$, which can then be rescaled to give the pressure as required.

A.3 Maximal column pivoting.

This idea works in principle, but difficulties can arise if no "pivoting"† is used to reduce the effect of round-off error which can occur if ill-conditioned equations lead to zero or very small values on the leading diagonal of the matrix. This can be quite significant

†Pivoting is a method of swapping rows and/or columns of the matrix so that the diagonal elements are of the correct order.

when finite precision arithmetic is used, and any efficient algorithm must take this effect into consideration. Although it is difficult to use column interchange with the algorithms given in the literature, they can be readily altered to include a row-interchange technique such as the maximal column pivoting described in Algorithm 6.2 of Burden and Faires (1985).

A.4 Sky-lining.

An important feature of the quindiagonal matrix, $\mathcal{A}$, is that it is sparse. This is due to the use of a five point star to represent the Laplacian, which means only five elements can be non-zero on each row of the matrix. This sparse structure is carried over into the lower and upper triangular factors and, if account is carefully taken of this, then the forward and backward substitutions can be greatly accelerated. Analysis of the Choleski factorisation algorithm shows that a given element of $\mathcal{L}$ can only be non-zero if $\mathcal{A}$ contains a non-zero element in the same position, or in the same row to the left of it. This prompts the definition of a "left sky-line" of $\mathcal{A}$ as the set of lowest column indices of non-zero elements in each row,

$$\vec{LS}(i) = \{\mathcal{A}[i,j] \neq 0\ ,\ \mathcal{A}[i,k] = 0 \ \text{ if } k < j\ \} \tag{A.10}$$

where $\vec{LS}(i)$ is the i^{th} element of the left sky-line and $\mathcal{A}[l,m]$ is the element of $\mathcal{A}$ which is i rows down and j columns across. In a similar way we can define a right sky-line as

$$\vec{RS}(i) = \{\mathcal{A}[i,l] \neq 0\ ,\ \mathcal{A}[i,m] = 0 \ \text{ if } m > l\ \} \tag{A.11}$$

Zeros to the left of the left sky-line are preserved in the lower triangular matrix and those to the right of the right sky-line are preserved in the upper triangular matrix.

When multiplications involving $\mathcal{A}$, $\mathcal{L}$ or $\mathcal{U}$ and a vector are performed, it is only necessary to work from the left sky-line to the right sky-line. In the case of a typical multiplication involving $\mathcal{A}$ this will mean $2n+1$ terms instead of n^2 for each row of the matrix, where n is the number of grid points in each directions. For an internal pressure grid of 7 by 7 nodes such as the one used in SARA this represents a saving of $\sim 70\%$.

Using all these techniques (Choleski factorisation, sky-lining, maximal column pivoting and forward and backward substitution) gives a robust and moderately fast solver for solution of linear equation sets. This is the matrix solver currently used in SARA and is as fast as the best iterative solver that has been tested for this problem. This result is due to the size of the matrix involved, which has only 49 by 49 elements. Iterative solvers are designed for systems with much larger matrices, where they are indeed much quicker and where a direct solver often will not work. When the model is extended into three dimensions with 64 zonal grid points, the matrix will be 3136 by 3136 elements with seven non-diagonal elements per row. In this case sky-lining will obviously be far more important, since the relative sparseness will be even greater ($\sim$ 97% of each row will lie outside the sky-lines), and the storage requirements more demanding. Iterative solvers are used *de rigeur* for matrices of this size and for completeness a brief discussion of these is given.

A.5 Iterative solvers.

Suppose a matrix $\mathcal{B}$ is difficult to invert, which will generally be the case for matrices such as $\mathcal{A}$, the Laplacian operator. It is possible to find a matrix $\mathcal{C}$ which is easy to invert and is in some sense close to $\mathcal{B}$. Then we can write a general iterative procedure

$$\mathcal{B}\,\vec{x} = \vec{d} \tag{A.12}$$

$$(\mathcal{C}+\mathcal{B}-\mathcal{C})\,\vec{x} = \vec{d} \tag{A.13}$$

$$\mathcal{C}\,\vec{x} = \vec{d}-(\mathcal{B}-\mathcal{C})\,\vec{x} \tag{A.14}$$

$$\mathcal{C}\,\vec{x}_{k+1} = \vec{d}-(\mathcal{B}-\mathcal{C})\,\vec{x}_k \tag{A.15}$$

$$\mathcal{C}(\vec{x}_{k+1}-\vec{x}_k) = \vec{d}-\mathcal{B}\,\vec{x}_k \tag{A.16}$$

where $\vec{x}$ is the answer (the scaled pressure in the case of SARA) and the subscripts denote the iteration number. This general procedure only involves forming matrix-vector products (where full account can be taken of sparseness) and the inversion of a matrix which is *a priori* easy to invert. Three possible choices of $\mathcal{C}$ are described below:

- $\mathcal{C} = \mathcal{D}$ = diag $\mathcal{B}$. When this choice is applied to the procedure given above, the result is the Jacobi iterative method. It can be shown that this will converge if $\mathcal{B}$ is diagonally dominant, that is if

$$|\mathcal{B}[i,i]| > \sum_{j \neq i} |\mathcal{B}[i,j]| \qquad i = 1, 2, \ldots, n \tag{A.17}$$

  This condition will be satisfied for the Laplacian matrix, $\mathcal{A}$, used in SARA .

- $\mathcal{C} = \mathcal{L} + \mathcal{D}$ where $\mathcal{B} = \mathcal{L} + \mathcal{D} + \mathcal{L}^t$. Here $\mathcal{L}$ is the lower triangle of $\mathcal{B}$ with zeros on the leading diagonal. This is called the Gauss-Seidel method.

- The Gauss-Seidel method spoils the symmetry if the matrix has any, so it is possible to alternate between $\mathcal{C} = \mathcal{L} + \mathcal{D}$ and $\mathcal{C} = \mathcal{L}^t + \mathcal{D}$ on successive iterations. This is equivalent to $\mathcal{C} = (\mathcal{L} + \mathcal{D})\mathcal{D}^{-1}(\mathcal{L}^t + \mathcal{D})$ at a single iteration

These methods will all converge to the solution, but not very quickly. This is because at each iteration, only one of the unknowns is updated using the current approximations to the values of the others. The methods can be speeded up by using over-relaxation, where the increment to the unknown being calculated is multiplied by a factor $\omega > 1$ before being added. When this is done to the three methods above they are known as simultaneous, successive and symmetric over-relaxation respectively. The convergence is a sensitive function of ω and various values need to be tried until an optimum is found (it is necessary to choose $1 < \omega < 2$ or there will be no convergence).

A.6 Incomplete factorisation.

The iterative procedures described so far all rely on splitting the original matrix to get a version which is easy to invert. There is a second option, which is to use incomplete (or reduced) factorisation (Laurie, 1982). If we use Choleski factorisation to say $\mathcal{B} = \mathcal{L}\,\mathcal{L}^t$ as before, then force the same sparsity pattern on $\mathcal{L}$ as the lower triangle of $\mathcal{B}$ by setting the necessary elements to zero, and then force $\mathcal{L}^t$ in the same way, we can construct $\mathcal{C}$

$$\mathcal{C} = \tilde{\mathcal{L}}\,\tilde{\mathcal{L}}^t \tag{A.18}$$

where tilde denotes incomplete factors with forced sparsity structure. This procedure allows full account to be taken of the fill-in zeros between the sky-lines and the other non-zero elements, meaning that as the grid size used in the Poisson solver gets larger, the incomplete factorisation becomes more and more efficient†.

A.7 Conjugate gradient solvers.

If we assume that we have an initial guess, $\vec{\Pi}_0$ to the solution of Eq. A.4 and an initial search direction, $\vec{p}_0$ (a vector in the space $\mathbb{R}^{n^2}$ of $\vec{\Pi}$ and $\vec{b}$ which points from the current guess towards the solution), then we can describe the conjugate gradient (Horn and Johnson 1990, for example) method as follows:

1. Calculate $\vec{\Pi}_{k+1} = \vec{\Pi}_k + \alpha\vec{p}_k$ with α fixed so that $\vec{r}_{k+1} = \vec{b} - \mathcal{A}\,\vec{\Pi}_{k+1}$ is orthogonal to the current search direction, $\vec{p}_k$.

2. A new search direction, $\vec{p}_{k+1}$, is chosen by subtracting the necessary component of $\vec{p}_k$ from $\vec{r}_{k+1}$ so that $\vec{p}_{k+1}$ is conjugate to $\vec{p}_k$ (ie. so that $\vec{p}_{k+1}$ is orthogonal to $\mathcal{A}\,\vec{p}_k$). Then return to the previous step.

This procedure is in one sense iterative since as it is repeated the solution converges, but it can also be considered a direct method since, in theory after n^2 iterations, the error will be orthogonal to the whole space, and so must effectively be zero. This would mean that a finite number of iterations would give the exact solution. In practice, however, round-off errors become prevalent and and the residual, $\vec{r}$, is never precisely zero.

†For the 7 by 7 internal pressure grid currently used, it was found that the increase in efficiency was countered by the extra numerical procedure and so there was no net gain in speed. The effect of increased number of grid points is discussed in the last section of this appendix.

A.8 Preconditioned conjugate gradient solvers

The over-relaxation methods described earlier arrive at a search direction by mapping the residual from the last iteration from "answer" space via the inverse close matrix back to "guess" space. This then gives the direction, the step size being determined by ω. This idea can be used in the conjugate gradient method to speed up convergence by using better search directions. Step 2 above is replaced by

$$\vec{p}_{k+1} = C^{-1}\,\vec{r}_{k+1} \tag{A.19}$$

but the calculation of the step size is slightly more complicated in this case (see Laurie, 1982). The two most common choices of pre-conditioning for C are symmetric over-relaxation, or incomplete Choleski factorisation, and these have been found to run from 5% to 50% faster than a standard conjugate gradient solver, depending upon the complexity of the particular problem on which they are tested.

A.9 A comparison of techniques.

In order to decide which of the techniques described in this section should be used for different kinds of problems, detailed intercomparisons were made. All the solvers were written in the same style, so that they were easy to follow, though not necessarily optimally coded. The test problem used for each algorithm was to calculate the steady state temperature distribution on a square plate with one edge at 200 degrees and the others at 100 degrees. The solution is illustrated in Figure A.1. The solver was deemed to have finished when the fractional change in the l_2 norm of the solution vector was $< 0.5\%$. Four solvers were coded: a direct solver using Choleski factorisation and maximal column pivoting, and three iterative solvers using successive over-relaxation, the conjugate gradient method and the conjugate gradient method pre-conditioned using incomplete Choleski factorisation, respectively. All these solvers were speeded up by using sky-lining. This final scheme is one of the innovative techniques originally suggested for use in SARA (see Chapter 1).

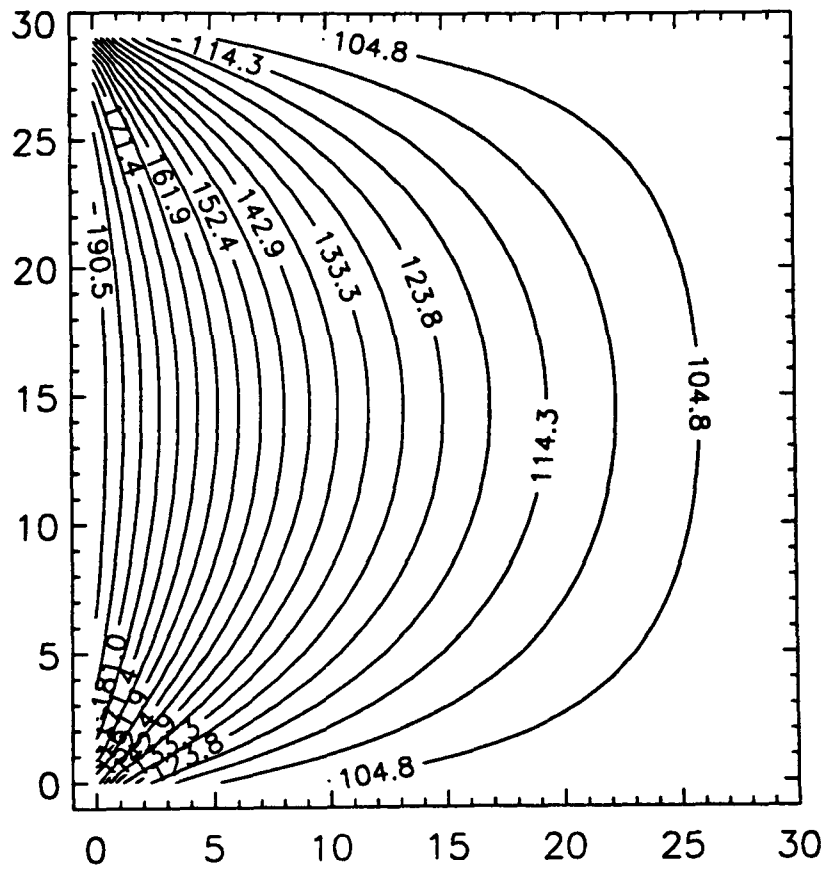


Figure A.1: The steady state temperature distribution in a square plate with one side at 200 degrees and the other sides held at 100 degrees. This is the solution to the test problem used for comparing the various matrix solvers.

Table A.1: Run time in CPU seconds for various matrix solvers at three different grid resolutions, for the heated square plate test problem. The code was run on a DEC VAXstation 3100-M76.

Matrix solver used	Square grid resolution		
	10×10	20×20	30×30
Direct factorisation with maximal column pivoting	0.5	6.8	45.2
Successive over-relaxation $\omega = 1.01$, 50 iterations	1.1	19.0	194.2
Conjugate gradient (solution converged to $< 0.5\%$)	0.5	3.9	20.0
Pre-conditioned conjugate gradient (converged as above)	0.5	5.9	37.3
As above excluding initial factorisation	(0.4)	(3.7)	(10.7)

The results are shown in Table A.1. When considering these statistics it is important to consider that: (i) the direct solver is the only one which gives the exact solution (to within rounding errors), (ii) the rate of convergence of the successive over-relaxation solver is dependent on the choice of the relaxation parameter, ω, so there may be some further slight improvement possible, and (iii) the incomplete factors only need to be calculated once at the start of the model, and so the bracketed times are the most important for the pre-conditioned conjugate gradient solver. For problems where the matrix is small, typified here by the 10×10 grid, there is not much difference between the various solvers. In this case it is preferable to use the direct method because it gives the exact solution. As the matrix becomes larger, the iterative schemes become quicker than the direct solver. A notable exception, however, is the successive over-relaxation solver, which was the slowest for all three test cases. For the largest matrix, the pre-conditioned conjugate gradient solver was four times as fast as the direct solver and twice as fast as the standard conjugate gradient solver. For even larger matrices, this relative improvement would become greater still.

In light of these results, the current axisymmetric model uses a direct solver with Choleski factorisation, sky-lining, and forward and backward substitution for the 7×7 grid of internal pressure nodes. If the matrix should become any larger (because of either increased resolution, or extending the model into three dimensions), then the pre-conditioned conjugate gradient solver would be the most appropriate. In this case, the modular structure of SARA would make it very simple to make this modification.

Appendix B

Bibliography

Andrews D. G., Holton J. R. and Leovy C. B. (1987) "Middle atmosphere dynamics." Academic Press.

Antar B. N. and Fowlis W. W. (1980) "Eigenvalues of a baroclinic stability problem with Ekman damping." *J. Atmos. Sci.*, **37**, 1399–1404.

Baliga B. R. and Patankar S. V. (1980) "A new finite element formulation for convection-diffusion problems." *Numerical Heat Transfer*, **3**, 393–409.

Baliga B. R. and Patankar S. V. (1983) "A control volume finite-element method for two-dimensional fluid flow and heat transfer." *Numerical Heat Transfer*, **6**, 245–261.

Baliga B. R., Phan T. T. and Patankar S. V. (1983) "Solutions of some two-dimensional incompressible fluid flow and heat transfer problems, using a control volume finite-element method." *Numerical Heat Transfer*, **6**, 263–282.

Bastin M. J. (1994) "Baroclinic waves and geostrophic turbulence in a convection chamber with sloping end walls." D. Phil. Thesis, S-DAOPP, Oxford.

Bell M. J. (1986) "Documentation of the poisson solver used by the Met. O. 21 2D Navier-Stokes numerical model." Meteorological Office Internal Report, Met. O. 21 IR86/2.

Bell M. J. (1989) "Theoretical investigations prompted by experiments with baroclinic fluids." Ph. D. Thesis, Imperial College, London.

Bell M. J. and White A. A. (1988) "The stability of internal baroclinic jets: some analytical results." *J. Atmos. Sci.*, **45**, 2571–2590.

Bjerknes V. (1937) "Application of line integral theorems to hydrodynamics of terrestrial and cosmic vortices." *Astrophys. Norv.*, **2**, 263–339.

Burden R. L. and Faires J. D. (1985) "Numerical analysis." Prindle, Weber and Schmidt.

Burns A. D. and Wilkes N. S. (1987) "A finite difference method for the computation of fluid flows in complex three dimensional geometries." AERE–R 12342, Harwell Laboratory.

Charney J. G. (1947) "The dynamics of long waves in a baroclinic westerly current." *J. Meteor.*, **4**, 135–163.

Charney J. G. and Drazin P. G. (1961) "Propagation of planetary scale disturbances from the lower into the upper atmosphere." *J. Geo. Res.*, **66**, No. 1.

Charney J. G. and Stern M. E. (1962) "On the stability of internal baroclinic jets in a rotating atmosphere." *J. Atmos. Sci.*, **19**, No. 2, 159–172.

Crank J. (1956) "The mathematics of diffusion" Oxford University Press

Davies T. V. (1956) "The forced flow due to heating of a rotating liquid." *Phil. Trans. Roy. Soc. London*, (A), **249**, 27–64.

Drazin P. G. (1978) "Variations on a theme of Eady." Rotating fluids in geophysics, Ed. P. H. Roberts and A. M. Soward, Academic Press.

Eady E. T. (1949) "Long waves and cyclone waves." *Tellus*, **1**, 33–52.

Ertel H. (1942) "Ein Neuer hydrodynamischer Wirbelsatz." *Met. Z.*, **59**, 271-281.

Farnell L. and Plumb R. A. (1975) "Numerical integration of flow in a rotating annulus I: axisymmetric model." Occasional Note, Met O 21 75/3.

Farnell L. and Plumb R. A. (1976) "Numerical integration of flow in a rotating annulus II: three dimensional model." Occasional Note, Met O 21 76/1.

Farnell L. (1980) "Solution of poisson equations on a nonuniform grid." *J. Comp. Phys.*, **35**, No. 3, 408–425.

Fowlis W. W. (1964) "An experimental study of the transitions between the flow regimes of thermal convection in a rotating annulus of liquid." Ph.D. Thesis, Massachusets Institute of Technology, USA.

Fowlis W. W. and Hide R. (1965) "Thermal convection in a rotating annulus of liquid: effect of viscosity on the transition between axisymmetric and non-axisymmetric flow regimes." *J. Atoms. Sci.*, **22**, 541–558.

Fultz D. (1951) "Experimental analogies to atmospheric motions." Compendium of meteorology, Am. Met. Soc., 1235–1248.

Gill A. E. (1966) "The boundary-layer regime for convection in a rectangular cavity." *J. Fluid Mech.*, **26**, part 3, 515–536.

Gill A. E. (1982) "Atmosphere–Ocean dynamics." Academic Press.

Greenspan H. P. (1968) "The theory of rotating fluids." Cambridge University Press.

Hide R. (1953) "Some experiments on thermal convection in a rotating fluid." Ph.D. Thesis, Cambridge University.

Hide R. (1958) "An experimental study of thermal convection in a rotating fluid." *Phil. Trans. Roy. Soc. London*, (A), **250**, 442–478.

Hide R. (1969) "Some laboratory experiments on free thermal convection in a rotating fluid subject to a horizontal temperature gradient and their relation to the theory of global atmospheric circulation." The global circulation of the atmosphere, Ed. G .A .Corby, London, Roy. Met. Soc.

Hide R. (1977) "Experiments with rotating fluids." *Quart. J. R. Met. Soc.*, **103**, No. 435, 1–28.

Hide R. and Mason P. J. (1978) "On the transition between axisymmetric and non axisymmetric flow in a rotating annulus subject to a horizontal temperature gradient." *Geophys. Astrophys. Fluid Dynamics*, **10**, 121–156.

Hignett P., Ibbetson A. and Killworth P. K. (1981) "On rotating thermal convection driven by non-uniform heating from below." *J. Fluid Mech.*, **109**, 161–187.

Hignett P., White A. A., Carter R. D., Jackson W. D. N. and Small R.M. (1985) "A comparison of laboratory measurements and numerical simulations of baroclinic wave flows in a rotating cylindrical annulus." *Quart. J. R. Met. Soc.*, **111**, 131–154.

Hookey N. A., Baliga B. R. and Prakash C. (1988) "Evolution and enhancement of some control volume finite-element methods – Part 1. Convection-diffusion problems." *Numerical Heat Transfer*, **14**, 255–272.

Horn R. A. and Johnson C. R. (1990) "Matrix Analysis." Cambridge University Press.

Hoskins B. J., McIntyre M. E. and Robertson A. W. (1985) "On the use and significance of isentropic potential vorticity maps." *Quart. J. R. Met. Soc.*, **111**, 877–946.

Houghton J. T. (1986) "The Physics of Atmospheres." Cambridge University Press.

James I. N., Jonas P. R. and Farnell L. (1981) "A combined laboratory and numerical study of fully developed steady baroclinic waves in a cylindrical annulus." *Quart. J. R. Met. Soc.*, **107**, 51–78.

Jonas P. R. (1980) "Laboratory experiments and numerical calculations of baroclinic waves resulting from potential vorticity gradients at low Taylor number." *Geophys. Astrophys. Fluid Dynamics*, **15**, 297–315.

Kaye G. W. C. and Laby T. H. (1956) "Tables of physical and chemical constants and some mathematical functions." Longmans, Green and Co.

Lanczos C. B. (1966) "The variational principle of mechanics." Toronto Press, Toronto.

Laurie D. P. (1982) "Solution of large systems of linear equations." Source unknown, 204–224.

Lewis S. R. (1992) "A quasi-geostrophic numerical model of a rotating internally heated fluid." *Geophys. Astrophys. Fluid Dynamics*, **65**, 31–55.

Li X. (1993) "Equatorial waves in planetary atmospheres." D. Phil. Thesis, S-DAOPP, Oxford.

Mason P. J. (1975) "Baroclinic waves in a container with sloping endwalls." *Phil. Trans. Roy. Soc. London*, (A), **278**, 397–445.

McIntyre M. E. (1968) "The axisymmetric convective regime for a rigidly bound rotating annulus." *J. Fluid Mech.*, **32**, part 4, 625–655.

Miller T. L. and Gall R. L. (1983a) "Thermally driven flow in a rotating spherical shell: axisymmetric states." *J. Atmos. Sci.*, **40**, 856–868.

Miller T. L. and Gall R. L. (1983b) "A linear transition curve for the baroclinic annulus." *J. Atmos. Sci.*, **40**, 2293–2303.

Moore I. (1992) "2-D building flow model." Private communication.

Patankar S. V. (1980) "Numerical heat transfer and fluid flow." Hemisphere, Washington, D.C.

Pedlosky J. (1987) "Geophysical fluid dynamics." Springer-Verlag.

Pedlosky J. and Klein P. (1991) "The nonlinear dynamics of slightly supercritical baroclinic jets." *J. Atmos. Sci.*, **48**, 1276–1286.

Potter D. (1973) "Computational Physics." Wiley.

Prakash C. and Patankar S. V. (1985) "A control volume finite-element method for solving the Navier-Stokes equations using equal-order velocity-pressure interpolation." *Numerical Heat Transfer*, **8**, 259–280.

Prakash C. (1986) "An improved control volume finite-element method for heat and mass transfer, and for fluid flow using equal-order velocity-pressure interpolation." *Numerical Heat Transfer*, **9**, 253–276.

Prakash C. (1987) "Examination of the upwind (donor-cell) formulation in control volume finite-element methods for fluid flow and heat transfer." *Numerical Heat Transfer*, **11**, 401–416.

Press W. H., Flannery B. P., Teukolsky S. A. and Vetterling W. T. (1989) "Numerical recipes: The art of scientific computing." Cambridge University Press.

Quon C. (1976) "A mixed spectral and finite difference model to study baroclinic annulus waves." *J. Comp. Phys.*, **20**, 442–479.

Rayer Q. G. (1992) "An experimental investigation of heat transfer by large scale motions in rotating fluids." D.Phil. Thesis, S-DAOPP, Oxford.

Read P. L. (1985) "Finite amplitude, neutral baroclinic waves and mean flows in an internally heated rotating fluid: 1. Numerical simulations and quasi-geostrophic 'free modes'." *Dyn. Atmos. Oceans*, **9**, 135–207.

Read P. L. (1986a) "Regions of axisymmetric flow in an internally heated rotating fluid." *J. Fluid Mech.*, **168**, 255–289.

Read P. L. (1986b) "Stable, baroclinic eddies on Jupiter and Saturn: A laboratory analog and some observational tests." *ICARUS*, **65**, 304–334.

Read P. L. (1986c) "Super-rotation and diffusion of axial angular momentum: I. "Speed limits" for axisymmetric in a rotating cylindrical fluid annulus." *Quart. J. R. Met. Soc.*, **112**, 231–251.

Read P. L. (1988a) "The dynamics of rotating fluids: the 'philosophy' of laboratory experiments and studies of the atmospheric general circulation." *Met. Mag.*, **117**, 35–44.

Read P. L. (1988b) "Finite amplitude, neutral baroclinic waves and mean flows in an internally heated rotating fluid: 2. Effects of spatially varying N^2." *Dyn. Atmos. Oceans*, **11**, 211–264.

Richardson L. F. (1922) "Weather prediction by numerical process." Cambridge University Press.

Rossby C. -G. (1940) "Planetary flow patterns in the atmosphere." *Quart. J. R. Met. Soc.*, **66**, Suppl., 68–87.

Schneider G. E. and Raw M. J. (1986) "A skewed, positive influence coefficient upwinding procedure for control-volume-based finite-element convection-diffusion computation." *Numerical Heat Transfer*, **9**, 1–26.

Schneider G. E. and Raw M. J. (1987a) "Control volume finite-element method for heat transfer and fluid flow using colocated variables – 1. Computational results." *Numerical Heat Transfer*, **11**, 363–390.

Schneider G. E. and Raw M. J. (1987b) "Control volume finite-element method for heat transfer and fluid flow using colocated variables – 2. Application and Validation." *Numerical Heat Transfer*, **11**, 391–400.

Simmons A. J. and Hoskins B. J. (1976) "Baroclinic instability on the sphere : normal modes of the primitive and quasi-geostrophic equations." *J. Atmos. Sci.*, **33**, 1454–1477.

Smith G. D. (1989) "Numerical solution of partial differential equations: finite difference methods" Oxford University Press.

Staniforth A. N. and Mitchell H. L. (1977) "A semi-implicit finite-element barotropic model." *Monthly Weather Review*, **105**, 154–169.

Staniforth A. N. and Mitchell H. L. (1978) "A variable resolution finite-element technique for regional forecasting with the primitive equations." *Monthly Weather Review*, **106**, 439–447.

Swanson Analysis Systems, Inc. (1989) "ANSYS engineering analysis system user's manual." Volume 1, Swanson Analysis Systems, Inc., Houston.

Tritton D. J. (1989) "Physical Fluid Dynamics." Oxford University Press.

White A. A. (1986) "Documentation of the finite difference schemes used by the Met. O. 21 2D Navier-Stokes numerical model." Meteorological Office Internal Report, Met. O. 21 IR86/3.

White A. A. (1988) "The dynamics of rotating fluids: numerical modelling of annulus flows." *Met. Mag.*, **117**, 54–62.

Williams G. P. (1969) "Numerical integration of the three-dimensional Navier-Stokes equations for incompressible flow." *J. Fluid Mech.*, **37**, part 4, 727–750.

Winslow A. M. (1966) "Numerical solution of the quasilinear Poisson equation in a nonuniform triangle mesh." *J. Comp. Phys.*, **1**, 149.

Wyatt L. R. (1981) "Linear and nonlinear baroclinic instability of the northern hemisphere winter zonal flow." *J. Atmos. Sci.*, **38**, 2121–2129.