METEOROLOGICAL OFFICE

Forecasting Research

Met O 11 Scientific Note No 12

Properties of the Lagrangian semi-geostrophic equations

by

M.J.P.Cullen and R.J.Purser

January 1989

Meteorological Office (Met O 11)
London Road, Bracknell, Berkshire RG12 2SZ, England

MET O 11 SCIENTIFIC NOTE Nº 12

PROPERTIES OF THE LAGRANGIAN SEMI-GEOSTROPHIC EQUATIONS

M.J.P. Cullen and R.J.Purser

Met O 11 (Forecasting Research)
Meteorological Office
London Road
Bracknell
Berkshire RG12 2SZ
ENGLAND

This paper has been submitted for publication in The Journal of the Atmospheric Sciences.

January 1989.

ABSTRACT

The Lagrangian form of the semi-geostrophic equations has been shown to possess discontinuous solutions which have been exploited as a simple model of fronts and other mesoscale flows. In this paper, it is shown that these equations can be integrated forward in time for arbitrarily long periods without breaking down, to give a 'slow manifold' of solutions. In the absence of moisture, orography and surface friction, these solutions conserve energy, despite the appearance of discontinuities.

In previous work these solutions have been derived by making finite parcel approximations to the data. This paper shows that there is a unique piecewise smooth solutions to the equations with general data, to which the finite parcel approximation converges. It is also shown that the time integration procedure is well-defined, and that the solutions remain bounded for all finite times.

Most previous results on the finite parcel solutions are restricted to the case of a Boussinesq atmosphere on an f plane with rigid wall boundary conditions. In this paper the results are extended to non-Boussinesq fluids, free-surface and periodic boundary conditions, and variable Coriolis parameter. The behaviour of the equations on a sphere and the effects of external forcing are discussed.

1. Introduction

This paper extends the theory of Lagrangian semi-geostrophic solutions introduced by Cullen and Purser (1984), henceforward referred to as CP. In CP, the theory was used to give a simple model of mature fronts. It was, however, conjectured that the theory had much greater potential to provide a simplified model of a wide variety of atmospheric flows. In this paper, this conjecture is followed up to show that the Lagrangian semi-geostrophic equations can be integrated in time to give a 'slow manifold' of atmospheric behaviour. Any balanced system of equations can in principle be used in this way, provided that it can be proved that solutions continue to exist for all time. Bennett and Kloeden (1981) proved such results for the quasi-geostrophic equations if boundary potential temperature gradients were excluded. In this paper we prove that the semi-geostrophic equations can be solved for arbitrarily long times, and conserve energy in the absence of moisture, friction, and mountains.

The concept of a 'slow manifold' has usually been used to describe a hypothetical subset of solutions of the full equations of motion which behave in a slowly varying manner, Leith (1980). This manifold is then used as a convenient way of describing those atmospheric motions which are directly associated with weather systems. It is now generally agreed, e.g. Vautard and Legras (1986), that no such manifold can be derived in a rigorous way from the equations of motion. An alternative procedure is to use a balanced system of equations to define the manifold. A precise procedure such as this is necessary if, for instance, the concept is being used to design data assimilation systems. However, whatever balanced system is used, it will not be an accurate approximation to the equations of motion everywhere. Its usefulness can therefore not be determined *a priori* but has to be established by experiment.

In CP the equations were written as a set of ordinary differential equations for the potential temperature and absolute momentum components along parcel trajectories. By approximating the data by a finite set of parcels, it was proved that a symmetrically stable solution of the equations could always be derived as a sequence of rearrangments of the parcels. This rearrangement is usually smooth, and determines the ageostrophic circulation. CP showed that it could nevertheless generate discontinuities in potential temperature and absolute momentum which formed a simple model of fronts. The rearrangement can also include cases where parcels have to jump to new stable positions, in which case it forms a simple model of penetrative convection. It was conjectured that by taking progressively larger numbers of elements, a sequence of approximate solutions would be generated which converges to the solution of the 'continuous' problem. This result is proved in this paper, by making use of the extensive theory of generalised solutions of the Monge-Ampere equation.

The finite parcel construction has been implemented as a numerical method, usually called the 'Geometric' method, by Chynoweth (1987). It has been used to explain aspects of mountain flow, Cullen et al. (1987), Shutts (1987a) and penetrative convection, Shutts (1987b), Shutts et al. (1988). These applications include cases where the solutions appear to contain discontinuities representing fronts. There is always a difficulty, however, in distinguishing the discontinuities between individual parcels in the approximation, and discontinuities which remain in the limit of infinitely small parcels. This is a situation where it is important to understand the convergence properties of the method. An even more difficult situation, not treated in this paper, is that of penetrative convection, where parcels 'jump' in physical space. In that case it is necessary to show that the method converges to a solution giving continuous mass transfer at a well-defined rate from a source region to a sink region.

The next aim of the paper is to establish results about the time evolution of the system. In CP it was conjectured that a solution could be constructed by integrating the ordinary differential equations in time for the parcel properties, and finding the arrangement of the parcels in physical space at each time. This is the implementation for the semi-geostrophic case of the general procedure for solving balanced equations set out by Hoskins, McIntyre and Robertson (1985) where the potential vorticity is advected along trajectories and the remaining fields found using an 'invertibility' principle. The Geometric method can be regarded as the inversion procedure in the special case of piecewise constant data. In simple problems, such as most of the published applications of the Geometric method, the evolution equations can be solved analytically, and the solution derived by a single construction at output time. In general, the evolution equations can only be solved with a knowledge of the current position of the parcels, and the solution must be obtained by time-stepping. It is necessary to ensure that a well-defined solution is obtained as the timestep tends to zero. We first prove that this procedure is well-defined for strictly positive potential vorticity. The limit solution then defines the ageostrophic circulation. If the potential vorticity is allowed to approach zero, the ageostrophic circulation is not well-defined but the evolution of the geostrophic field still converges to a well defined limit. This covers such cases as the evolution of well-mixed boundary layers. Such a solution is physically relevant because changes in the positions of individual parcels in a well-mixed layer do not affect the pressure field. This extension is the first step towards ensuring that the penetrative convection solutions are well-defined.

Salmon (1985) and Shutts (1988) have derived forms of the semi-geostrophic equations from Hamilton's principle. Both make use of the Lagrangian form of the equations. This means that the equations have conservation properties which make them a suitable candidate for extended time integrations. The next step is therefore to show that the Lagrangian equations can be integrated forward in time for indefinitely long periods. It is shown using conservation of potential vorticity and the energy equation for individual parcels that a solution exists for all finite times. This analysis is most conveniently carried out by using the 'dual' form of the equations set out in Purser and Cullen (1987), referred to hence as PC. In this formulation we study how parcels move round data space. The constraints which are used in the analysis must be respected by numerical methods if their nonlinear stability and accuracy is to be proved for long time integrations.

The results in CP and the extensions set out above apply to the f-plane equations in a bounded region, using the form of Boussinesq approximation and lower boundary condition introduced by Hoskins and Bretherton (1972). The rest of the paper shows how much of the theory can be extended to the case where these extra approximations are not made. We thus consider the non-Boussinesq case, free-surface boundary conditions, periodic boundary conditions, and variable Coriolis parameter f. An extra approximation introduced by Salmon (1985) is crucial in applying the results directly to the variable f case. Slightly weaker results, however, can still be proved by iteration without this approximation. The theory has been extended to spherical geometry by Shutts (1988) using a modified version of the equations in which the component of geostrophic wind parallel to the axis of rotation is neglected in the Hamiltonian. In this paper we study the unmodified version of the equations in spherical geometry. We finally illustrate how external forcing terms are included in the solution, and discuss their effect on the long-term evolution.

2. Solutions of the semi-geostrophic equations in a bounded region on an f-plane.

a. *Basic equations*

The equations are written in their usual form with z a function of pressure and the Boussinesq approximation made, Hoskins and Draghici (1977).

$$Du_g/Dt + f(v_g - v) = 0, \tag{2·1}$$

$$Dv_g/Dt + f(u - u_g) = 0, \tag{2·2}$$

$$D\theta/Dt = 0, \tag{2·3}$$

$$DV/Dt = 0, \tag{2·4}$$

$$(fv_g,\ -fu_g,\ g\theta/\theta_o) = \nabla\phi, \tag{2·5}$$

where $D/Dt \equiv \partial/\partial t + \underline{u}\cdot\nabla$. The equations are to be solved in a closed region Ω in $\mathbf{x}\equiv(x,y,z)$ with zero mass flux through the boundary. No boundary conditions can be given on the geostrophic variables. The continuity equation has been written in Lagrangian form with V the specific volume. For this first application, we rewrite the equations using the geostrophic coordinates as dependent variables, as in PC:

$$\mathbf{X} \equiv (X,Y,Z) = (x+v_g/f, y-u_g/f, g\theta/(f^2\theta_o)), \tag{2·6}$$

$$DX/Dt + f(Y-y) = 0, \tag{2·7}$$

$$DY/Dt + f(x-X) = 0, \tag{2·8}$$

$$DZ/Dt = 0, \tag{2·9}$$

$$D\rho/Dt = 0. \tag{2·10}$$

ρ is the inverse potential vorticity, $\partial(\mathbf{x})/\partial(\mathbf{X})$. If the equations have differentiable solutions, (2.10) was shown to follow from (2.1) to (2.4) by Hoskins and Draghici (1977). (2·7) to (2·10) can be reinterpreted as equations defining the motion of points in geostrophic and isentropic coordinates (X,Y,Z). We henceforward refer to $\mathbf{X}$=(X,Y,Z) space as *data space*. The equations in data space are then:

$$\mathbf{U} = (f(y-Y), f(X-x), 0) \qquad (2\cdot11)$$

$$\partial\rho/\partial t + \nabla_x\cdot(\rho\mathbf{U}) = 0. \qquad (2\cdot12)$$

As indicated, all derivatives are now with respect to $\mathbf{X}$ rather than $\mathbf{x}$, so (2·12) does not take the same form as (2·10). However, for the particular $\mathbf{U}$ given by (2·11), $\nabla_x\cdot\mathbf{U}=0$, so that $D_x\rho/Dt$ is zero. $\mathbf{U}$ must be redefined when forcing terms are included, and the more general equation (2·12) will be needed to calculate $D_x\rho/Dt$.

As shown in **PC**, the physical coordinates ($\mathbf{x}$) can be derived as the gradients of a potential function R, where

$$R = f^2\mathbf{x}\cdot\mathbf{X} - P, \qquad (2\cdot13)$$

and

$$P = \phi + \tfrac{1}{2}f^2(x^2+y^2). \qquad (2\cdot14)$$

P is the generalised geopotential function introduced by **CP**. It was shown there that the condition for static, inertial and symmetric stability is that P is a convex function.

b. *Representations of atmospheric states at a fixed time*

We follow the general prescription of Hoskins et al. (1985) for solving balanced systems of equations. An invertibility principle allows all variables to be derived from the potential vorticity. The potential vorticity equation is then used to advance the solution in time. We first show how the atmospheric state is calculated from the potential vorticity, using the semi-geostrophic definition of balance. The geometric construction of CP is the invertibility principle for the special case when ρ is made up of delta functions. Since the general case will be derived as the limit of such constructions, it will turn out to be most convenient to use ρ rather than q as the history carrying variable.

Suppose that at some time t, ρ is given as a function of **X**, subject to the condition that integrating ρ over all **X** gives the volume of the given region Ω in physical space. ρ is zero for all values of the data **X** which are not taken by any point in the fluid. Then the remaining variables can be computed in principle by the following steps:

(i) Use the definitions of ρ and R to give

$$\det(\partial^2 R/\partial x_i \partial x_j) \equiv \partial(\nabla R)/\partial(\mathbf{X}) = \rho. \qquad (2\cdot 15)$$

(ii) (2·15) is a Monge-Ampere equation for R in terms of ρ. It must be solved given the boundary condition that $\nabla R \equiv \mathbf{x}$ is always within Ω for all **X**. Note that this avoids the need to specify any information about the geostrophic variables on the domain boundary. This is an important advantage over some of the systems discussed by Gent and McWilliams (1983).

(iii) The solution for R(**X**) allows ∇R to be calculated, and hence the mapping $\mathbf{X} \to \mathbf{x}$ which assigns data values to points in physical space.

(iv) All information required to advance the solution of (2·11) - (2·12) in time is now available. The values of the total velocity **u** are not needed for the time integration. Where they are well defined, they can be diagnosed during the time integration by calculating $(\mathbf{x}(t+\Delta t)-\mathbf{x}(t))/\Delta t$.

The boundary value problem (ii) is a standard one in the theory of the Monge-Ampere equation. The theory of this equation is set out in Pogorelov

(1964), henceforward referred to as **PG**. The extensions needed to apply this theory to the meteorological problem are proved in Cullen, Norbury and Purser (1988), hereafter referred to as **CNP**. In this paper we give an intuitive geometrical description of the results and proofs.

In the geometrical interpretation, the potential function P is interpreted as a surface, and the extra condition is imposed that the surface must be convex when viewed from below. This is equivalent to requiring general parcel stability for the atmospheric state, Shutts and Cullen (1987). The curvature of P is related to the potential vorticity, and convexity implies non-negative potential vorticity. It implies that X is monotonically increasing in x, Y in y and Z in z. It can then be shown that, if the physical domain Ω is convex, the 'dual' potential function R is also convex. The inverse potential vorticity ρ is related to the curvature of R, and convexity of R implies non-negative ρ. Since ρ is defined for all **X**, R is an infinite convex surface. These ideas are now developed more formally.

Definition.

Given any point O of an infinite convex surface, construct the cone of rays extending upwards from the surface which do not intersect it again. The set of directions taken by these rays is called the *limit cone* of the surface, Fig. 1. It can be shown that it is independent of the choice of O.

Definition.

Given a segment of a surface containing a point O, construct the normals to the surface at all points within the segment. Move these normals to O, thus constructing a cone at O, Fig. 2. Calculate the ratio of the solid angle of this cone to the area of the segment. This is the *curvature* of the segment. The limit, λ, of this ratio as the segment shrinks towards

O is called the *curvature* of the surface at O. (A rather fuller definition is given by **PC**, p.3453.)

The inverse potential vorticity ρ associated with the segment can be calculated directly from the curvature. It is shown in **PC** (p.3454) that

$$\rho = \lambda/\surd(1+|\nabla R|^2)^5 = \lambda/\surd(1+x^2+y^2+z^2)^5.$$

This is an example of a *generalised curvature*. **PG** defines a generalised curvature to be λ multiplied by any positive function of $\mathbf{x}$, $\mathbf{X}$, and R.

Theorem 1.

There is a unique infinite convex surface R with given generalised curvature ρ and limit cone **W** up to an arbitrary additive constant, given the condition that the integral of ρ over **X** is the solid angle contained in **W**.

Proof.

The formal statement and proof of existence are given by **PG** (p.33), and the uniqueness result is given by **CNP**. The method used is to approximate the surface by polyhedra with successively more faces, and thus show that the geometric construction introduced by **CP** converges. The uniqueness result is proved by contradiction. If there are two solutions R_1 and R_2, the arbitrary constant and the common limit cone V means that we can assume that $R_2 \geqslant R_1$ everywhere, with equality at at least one point O. Then R_2 can only diverge from R_1 if there is some neighbourhood of O where the solid angle contained in R_2 over the neighbourhood is strictly less than that of R_1. The generalised curvature of R_2 over the neighbourhood is therefore also strictly less than that of R_1, contrary to hypothesis. (The rigorous mathematical proof requires care because R_1 and R_2 may contain 'creases' which make the solid angle contained in certain regions ill-defined).

Definition.

A *displacement* of a field $\rho(\mathbf{X})$ is a mapping $\mathbf{X}$ to $\mathbf{X}'=\mathbf{X}+\Delta\mathbf{X}$ and a new function ρ' such that $\rho'(A')=\rho(A)$ where A is any subset of $\mathbf{X}$ space and A' is its image under the displacement. The case where the displacement maps a finite region onto a single point is covered.

Theorem 2.

The surface R changes continuously under changes to ρ which maintain the physical requirement $\rho\geqslant 0$ and conserve the integral of ρ over $\mathbf{X}$.

Proof

The mathematical version is given in **CNP**. Any change to ρ which satisfies these conditions can be generated by a displacement of the ρ field as defined above. Measure the displacement required to achieve the change by its maximum value of $\Delta\mathbf{X}$. Consider a sequence of displacements with $\max(\Delta\mathbf{X})$ tending to zero, and associated functions ρ' and R'. A standard argument shows that the functions R' must have a limit R". Let ρ'' be the generalised curvature of R". The definition of displacement, however, means that the generalised curvatures $\rho'(A)$ must tend to $\rho(A)$ over any subset A of $\mathbf{X}$ with boundary ∂A where $\rho(\partial A)=0$. The boundary of A has measure zero in $\mathbf{X}$, and so for almost all choices of A, $\rho(\partial A)$ will be zero. Thus $\rho''(A)=\rho(A)$ for almost all A, and the argument used to prove uniqueness in Theorem 2 shows that R"=R.

These results show that given any $\rho(\mathbf{X})$, the geostrophic pressure and wind fields can be constructed uniquely in physical space. This solution can be achieved by taking progressively more elements in the Geometric method, and may contain discontinuities in the limit solution. It is these discontinuities which form a simplified model of mature fronts. Unlike the more familiar quasi-geostrophic application of the invertibility principle, no boundary values can be given for the geostrophic fields. In the quasi-geostrophic case, the isentropic potential vorticity must be given, which means that ρ is given as a function of x,y and Z. In the semi-geostrophic case ρ is given as a function of X,Y and Z. The difference is the familiar one of replacing physical by geostrophic coordinates. The quasi-geostrophic problem requires Z to be specified as a function of x and y on the physical boundary. In the semi-geostrophic case , consider the region of **X** space with $\rho \neq 0$. The monotonicity condition means that any point on the convex hull of this region corresponds to a point on the physical boundary (CP, p.1486). In particular, the point with the largest X, Y or Z does so. At these points any one of the **X** coordinates can be considered as given in terms of the other two, so we can write $Z=Z(X,Y)$. This is the semi-geostrophic version of the quasi-geostrophic requirement $Z=Z(x,y)$. However, the practical effect is very different. In the semi-geostrophic case no values of the geostrophic variables can be specified in advance at particular points on the physical boundary, the entire solution is constructed from ρ. No unnatural boundary conditions of the type needed in the various systems examined by Gent and McWilliams (1983) are required.

It was shown in Shutts and Cullen (1987) that the atmospheric state obtained by the above inversion procedure is a minimum energy state with respect to smooth volume preserving rearrangements of the fluid conserving **X**. This characterisation also applies when the rearrangement is not smooth, as may be the case for general data. The necessary extension of the result is given in the following theorem:

Theorem 3

The minimum geostrophic energy

$$E = \int_\Omega \tfrac{1}{2}(u_g^2+v_g^2)-g\theta z/\theta_0 \, d\mathbf{x} \qquad (2\cdot16)$$

obtainable by rearranging the fluid within Ω and preserving the value of **X** following each particle is given by solving for R as described in Theorems 1 and 2, and then finding the particle positions from ∇R.

Proof

The full mathematical version is given in **CNP**. The energy integral can be written

$$f^2\int_\Omega\{\tfrac{1}{2}((y-Y)^2+(x-X)^2)-zZ\}\,d\mathbf{x}. \qquad (2\cdot17)$$

The only terms which change under the prescribed rearrangements are

$$f^2\int-\{yY+xX+zZ\}\,d\mathbf{x}. \qquad (2\cdot18)$$

This integral is bounded under rearrangements because both Ω and the data are bounded, and it can then be shown that there must be some rearrangement that minimises E. In this configuration, any cyclic interchange of particles originally at $(\mathbf{x}_1,\cdots,\mathbf{x}_n)$ must not decrease E, so that

$$-\mathbf{X}_1\cdot(\mathbf{x}_2-\mathbf{x}_1)-\cdots-\mathbf{X}_n\cdot(\mathbf{x}_1-\mathbf{x}_n) \geqslant 0. \qquad (2.19)$$

A standard theorem proved in Rockafellar (1970,p.238) states that **X** must then be the gradient mapping of a convex function $P(\mathbf{x})$ defined on Ω. (This mapping may be multivalued at some points in Ω). Since there is a unique convex function constructed according to Theorems 1 and 2, this must be the desired minimising function.

c. *Integrations of the equations in time*

Now consider the integration of (2·11) and (2·12) in time. (2·12) does not make sense when the data take a uniform value $\mathbf{X}$ over a region Ξ of positive volume in Ω because $\mathbf{x}-\mathbf{X}$ is then multivalued. The potential vorticity is then zero over this region. ρ is a delta function at $\mathbf{X}$ whose integral is the volume of Ξ. (2·12) must be replaced by the integral form

$$\partial/\partial t(\textstyle\int_V \rho d\tau) + \int_{\partial V} \rho \mathbf{U}\cdot\mathbf{n}ds = 0, \qquad (2\cdot 20)$$

where the integrals are taken over fixed volumes V in $\mathbf{X}$ space with boundaries ∂V. Physically, the semi-geostrophic equations do not define the motion within a region where the data are uniform, such as a well-mixed layer with zero absolute vorticity. Motions within such a region, however, have no effect on the remainder of the balanced solution, so the equations have in effect been reformulated to ignore them.

The standard method of showing that equations like (2·12) and (2·20) can be integrated forward in time over an interval (0,T) in a well-defined way is to divide the interval into timesteps and prove that the solutions obtained converge to a unique limit as the timestep tends to zero. This is a trivial matter for conventional evolution equations with smooth solutions, but is not trivial for balanced equations where some of the variables are determined implicitly. Essentially it then requires that an ellipticity condition is satisfied.

The proof requires $\nabla_x\cdot(\rho\mathbf{U})$ to vary in a continuous way as the ρ field evolves under the conservation law (2·20). In the simple case where ρ is initially bounded and remains so throughout (0,T), it is shown in **CNP** that the continuity result in Theorem 2 is sufficient. (2·11) and (2·12) are used. The $\mathbf{U}$ defined by (2·11) is non-divergent, even if it contains discontinuities. This is because the integral $\int_{\partial V}\mathbf{U}\cdot\mathbf{n}ds$ around the boundary of any subset V of **X** space can be written as

$\int_{\partial V}(\partial/\partial s(R-\frac{1}{2}(X^2+Y^2))ds,$

which vanishes. Hence we obtain

Theorem 4

(2·11) and (2·12) can be integrated forward for a finite time interval, given that ρ is initially bounded.

In the general case where ρ contains delta functions, corresponding to regions of zero potential vorticity, it is still possible to show that the evolution of the potential functions R and P is well-defined. Thus the geostrophic part of the flow can still be calculated. Individual particle trajectories are undefined within regions where the potential vorticity is zero, since two particles with the same **X** can be exchanged without any effect on the geostrophic fields and hence the balance requirements.

In order to do this **U** must be redefined, since (2·11) will be multivalued at points where ρ is unbounded. In order to preserve energy conservation, the appropriate choice of **U** will be shown to be

$$\mathbf{U}(\mathbf{X}) = \int_V f(y-Y, X-x, 0)d\mathbf{x}/\mu(V), \qquad (2\cdot21)$$

where V is the region of Ω with data values **X**, and μ(V) is its volume.

Theorem 5

R can be integrated forward in time for a finite interval.

Proof

The procedure of updating ρ according to (2·20) and (2·21), and finding R, can be set out as solving the equation

$$\partial R/\partial t = F(R). \qquad (2\cdot22)$$

Theorem 2 shows that F(R) is continuously dependent on any changes to ρ that can arise from solving (2·20), so the standard time integration argument used for Theorem 4 proves the result.

It is then natural to ask if the equations can be integrated forward indefinitely. This is necessary if they are to be used as a 'slow manifold', or a model describing the evolution of the balanced part of the atmospheric flow over long time periods. Hoskins (1975) shows that they conserve energy, at least for smooth solutions, and they also conserve potential vorticity. In the quasi-geostrophic case, such properties can be used to show that the equations can be integrated forwards indefinitely, under restrictive assumptions on the boundary conditions, Bennett and Kloeden (1981).

We first show that the equations still conserve energy when the solutions are discontinuous. Since Theorem 1 shows that a general solution can be treated as the limit of a sequence of piecewise constant approximations, we prove the result for piecewise constant **X**. The general case can be recovered in the limit.

Theorem 6

Given piecewise constant data $\mathbf{X}_i$ at t=0, the integration of (2·20) and (2·21) in time conserves the total geostrophic energy (2·16).

Proof

Using the form (2·17), and dropping the factor f^2, the integral of $\frac{1}{2}(x^2+y^2)$ is constant in time. Consider the rate of change of

$$\int_\Omega\{\tfrac{1}{2}(X^2+Y^2)-xX-yY-zZ\}d\mathbf{x}.$$

$$= \sum_i \int_{\Omega i}\{\tfrac{1}{2}(X_i{}^2+Y_i{}^2)-x_iX_i-y_iY_i-z_iZ_i\}\rho_i dx, \qquad (2\cdot 23)$$

where $\mathbf{x}_i$ is the centroid of the region Ω_i of Ω with data $\mathbf{X}_i$, and ρ_i is its volume. According to (2·21)

$$dX_i/dt = f(y_i-Y_i)$$

$$dY_i/dt = f(X_i-x_i) \qquad (2\cdot 24)$$

$$dZ_i/dt = 0.$$

Differentiating (2·23) with respect to time and substituting (2·24), all terms within the integral cancel except

$$-\mathbf{X}_i \cdot d\mathbf{x}/dt. \qquad (2\cdot25)$$

This is a discrete representation of the term $\mathbf{u}\cdot\nabla\phi$ which is derived in the energy equation for the continuous case. Since the volume of the regions Ω_i is conserved, the physical velocity within each region can be represented as a non-divergent $\mathbf{u}$ with mean value $d\mathbf{x}_i/dt$. The component of $\mathbf{u}$ normal to region interfaces is continuous across them but the tangential component need not be. Then the term $\int_{\Omega i}\{-\mathbf{X}_i\cdot d\mathbf{x}_i/dt\}d\mathbf{x}$ becomes

$$\int_{\Omega i} -\mathbf{u}\cdot\nabla P \; d\mathbf{x}$$

$$= -\int_{\partial\Omega i}(P\mathbf{u})\cdot\mathbf{n} \; ds$$

around the boundary of Ω_i. When summed over the Ω_i, this term cancels at region interfaces by continuity of P and $\mathbf{u}\cdot\mathbf{n}$, and vanishes on the boundary of Ω because no fluid can flow across it. This proves the result.

In order to study the long time behaviour of the system, consider the evolution of $\mathbf{X}$ following a fluid particle. The aim is to show that $\mathbf{X}$ remains bounded indefinitely, which means that the equations will have a solution for all time with total energy conserved. The Z component never changes, so consider the evolution of the (X,Y) components. Let $W=\sqrt{(X^2+Y^2)}$ and α be the angle made by (X,Y) with the X axis, Fig. 3. Denote the position of the particle in the (X,Y) plane by A. If $\mathbf{x}$ is the position of the particle in Ω, let B be the point in the (X,Y) plane with coordinates (x,y) and let β be the angle between (x,y) and the x axis. Then (2·7) and (2·8) show that A rotates around B with period $2\pi f^{-1}$ and that W satisfies the equation

$$dW/dt = f\sqrt{(x^2+y^2)}\sin(\beta-\alpha). \qquad (2\cdot26)$$

W can only increase at a rate bounded by the diameter of Ω. Furthermore, if

$|\mathbf{X}|$ becomes large, it will have to be close to the point in Ω where the normal direction is $\mathbf{X}$, by monotonicity of any component of $\mathbf{X}$ in the corresponding component of $\mathbf{x}$. This is illustrated in Fig. 4 which shows contours of P, recalling that $\mathbf{X}=\nabla P/f^2$. Energy conservation means that a region with large $|\mathbf{X}|$ and hence large W must have volume proportional to $|\mathbf{X}|^{-2}$. Let Π be the closed convex curve in the (X,Y) plane whose coordinates are equal to the (x,y) coordinates of the points of the boundary of Ω where the tangent plane is parallel to the z axis. This curve and the subsequent construction are illustrated in Fig. 5. For large W, the point B will be close to that point of Π where the tangent is normal to the direction (X,Y). Now let C be the point where the direction (X,Y) intersects Π and let r be the distance OC, where O is the origin in the (X,Y) coordinate system. Let the angle α that $\mathbf{X}$ makes to the X axis increase. If $dr/d\alpha>0$, as shown in Fig. 5, it is clear that OB makes a greater angle to the X axis than OC and so $(\beta-\alpha)>0$. As α rotates round $(0,2\pi)$, the integral of $\sin(\beta-\alpha)$ will be approximately the integral of $dr/d\alpha$, which vanishes. If W is large, (2·7) and (2·8) show that α will indeed rotate in this way, and so W is likely to tend to a constant rather than increase indefinitely. Though this argument is not rigorous, it suggests very strongly that the solutions will remain bounded for all time.

The very strict bound (2·26) on the growth of W, and the likelihood that it remains bounded for all time, mean that the semi-geostrophic equations (2·1) to (2·5) can be integrated for indefinitely long periods while conserving total energy. It is known that a slightly simplified version of these equations can describe the growth of baroclinic waves including fronts, Hoskins and West (1979). The existence result therefore suggests that a complete inviscid baroclinic lifecycle exists. When such a lifecycle is computed using a primitive equation model, up to half the net available potential energy depletion is accounted for by dissipation in horizontal diffusion terms, Simmons and Hoskins (1978). The semi-geostrophic lifecycle is thus fundamentally different.

Since (2·1) to (2·5) are inviscid, they can be run backwards as well as forwards in time. This suggests that the behaviour of synoptic systems with embedded fronts can be modelled as a completely reversible system in which fronts are regions of discontinuity that appear and disappear from time to time, and are entirely controlled by the large scale flow.

3. Extensions of the theory

The f plane Boussinesq system discussed in section 2 takes a particularly simple form which makes it suited to analytic solutions for smooth data using the geostrophic coordinate transformation and to simple construction of solutions for piecewise constant data by the geometric method. However, for assessing the applicability of the theory, it is important to know if the basic qualitative structure of the solutions is affected by these approximations. In this section they are withdrawn one by one to see if the theory still applies.

a. *Non-Boussinesq fluids*

As shown in the original derivation of (2·1) to (2·5) by Hoskins and Bretherton (1972), (2·4) takes the general form

$$D(r(z)V)/Dt = 0. \tag{3·1}$$

The equations are now to be solved in a region Ω of the form $\Omega(x,y)\times(0,H)$. $r(z)$ is a prescribed function (a pseudo-density) which decreases with z. If the whole mass of the atmosphere is included in the model, then $r(z)$ tends to zero as z tends to H. The special form of Boussinesq approximation used neglects the variation of $r(z)$ following a trajectory. This is not too serious for dry adiabatic processes where conservation of θ makes large excursions in z unlikely. Since the change from (2·4) to (3·1) is essentially a rescaling of the measure of mass in the z coordinate, it is not likely that the qualitative nature of the solution will change. The only difficulty is that $r(z) \to 0$ at the upper boundary.

In order to extend the theory to use (3·1) instead of (2·4), it is sufficient to prove that the inversion procedure of section 2b can still be applied. (2·12) becomes

$$\partial/\partial t(\rho r(z)) + \nabla\cdot(\rho r(z)\mathbf{U}) = 0. \qquad (3\cdot 2)$$

Thus $\rho r(z)$ is the predicted quantity rather than ρ. The same procedure for constructing the solution as in section 2 can be followed. The theory given by **PG**, p.33, proves that a convex surface R can be constructed if ρr is the given quantity rather than ρ itself because the definition of generalised curvature includes the form $\rho\mu(x,y,z)$, where μ is an arbitrary prescribed positive function.

PG does not give a uniqueness proof for this case, so the uniqueness proof in **CNP** must be extended. Supposing that there are two solutions, R_1 and R_2, the critical part of the proof, as summarised in the proof of Theorem 1 above, is the identification of a region where the solid angle contained in R_2 is strictly greater than that in R_1. The integral over this region of ρ multiplied by $\mu(\mathbf{x})$ will then be different for R_2 and R_1, establishing the required contradiction. The only difficulty is near the upper boundary where $r(z)$ tends to zero. If, however, the only region where $R_1 \neq R_2$ adjoins $z=H$, there must still be a z strictly less than H where $r(z_1)$ is strictly positive and $R_1(z_1) \neq R_2(z_1)$. The same argument can then be applied using z_1 as the upper boundary rather than H.

b. *Free surface boundary conditions*

The easiest extension from the case of a fluid filling a prescribed volume is to the shallow water equations, where the fluid is within a region Ω in the (x,y) plane but its depth can vary. The theory can treat the important cases where the fluid does not cover the whole of Ω, or where it detaches from the boundary of Ω during the time evolution. Neither are easily treated by conventional shallow water theory.

Consider first the 'shallow water' form of the semi-geostrophic equations in two space dimensions:

$$Du_g/Dt + f(v_g - v) = 0, \tag{3·3}$$

$$Dv_g/Dt + f(u - u_g) = 0, \tag{3·4}$$

$$D(\phi V)/Dt = 0, \tag{3·5}$$

$$(fv_g, -fu_g) = \nabla\phi. \tag{3·6}$$

The potential vorticity equation now becomes

$$D(\rho\phi)/Dt = 0. \tag{3·7}$$

ρ is still defined as $\partial(\mathbf{x})/\partial(\mathbf{X})$, with $\mathbf{X}$ defined by the first two components of (2·6). The equations are to be solved in a closed region Ω in $\mathbf{x}=(x,y)$ with zero mass flux across the boundary. In order to understand how the solution is constructed, consider piecewise constant data as in CP.

Consider first the problem solved in section 2 where (3·5) is repalced by $DV/Dt = 0$, and the fluid fills Ω. The solution is constructed as a polyhedral shell $P=\phi+\frac{1}{2}f^2(x^2+y^2)$ whose faces have given gradients $\mathbf{X}_i$ and areas V_i. The areas V_i projected onto the (x,y) plane add up to the area of Ω. The solution is only determined up to an additive constant.

Given the same form of data, the free surface construction is illustrated in Fig. 6. For simplicity, the diagram is drawn for one space dimension only. It is now the fluid volumes μ_i, equal to the integrals of

$(P-\frac{1}{2}f^2(x^2+y^2))V_i$ over each face, which are given. The values of V_i are not given. The construction is carried out by starting from the convex cylinder with lower end surface $s=\frac{1}{2}f^2(x^2+y^2)$, $(x,y)\in\Omega$, and vertical sides intersecting the boundary of Ω. The surface P is then constructed as the intersection of tangent planes with this cylinder. These planes are displaced parallel to the s axis a sufficient distance to obtain the correct μ_i. As in the incompressible case, the crucial property is that increasing the value of s associated with one face increases its value of μ at the expense of neighbouring faces. The argument of **CP** can then be followed through to prove the existence of a unique solution. If the total volume of fluid is sufficient, the whole of Ω will be covered. If not, the fluid will only partially fill Ω as in the example shown in Fig. 6. In a time dependent calculation, this property of the model allows boundary separation to be represented. This may be particularly important if the model is applied to the ocean. This method of construction is equivalent to the procedure used by Shutts (1987a) to solve the problem of two-fluid flow over a weir. In his case the fluid interface plays the role of the free surface.

This result can be extended to general data by using methods similar to those in section 2. Define the conserved quantity $\rho\phi$ in (3·7) to be

$$\mu(\mathbf{X}) = \max(f^2\mathbf{x}\cdot\mathbf{X}-R(\mathbf{X})-\tfrac{1}{2}f^2(x^2+y^2),0)\rho(\mathbf{X}). \qquad (3\cdot 8)$$

This expression is obtained by substituting for ϕ in terms of R using (2·13) and (2·14). The zero value represents values of **X** with no associated fluid volume. **PG**, pp. 29-33, includes functions like (3·8) in his definition of generalised curvature, so the existence theorem holds for this case. The uniqueness proof can be carried out by similar methods to those used in **CNP** for the fixed boundary problem, though it is easier to work in physical space rather than data space.

The second extension to the theory made in this section withdraws the approximation to the lower boundary condition made in section 2, where zero mass flux is imposed at $z=0$, though this is not the physical lower boundary. The correct lower boundary condition in the z coordinate case be shown to be

$$\phi = 0 \text{ at } z = H(1 - (p_*/p_o)^{(\gamma-1)/\gamma}), \qquad (3\cdot8)$$

using the definition of z in Hoskins and Bretherton (1972). p_* is the surface pressure. Only the integral of p_* over Ω is given. The other boundary conditions are unaltered, so that there is zero mass flux through the boundary of a closed region Ω of the (x,y) plane and through $z=H$. The interior equations are still (2·1) to (2·5).

The method of solution is illustrated for a version of the problem in a (x,z) cross-section, so that P and ϕ are surfaces in three dimensions and can be easily visualised. In this case Ω is a closed line segment in x. Piecewise constant data is used as in **CP**. In the fixed boundary problem, the construction of **CP** starts with an infinite cylinder Π in (x,z,s) space, whose cross section is $\Omega\times(0,H)$. Then P is the lower boundary of the intersection of Π with the half spaces

$$s \geqslant xX_i + zZ_i + s_i. \qquad (3\cdot9)$$

It can then be proved that there is a unique construction in which the faces of P have prescribed areas whose sum is the area of $\Omega\times(0,H)$.

In the free surface case, the only change is that Π is the curved half cylinder

$$s \geqslant \tfrac{1}{2}f^2x^2,\ z \leqslant H,\ x \in \Omega. \qquad (3\cdot10)$$

This is illustrated in Fig. 7. It is convex, and so a convex surface P can be constructed as the intersection of Π with half spaces (3·9). Assume that

$Z_i > 0$ for all i, which simply requires positive θ. Then it can be seen from Fig. 7 that the resulting surface is equal to $\frac{1}{2}f^2x^2$ for $z \leq z_o$, for some z_o, and otherwise has the required boundaries $z=H$ and $x \in \partial\Omega$. The required solution of the equations is that part of the surface with $z \geq z_o$. The proof that a solution in which each face has the required area can be constructed is then exactly the same in **CP**. To see that this solution is unique, suppose there are two solutions P_1 and P_2, with lower boundaries $z_1(x)$ and $z_2(x)$ where $P = \frac{1}{2}f^2x^2$. Choose $z_o \leq z_1(x), z_2(x)$ for all x in Ω. Extend P_1 and P_2 to the region $\Omega \times (z_o, H)$ by setting $P_1=\frac{1}{2}f^2x^2$ for $z_o \leq z \leq z_1(x)$, $P_2=\frac{1}{2}f^2x^2$ for $z_o \leq z \leq z_2(x)$. The area of this extension is the same for P_1 and P_2, because the area of each face is given. Then P_1 and P_2 are two different solutions to the fixed boundary problem, and can only differ by a constant. Since the two surfaces are equal for $z=z_o$, they are equal everywhere.

The proof in the three-dimensional case is exactly the same, though not easy to visualise. The extension to general data can be carried out by the same methods as in section 2. The only extra step needed is to ensure that it is possible to choose a z_o in the uniqueness proof. This fact follows from boundedness of the data.

c. *Periodic boundary conditions*

The next extension is from the fixed boundary problem to periodic boundary conditions. This is needed to treat the important idealised problem of flow in a periodic channel, as well as a first step to extending the theory to spherical geometry. The only difficulty arises with the definition of absolute momentum components, which have to be redefined allowing the coordinate to 'wrap around'.

The geometrical construction of **CP** was extended to a domain in an

(x,z) cross-section with periodic boundary conditions in x by Chynoweth (1987). His method is used here to extend the result to general periodic boundary conditions. The geopotential $\phi(x,y,z)$ is defined for all $\mathbf{x}$, with the periodicity condition

$$\phi(x+kK, y+lL, z+mM) = \phi(x,y,z), \tag{3·11}$$

for all integers k,l,m and fixed values (K,L,M). The potential P is now defined by

$$P(x+kK, y+lL, z+zZ) =$$

$$P(x,y,z) + f^2(kKx+lLy+mMz) + \tfrac{1}{2}f^2(k^2K^2+l^2L^2+m^2M^2). \tag{3·12}$$

This definition retains the convexity of P. Now given a set of n values of piecewise constant data $\mathbf{X}_i$ with total area KLM, construct a set of 'ghost' values

$$(X_i+kK, Y_i+lL, Z_i+mM) \tag{3·13}$$

indexed by the integers (k,l,m) for each i. It is then shown that if $\mathbf{x}$ lies on the intersection of values $\mathbf{X}_i$ and $\mathbf{X}_j$, $\mathbf{x}+\mathbf{k}\cdot\mathbf{K}$ lies on the intersection of values $\mathbf{X}_i+\mathbf{k}\cdot\mathbf{K}$ and $\mathbf{X}_j+\mathbf{k}\cdot\mathbf{K}$. The construction is now carried out in a similar way to that in **CP**. The solution is obtained over a single 'copy' of the domain $(0\leqslant x\leqslant K, 0\leqslant y\leqslant L, 0\leqslant z\leqslant Z)$. It is found as the intersection of the tangent planes

$$s = \mathbf{x}\cdot\mathbf{X} + s_i, \quad 1\leqslant i\leqslant n, \tag{3·14}$$

where each plane has 'ghost' planes

$$s = \mathbf{x}.(X_i+kK, Y_i+lL, Z_i+mM) + s_i + \tfrac{1}{2}f^2(k^2K^2+l^2L^2+m^2M^2)$$

$$+ f^2(kKX_i+lL_{Yi}+mMZ_i). \tag{3·15}$$

It is then clear that as each s_i is increased, the total area of the ith plane and its 'ghosts' will be increased at the expense of its neighbours, so that the proof of **CP** can be applied. The construction is illustrated for one space dimension in Fig. 8. The extension to general data is made by the methods of section 2.

d. *Variable Coriolis parameter*

This is the most difficult of the extensions treated in this paper, because the rewriting of the equations in a form where all the ageostrophic circulation is absorbed into the D/Dt operator is only possible for constsnt f. This procedure is crucial in the theory. The extension is necessary if the theory is to be used to describe more than localised mesoscale phenomena over short time periods, and has therefore received much attention in the literature, especially Salmon (1985). It is first analysed assuming that the domain does not approach the equator. The application of the theory at the equator is described in the next section. At least four approaches are possible. Three introduce extra approximations as the price for obtaining equations to which some of the theory can be applied. The fourth does not, but little can be proved about the solutions.

Shutts (1989) extended the theory to spherical geometry by considering the sphere as embedded in three-dimensional Cartesian space and using a Hamiltonian formulation in which the gravitational force does not have to be parallel to the axis of rotation. The theory applies quite naturally to this configuration. The effect is that equation (2·2) is replaced by

$$D(v_g \sin\lambda)/Dt + f(u - u_g) = 0, \qquad (3\cdot 16)$$

where λ is the latitude. This approximation replaces the horizontal wind by its geostrophic component in a plane perpendicular to the axis of rotation. The absolute momentum can now be written as

$$\mathbf{M} = 2\Omega(\mathbf{r} - (\mathbf{\Omega}\cdot\mathbf{r})\mathbf{\Omega}/\Omega^2) - \mathbf{\Omega}\times\mathbf{v}/\Omega, \qquad (3\cdot 17)$$

where $\mathbf{r}$ is a position vector and $\mathbf{\Omega}$ is the angular velocity vector. It is then shown that the absolute momentum components are conserved if the pressure gradient terms are removed from (2·1) and (3·16), geostrophic coordinates can be defined, and the remainder of the theory presented in this paper follows through without alteration.

Salmon (1985) derived a set of equations very similar to the semi-geostrophic equations from Hamilton's principle. By introducing approximations directly into the Hamiltonian, he was able to ensure that simple Lagrangian equations of the form (2·11) and (2·12) were obtained. We apply his results to the incompressible equations in the (x, y) plane with f a function of x and y. The standard semi-geostrophic equations are then (2·1), (2·2) ,(2·4), and the (x,y) components of (2·5). Salmon's equations (3·12), (3·21), and (3·22) can be shown to reduce to

$$f\dot{X} = f^2(y-Y) - f|\mathbf{x}-\mathbf{X}|^2 f_Y \qquad (3\cdot18)$$

$$f\dot{Y} = f^2(X-x) + f|\mathbf{x}-\mathbf{X}|^2 f_X, \qquad (3\cdot19)$$

where f is defined as a function of X and Y rather than of x and y. $(f\dot{X}, f\dot{Y})$ is non-divergent and can be derived from a stream function Φ. The inverse potential vorticity ρ satisfies (Salmon's (3·30))

$$f\rho = \partial(\mathbf{x})/\partial(\mathbf{X}) \qquad (3\cdot20)$$

and the evolution equation

$$D_X\rho/Dt = 0. \qquad (3\cdot21)$$

If ρ is given as a function of $\mathbf{X}$, then (3·20) must be solved for $\mathbf{x}(\mathbf{X})$ with the boundary condition that $\mathbf{x}$ is contained in a specified region of physical space. It is not yet known whether this can be done. There is an extension to the geometric method which can treat this case, but it is not known whether it gives a unique solution. Salmon derives an elliptic equation (his (3·34)) which can clearly be solved uniquely by making further approximations involving neglecting derivatives of f in certain terms. A rather similar approximation can be made here by writing $\mathbf{x}$ as ∇R, so that

$$\partial x/\partial Y - \partial y/\partial X = 0. \qquad (3\cdot 22)$$

This then allows $\mathbf{x}(\mathbf{X})$ to be found by the method of section 2. (3·22) is not compatible with the condition that $(f\dot{X}, f\dot{Y})$ is non-divergent unless the spatial derivatives of f are neglected.

Once $\mathbf{x}(\mathbf{X})$ is known, the solution can be completed by noting that ρf satisfies an equation of the form (2·12)

$$\partial/\partial t(\rho f) + \nabla_x\cdot(\rho f\mathbf{U}) = 0, \qquad (3\cdot 23)$$

where $\mathbf{U}\equiv(\dot{X},\dot{Y})$ is defined by (3·18-19). This part of the problem is exactly the same as that posed in section 2, with ρf replacing ρ.

The other option is to solve (2·1) to (2·5) as they stand. Hoskins (1975) proved that they conserve the energy integral (2·16) even with variable f, and Cullen et al. (1987) showed that the energy is stationary under smooth local rearrangements which can be written

$$\delta u_g = f\delta y \qquad (3\cdot 24)$$

$$\delta v_g = -f\delta x. \qquad (3\cdot 25)$$

This stationary point is a minimum if the geostrophic potential vorticity is positive, since the variation of f is not relevant to local parcel stability. However, the potential vorticity is no longer conserved. The rearrangements (3·24-25) cannot be extended to global rearrangements because the changes to u_g and v_g will be dependent on the trajectory used. It is thus clear that if the solution implies explicit convection in the horizontal it will not be unique, because the end state values of (u_g, v_g) will depend on the trajectory followed in the convection. It will thus only be possible to prove that the solution is unique if the potential vorticity is positive and the equations can be solved for a bounded total velocity $\mathbf{u}$. We consider only this case.

There are two approaches. Much of the theory can still be applied if the equations are written in spherical polar coordinates in a way which allows them to be decoupled into zonal and meridional problems to which the theory can be applied separately. The zonal problem can be treated by the periodic f-plane theory, and the meridional problem by the axisymetric formulation of Shutts et al. (1988). The results can then be extended to the coupled system by an iterative argument. This argument also allows the axisymmetric solutions of Shutts et al. to be extended to non-axisymmetric flows. The solutions should be useful if the trajectories are approximately straight in a polar coordinate system. This makes their theory much more useful in studying, for instance, hurricane structure.

The equations are written in spherical polar coordinates with λ longitude, μ latitude, a the radius of the earth and Ω the earth's angular velocity. The region of interest is assumed to be bounded away from the pole, to avoid difficulties with the coordinate system, as well as away from the equator. In order to decouple the equations into zonal and meridional parts, it can be seen from Shutts (1980) that the correct form is

$$Du_g/Dt - (u_g v\tan\mu)/a + 2\Omega\sin\mu(v_g - v) = 0 \qquad (3\cdot26)$$

$$Dv_g/Dt + 2\Omega\sin\mu(u - u_g) = 0 \qquad (3\cdot27)$$

$$D\theta/Dt = 0 \qquad (3\cdot28)$$

$$DV/Dt = 0 \qquad (3\cdot29)$$

$$u_g^2\tan\mu + 2\Omega a u_g\sin\mu + \partial\phi/\partial\mu = 0 \qquad (3\cdot30)$$

$$2\Omega a v_g\sin\mu\cos\mu = \partial\phi/\partial\lambda \qquad (3\cdot31)$$

$$g\theta/\theta_0 = \partial\phi/\partial z \qquad (3\cdot32)$$

$$D/Dt \equiv \partial/\partial t + (u/a\cos\mu)\partial/\partial\lambda + (v/\alpha)\partial/\partial\mu + w\partial/\partial z. \qquad (3\cdot33)$$

Though this form of the equations appears closer to the standard geostrophic momentum approximation than those used by Shutts and Salmon, it requires trajectories to be approximately straight in the latitude-longitude plane, an unreasonable requirement near the poles. The systems used by Shutts and Salmon are both accurate at the poles.

In order to clarify the method of solution, temporarily write the continuity equation (3·29) in Eulerian form

$$\partial u/\partial\lambda + \partial/\partial\mu(v\cos\mu) + a\cos\mu\,\partial w/\partial z = 0. \tag{3·34}$$

(3·26-34) are now split into two separate problems

$$\mathbf{u}_1\cdot\nabla \equiv (u/a\cos\mu)\partial/\partial\lambda + w_1\partial/\partial z \tag{3·35}$$

$$\mathbf{u}_2\cdot\nabla \equiv (v/a)\partial/\partial\mu + w_2\partial/\partial z, \tag{3·36}$$

where

$$(1/a\cos\mu)\partial u/\partial\lambda + \partial w_1/\partial z = 0 \tag{3·37}$$

$$(1/a\cos\mu)\partial(v\cos\mu)/\partial\mu + \partial w_2/\partial z = 0, \tag{3·38}$$

$$\partial u_g/\partial t + \mathbf{u}_1\cdot\nabla u_g + \left\{\mathbf{u}_2\cdot\nabla u_g - (u_g v\tan\mu)/a + 2\Omega(v_g-v)\sin\mu\right\} = 0 \tag{3·39}$$

$$\partial v_g/\partial t + \{\mathbf{u}_1\cdot\nabla v_g + 2\Omega(u-u_g)\sin\mu\} + \boldsymbol{\theta}_2\cdot\nabla v_g = 0. \tag{3·40}$$

$$\partial\theta/\partial t + \mathbf{u}_1\cdot\nabla\theta + \mathbf{u}_2\cdot\nabla\theta = 0 \tag{3·41}$$

$$(u_g^2/a)\tan\mu + 2\Omega u_g\sin\mu + (1/a)\partial\phi/\partial\mu = 0 \tag{3·42}$$

$$fv_g = (1/a\cos\mu)\partial\phi/\partial\lambda \tag{3·43}$$

$$g\theta/\theta_0 = \partial\phi/\partial z. \tag{3·44}$$

The continuity equations (3·37) and (3·38) can be written as Lagrangian conservation laws for specific cross-sections in the zonal and meridional directions. The east-west problem consists of (3·35), (3·37), (3·43), (3·44) and the relevant parts of the evolution equations (3·39-41). This problem can be solved uniquely for each value of μ, since it is a set of f plane problems with periodic boundary conditions in λ. The north-south problem consists of (3·36), (3·38), (3·42), (3·44) and the second parts of (3·39-41). It can be reduced to the equations for axisymmetric flow on an f plane by the substitution $r=a\cos\mu$. Then, in particular, we can write the second part of (3·39) as an equation foe zonal angular momentum

$$M = u_g r + \Omega a r^2 \qquad (3\cdot45)$$

$$DM/Dt = 2\Omega a v_g r \sin\mu \qquad (3\cdot46)$$

and (3·38) as

$$D(rA)/Dt = 0 \qquad (3\cdot47)$$

where A is the specific meridional cross-section. (3·42) becomes

$$u_g^2/r + 2\Omega u_g + \partial\phi/\partial r = 0. \qquad (3\cdot48)$$

These equations were solved by Lagrangian construction by Shutts, Booth and Norbury (1988). They showed that the transformed radial coordinate $(1/r_o^2-1/r^2)$, where r_o is a minimum value of r, must be used. The problem cannot be solved in this coordinate system if the domain includes the poles.

The solution of the full problem (3·26-33) is now obtained by iteration. Starting with a first guess for **u**, the east-west problem is solved for (u,w_1) with (v,w_2) fixed and then the north-south problem is solved for (v,w_2) with (u,w_1) fixed. If we suppose that these partial problems are solved over a time interval Δt, solving the east west problem minimises the total energy (2·16) at the end of the time interval under rearrangements in the zonal direction, and solving the north-south problem minimises the energy under rearrangements in the meridional direction. Each step of the iteration therefore reduces the total energy. Since the energy that can be reached by such an iteration is bounded below by the rest state energy of the system, the iteration must converge. Any limit satisfies the full problem (3·26-33), and in particular has the same total energy as the initial state. This limit is therefore unique for the given iteration strategy. It is possible that a different solution could be obtained if the north-south problem was solved first. However, if this happens, the second solution must be reached from the first by a rearrangement of the fluid, since only **u** is being iterated. If

$$\mathbf{Q} = \begin{pmatrix} \partial v_g/\partial\lambda + 2a\Omega\sin\mu\cos\mu & \partial v_g/\partial\mu & \partial v_g/\partial z \\ -\partial u_g/\partial\lambda & 2\Omega a\sin\mu + (u_g/a)\tan\mu - \partial u_g/\partial\mu & \partial u_g/\partial z \\ \partial\theta/\partial\lambda & \partial\theta/\partial\mu & \partial\theta/\partial z \end{pmatrix} \quad (3\cdot 49)$$

has strictly positive eigenvalues, giving strict local parcel stability everywhere, then any perturbation by rearrangement of the first solution must have more energy. Thus any second solution must be separated by a finite difference from the first and cannot bifurcate from the first. The initial-value problem therefore has a unique solution as long as the condition on (3·49) is satisfied everywhere. Since there is no potential vorticity conservation law for this system, we cannot be sure that (3·49) will remain positive definite for all time, even if it is for the initial data. However, the arguments of Shutts and Cullen (1987) suggest that any breakdown will occur in a parameter range where the semi-geostrophic approximation is not accurate.

It is important to note that this method of solution only works because f varies with latitude only and is axially symmetric. It would not work if f was an arbitrary function of position. This method can also be used to generalise the solutions of Shutts et al. (1988) for axisymmetric vortices to non-axisymmetric flow. Its accuracy depends on the trajectories being approximately straight in these coordinates, including the case of almost axisymmetric flow. Because of the loss of potential vorticity conservation and the indirect nature of the proofs, this method is suited to analytic solutions.

The problem at the pole is artificially created by the need to decompose the equations into two tractable sub-problems. If the geostrophic momentum approximation is applied directly in spherical polar coordinates, an extra term $(uu_g \tan\mu)/a$ appears on the left of (3·27) and the term $u_g^2 \tan\mu$ is omitted from (3·30). The term $-(u_g v \tan\mu)/a$ in (3·26) becomes $-(uv_g \tan\mu)/a$. The resulting equations conserve energy, and the identification of the solution with a minimum energy state still holds. The equations are very similar to (3·26-30) away from the poles, and are very close to f-plane equations at the poles. This suggests that the same qualitative properties of the solutions apply, though there is no easy way of proving it.

In terms of physical applicability, the loss of the global rearrangement property in (3·26-30) or the alternative discussed in the preceding paragraph only matters if there are regions of zero potential vorticity, slantwise convection, or mountain blocking effects, extending for a latitudinal distance comparable to the scale on which the Coriolis parameter varies. In order to make the solution unique, it is necessary to require any convective jump to take place by a particular route so that the dependence of the change in u_g on the trajectory can be allowed for. This route must be chosen on physical considerations outside the scope of semi-geostrophic theory.

e. *Behaviour at the equator*

While the geostrophic relation no longer gives a useful relation between height and wind fields near the equator, any theory which is applied to the planetary scale atmospheric circulation should, to be useful, be capable of giving at least a simplified description of the equatorial flow, if only to act as a proper boundary condition for the extra-tropics. Similarly, if a 'slow manifold' definition is to be useful in initialising data for numerical models, it must be capable of being used at the equator, even if the solution is oversimplified. Cullen et al. (1987) show that the interpretation of the semi-geostrophic equations as 'quasi-equilibrium' equations evolving through minimum energy states makes sense at the equator, since the concept of minimising the energy is just as applicable in non-rotating as in rotationg fluids. In this section we illustrate the behaviour of the solutions near the equator. A full existence and uniqueness proof for the global problem is, however, not yet possible.

The equatorial behaviour is illustrated by describing two simple problems, one in a (y,z) cross-equatorial cross-section and the other in a cross-section along the equator. The unmodified semi-geostrophic equations (2·1) to (2·5) are used. The Salmon method has not yet been extended to cross-equatorial flow. The behaviour of the system studied by Shutts (1988) is significantly different, and allows Kelvin waves at the equator. It is not yet possible to assess the usefulness of his system as compared to (2·1) to (2·5).

First consider flow in an (x,z) cross-section along the equator. Periodic boundary conditions are used, and the flow is driven by an atmospheric heat source such as latent heat release. The equations are

$$D\theta/Dt = Q \tag{3·50}$$

$$DV/Dt = 0 \tag{3·51}$$

$$(\partial\phi/\partial x, \partial\phi/\partial z) = (0, g\theta/\theta_0), \tag{3·52}$$

where V is the specific cross-section. The requirement of no horizontal gradient means that the solution is

$$w = (Q - \bar{Q}^{x})/(\partial\theta/\partial z), \tag{3·53}$$

with u given by (3·51). If there is topography present, the average over x extends only between topographic barriers at the same level in z, Fig. 9. Note that there is no way of distinguishing air parcels at the same z but different x. The zonal mean u is undefined, except below the height of the highest mountain barrier.

Now consider a cross-equatorial cross-section, bounded by walls at some latitude either side of the equator. If u_g satisfies

$$u_g + \Omega a\cos\lambda = 0, \tag{3·54}$$

then the associated absolute vorticity is zero, the momentum equation becomes trivial and the problem reduces to (3·50-52) solved in (y,z) rather than (x,z). In general cross-equatorial flow will result. If u_g differs from the value given by (3·54) by other than a constant, then the absolute vorticity component will become non-zero. The inertial stability condition requires it to be positive in the Northern hemisphere and negative in the Southern hemisphere. Cross-equatorial advection will result in the condition being violated and horizontal 'convective' inertial interchanges will result. These interchanges represent advection on time-scales faster than f^{-1} which cannot be 'seen' by semi-geostrophic dynamics. Thus a solution will be obtained and the mass transport across the equator will be well-defined. The cross-equatorial velocity field will not be easy to calculate. It is likely that energy will be lost in this jump, as it is when the jump results from orographic blocking or penetrative convection, Shutts (1987a,b).

Any practical solution of these equations by other than the geometric construction method is likely to require addition of some background friction to enable solutions to be obtained. The resulting system of equations then contains the simple systems used to describe the basic cross-equatorial monsoon circulation, for instance that used by Sashegyi and Geisler (1987). This solution depends critically on the presence of friction. Further study is required to see if the mass transport implied by this solution converges to that given by the geometric method as the friction tends to zero.

The solution procedure given above should give a unique solution in three dimensions. Even using the formulation of Shutts, (3·16), it cannot be found by global rearrangement arguments because parcels in the two hemispheres with equal angular momentum and potential temperature are interchangeable.

f. *Inclusion of forcing terms*

We analyse the effect of forcing terms for the simplest case of the bounded f plane problem (2·1) to (2·5). The analysis can then be extended to the other cases treated in previous sections. The equations are now

$$Du_g/Dt + f(v_g-v) = A \tag{3·55}$$

$$Dv_g/Dt + f(u-u_g) = B \tag{3·56}$$

$$D\theta/Dt = C, \tag{3·57}$$

together with (2·4) and (2·5). A,B and C are assumed to be prescribed forcing functions of $\mathbf{x},\mathbf{X}$ and t, but not of $\mathbf{u}$. In terms of the geostrophic coordinates, the equations are

$$DX/Dt + f(Y-y) = B/f \tag{3·58}$$

$$DY/Dt + F(x-X) = -A/f \tag{3·59}$$

$$DZ/Dt = C. \tag{3·60}$$

(2·12) still holds, but (2·11) becomes

$$\mathbf{U} = (f(y-Y)+B/f, f(X-x)-A/f, C) \tag{3·61}$$

and $\nabla_x \cdot \mathbf{U}$ is not necessarily zero. The inverse potential vorticity ρ is no longer a Lagrangian conservation property. The definition $\rho=\partial(\mathbf{x})/\partial(\mathbf{X})$ is unaffected, and the solution procedure of section 2 can be followed provided (2·12) can be solved. The difficult case is where zero potential vorticity is generated by the forcing, for instance by heating from below.

(2·12) can be described in terms of displacements of ρ in the sense of section 2 (Theorem 2) as $\mathbf{U}$ is a bounded function of $\mathbf{x}$ and $\mathbf{X}$. Theorem 5 therefore proves that (3·58-60) can be solved for the evolution of R and hence the geostrophic variables $\mathbf{X}$. The total physical velocity $\mathbf{u}$ will cease to be well defined if zero potential vorticity regions are generated.

4. Discussion

The results contained in this paper show that the Lagrangian form of the semi-geostrophic equations form a 'slow manifold', in the sense described in the introduction. They do not show how accurate this model is in describing real atmospheric behaviour. No equivalent results have been proved for other filtered sets of equations such as the nonlinear balance equations, it may be important to attempt to do so. The equations can be solved in principle for general smooth or piecewise smooth data, for free-surface and periodic boundary conditions, and for variable Coriolis parameter. In some cases the method may produce discontinuous solutions from smooth data. The equations can be integrated forward for arbitrary finite times and produce a solution which conserves the initial energy. It is thus possible to construct a model of the adiabatic atmospheric circulation without mountains which conserves energy. Shutts (1987a and b) and Cullen (1988) show that energy will be lost if there is orographic blocking, penetrative convection, and surface friction.

Further research is needed to determine which, if any, of the versions of the theory described here is most useful for describing planetary scale circulations. Further work is also needed to prove that solutions involving mountains, penetrative convection, or cross-equatorial transport are well-defined in the sense that the implied mass transports can be uniquely calculated and the geostrophic fields predicted. Such an extension is important, since if these mass transports are really determined by the larger scale flow in this fashion, the overall predictability of the flow will be much greater than if the transports depend on local effects.

These results also prove that the geometric construction method and an alternating direction method converge to the desired solutions as they are refined. Such proofs do not indicate whether these methods will converge fast enough to be of practical use for calculations, though they will produce reliable benchmark solutions.

ACKNOWLEDGMENTS

The authors would like to thank Dr. D.G.Andrews for bringing the work by Pogorelov to their attention, Dr. G.J.Shutts for a helpful review of the manuscript, and Mr. P.Cloke for assistance in the evaluation of Salmon's work,

REFERENCES

Bennett, A.F. and P.E.Kloeden, 1981: The quasi-geostrophic equations: approximation, predictability and equilibrium spectra of solutions. *Quart. J. Roy. Meteor. Soc.*, **107**, 121-136.

Chynoweth, S., 1987: *The semi-geostrophic equations and the Legendre transform*. Ph. D thesis, Dept. Maths., Univ. of Reading, U.K., 248pp.

Cullen, M.J.P., 1989: Implicit finite difference methods for modelling discontinuous atmospheric flows. *J. Comput. Phys.*, to appear.

——, and R.J.Purser, 1984: An extended Lagrangian theory of semi-geostrophic frontogenesis. *J. Atmos. Sci.*, **41**, 1477-1497.

——, J.Norbury, R.J.Purser and G.J.Shutts, 1987: Modelling the quasi-equilibrium dynamics of the atmosphere. *Quart. J. Roy. Meteor. Soc.*, **113**, 735-757.

——, J.Norbury and R.J.Purser, 1988: Generalised Lagrangian solutions for atmospheric and oceanic flows. Met O 11 Sci. Note no. 11, Met. Office, Bracknell, U.K., 27pp.(Submitted to SIAM J. Appl. Math.)

Gent, P.R. and J.C.McWilliams, 1983: Consistent balanced models in bounded and periodic domains, *Dyn. Atmos. Oceans*, **7**, 67-93.

Hoskins, B.J., 1975: The geostrophic momentum approximation and the semi-geostrophic equations. *J. Atmos. Sci.*, **32**, 233-242.

——, and F.P.Bretherton 1972: Atmospheric frontogenesis models: Mathematical formulation and solution. *J. Atmos. Sci.*, **29**, 11-37.

——, and I. Draghici, 1977: The forcing of ageostrophic motion according to the semigeostrophic equations and in an isentropic coordinate model. *J. Atmos. Sci.*, **34**, 1858-1867.

———, and N.V. West, 1979: Baroclinic waves and frontogenesis. Part II. Uniform potential vorticity jet flows in warm and cold fronts. *J. Atmos. Sci.*, **40**, 1595-1612.

———, M.E.McIntyre and A.W.Robertson: On the use and significance of isentropic potential vorticity maps. *Quart. J. Roy. Meteor. Soc.*, **111**, 877-946.

Leith,C.E., 1980: Nonlinear normal-mode initialisation and quasi-geostrophic theory. *J. Atmos. Sci.*, **37**, 958-968.

Pogorelov, A.V., 1964. *Monge-Ampere equations of elliptic type*. Noordhoff, Groningen, 114pp.

Purser, R.J. and M.J.P.Cullen, 1987: A duality principle in semi-geostrophic theory. *J. Atmos. Sci.*, **44**, 3449-3468.

Rockafellar, R.T., 1970. *Convex analysis*. Princeton University Press. 451pp.

Salmon, R., 1985: New equations for nearly geostrophic flow. *J. Fluid Mech.*, **153**, 461-477.

Shutts, G.J., 1980: Angular momentum coordinates and their use in zonal, geostrophic motion on a hemisphere. *J. Atmos. Sci.*, **37**, 1126-1132.

———, 1987a: The semi-geostrophic weir: a simple model of flow over mountain barriers. *J. Atmos. Sci.*, **44**, 2018-2030.

———, 1987b: Balanced flow states arising from penetrative slantwise convection. *J. Atmos. Sci.*, **44**, 3363-3376.

———, 1988: Planetary semi-geostrophic equations derived from Hamilton's principle, Met O 11 Sci. Note no. 9, Met. Office, Bracknell, U.K., 69pp. (Submitted to J. Fluid Mech.)

———, and M.J.P.Cullen, 1987: Parcel stability and its relation to semi-geostrophic theory. *J. Atmos. Sci.*, **44**, 1318-1330.

———, M.M.Booth and J.Norbury, 1988: A geometric model of balanced, atmospheric flows with embedded penetrative convection. *J. Atmos. Sci.*, **45**, 2609-2621.

Simmons, A.J. and B.J.Hoskins, 1978: The life cycles of some nonlinear baroclinic waves. *J. Atmos. Sci.*, **35**, 414-432.

Vautard, R. and B. Legras, 1986: Invariant manifolds, quasi-geostrophy, and initialisation. *J. Atmos. Sci.*, **43**, 565-584.

FIGURE CAPTIONS

Fig. 1 The limit cone of a convex curve at a point O.

Fig. 2 The construction which determines the curvature at O.

Fig. 3 The construction which determines the motion of the point A.

Fig. 4 Contours of P within Ω, showing that the extreme value of ∇P occurs on the boundary near where ∇P is normal to the boundary.

Fig. 5 The construction which illustrates the motion of an isolated large value of $\mathbf{X}$.

Fig. 6 The free surface construction with two elements of area μ_1 and μ_2.

Fig. 7 The free surface construction for an (x,z) cross-section. (a) Basic cylinder. (b) The intersection of one plane with the cylinder.

Fig. 8 The construction of periodic solutions.

Fig. 9 Horizontal averaging in the equatorial solution.

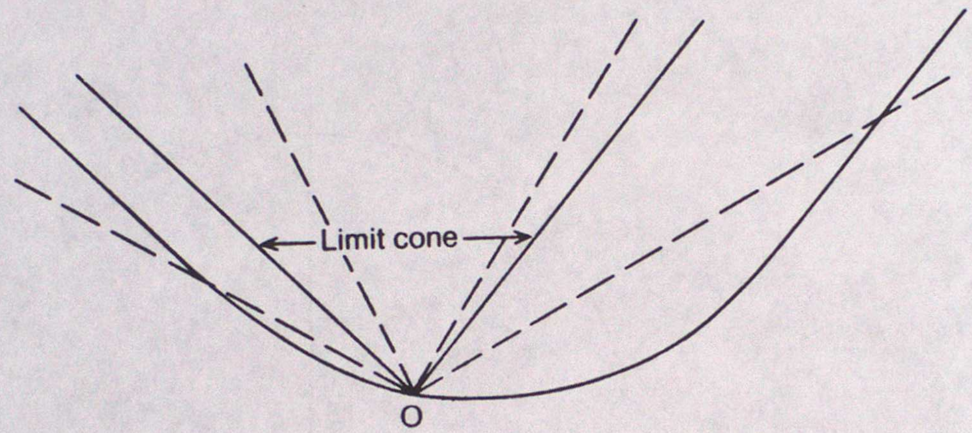

Fig. 1 The limit cone of a convex curve at a point O.

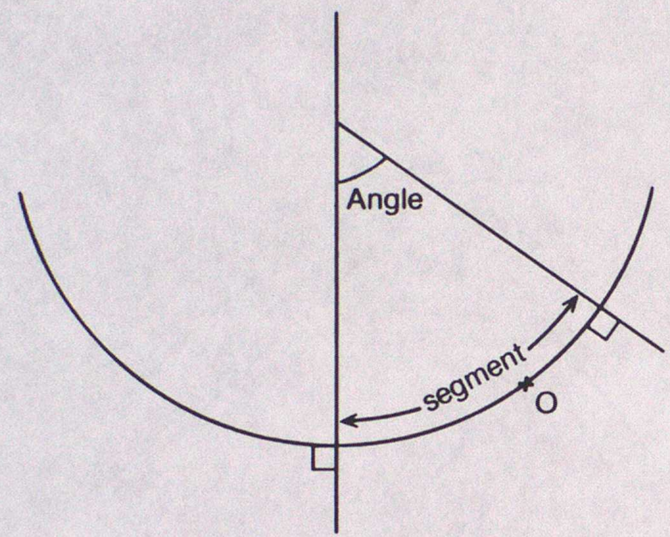

Fig. 2 The construction which determines the curvature at O.

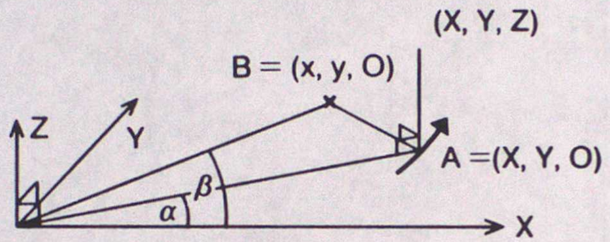

Fig. 3 The construction which determines the motion of the point A.

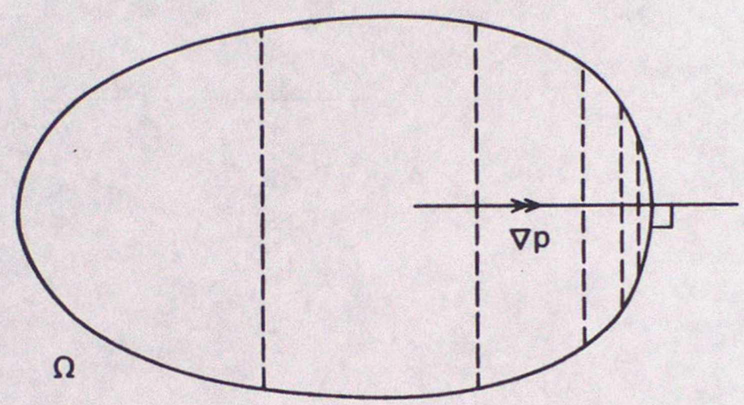

Fig. 4 Contours of P within Ω, showing that the extreme value of ∇P occurs on the boundary near where ∇P is normal to the boundary.

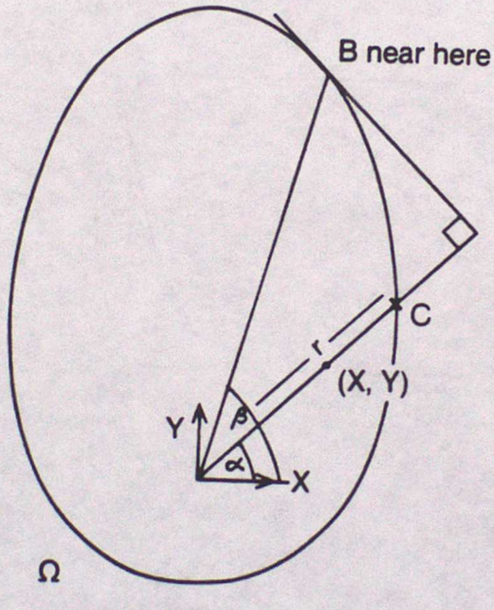

Fig.5 The construction which illustrates the motion of an isolated large value of X.

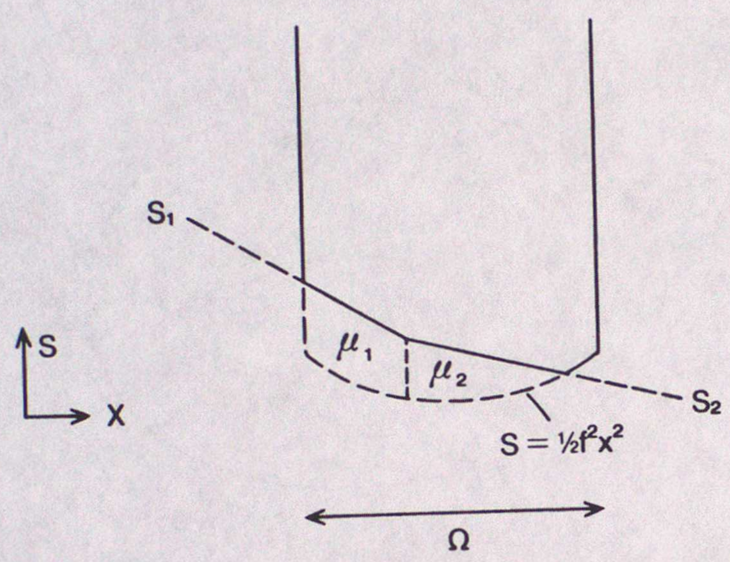

Fig. 6 The free surface construction with two elements of area μ_1 and μ_2.

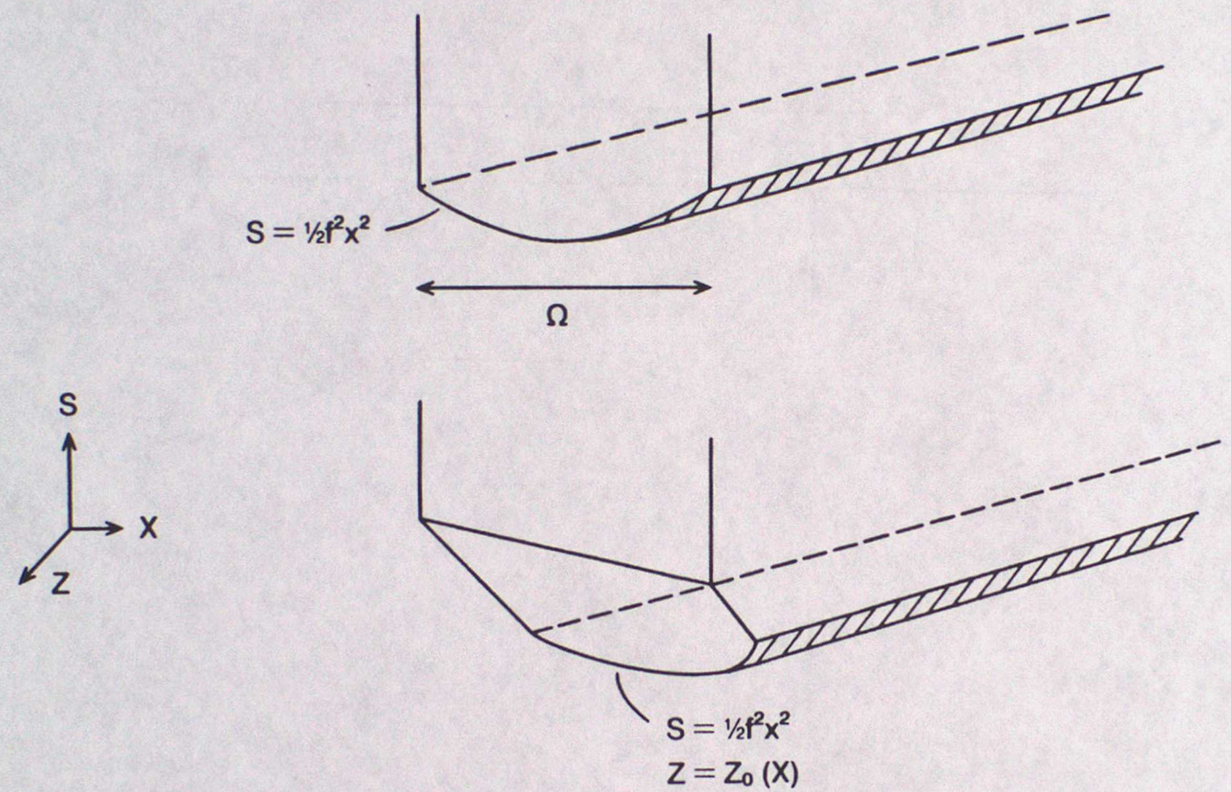

Fig. 7 The free surface construction for an (x, z) cross-section. (a) Basic cylinder. (b) The intersection of one plane with the cylinder.

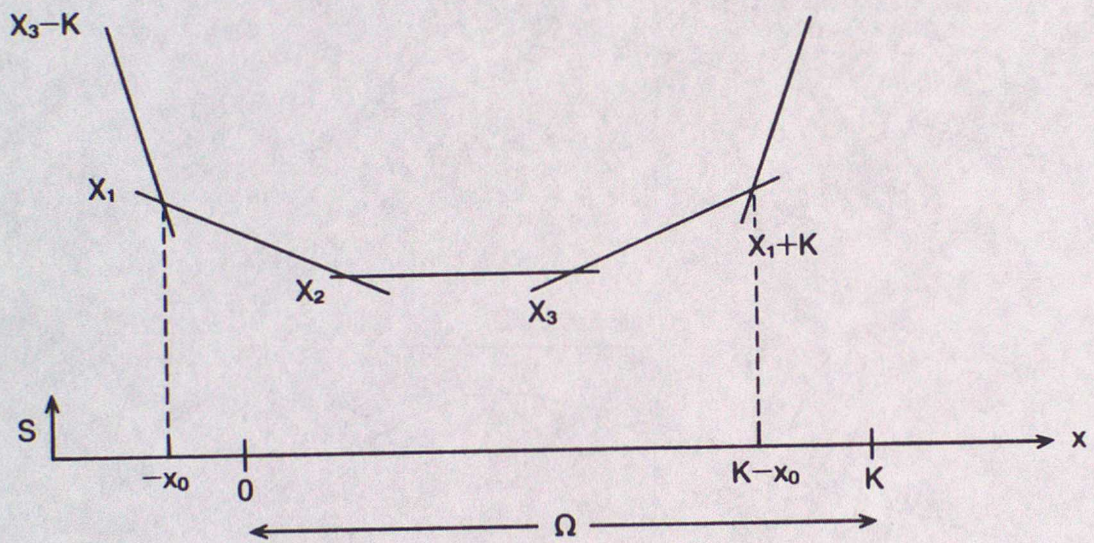

Fig. 8 The construction of periodic solutions.

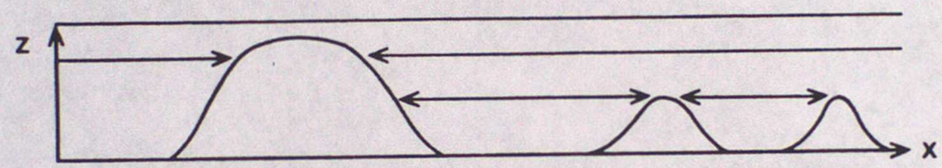

Fig. 9 Horizontal averaging in the equatorial solution.

MET O 11 SCIENTIFIC NOTES

1. **The theory of periodic solutions of the semi-geostrophic equations.**
   R.J. Purser
   October 1987

2. **Properties of the partial differential equations governing various types of atmospheric motions and implications for numerical methods.**
   M.J.P. Cullen
   December 1987

3. **A geometric model of balanced axisymmetric flows with embedded penetrative convection.**
   G.J. Shutts, M. Booth and J. Norbury
   February 1988

4. **Implicit finite difference methods for computing discontinuous atmospheric flows.**
   M.J.P. Cullen
   March 1988

5. **Variational aspects of semi-geostrophic theory.**
   R.J. Purser
   August 1988

6. **On the interpretation of boundary layer effects into a balanced model.**
   M.J.P. Cullen
   June 1988

7. **Implicit finite difference methods for modelling discontinuous atmospheric flows.**
   M.J.P. Cullen
   June 1988

8. **An analytical model of the growth of a frontal depression.**
   M.W. Holt and G.J. Shutts
   November 1988

9. **Planetary semi-geostrophic equation derived from Hamilton's principle.**
   G.J. Shutts
   August 1988

10. **Semigeostrophic moist frontogenesis in a Lagrangian model.**
    M.W. Holt
    October 1988

11. **Generalised Lagrangian solutions for atmospheric and oceanic flows.**
    M.J.P. Cullen, J. Norbury and R.J. Purser
    November 1988

12. **Properties of the Lagrangian semi-geostrophic equations.**
    M.J.P. Cullen and R.J. Purser
    January 1989