INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE: CENTRATOM TRIESTE

SMR/534-3

ICTP/WMO WORKSHOP ON EXTRA-TROPICAL AND TROPICAL
LIMITED AREA MODELLING
22 October - 3 November1990

"Initialisation"

by

W. Wergen
revised by W.A. Heckley

Presented by:
Prof. S. TIBALDI
Institute of Physics
Univ. of Bologna
Bologna, Italy

Please note: These are preliminary notes intended for internal distribution only.

Meteorological Training Course

Lecture Series

Initialisation

by

W. Wergen
revised by W.A. Heckley

April 1987

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

CONTENTS

1. INTRODUCTION

Primitive equation models, unlike quasi-geostrophic models, generally admit high frequency gravity wave solutions, as well as the slower moving Rossby wave solutions. If the results of the analysis scheme are used directly as initial conditions for a forecast, subtle imbalances between the mass and wind fields will cause the forecast to be contaminated by spurious high frequency inertial-gravity wave oscillations of much larger amplitude than are observed in the real atmosphere. The unrealistic initial amplitudes of these waves are caused by:

(a) erroneous observations not detected by data checking.

(b) simplified constraints imposed on the mass and wind analysis.

(c) various approximations in the model formulation.

Although these oscillations tend to die away slowly due to various dissipation mechanisms in the model, they make the forecast noisy and they may be quite detrimental to the analysis cycle, in which the six-hour forecast is used as a first-guess field for the next analysis. The synoptic changes over the six-hour period may be swamped by spurious changes due to the oscillations, with the consequence that at the next analysis time, good data may be rejected as being too different from the first-guess field. For this reason, an initialization step is performed between the analysis and the forecast, with the object of eliminating the spurious oscillations.

Before discussing how the initialisation is actually performed, it is worth illustrating the effects of initialisation. Fig. 1 shows how the surface pressure at a particular point (0°E, 52°N) exhibits high frequency

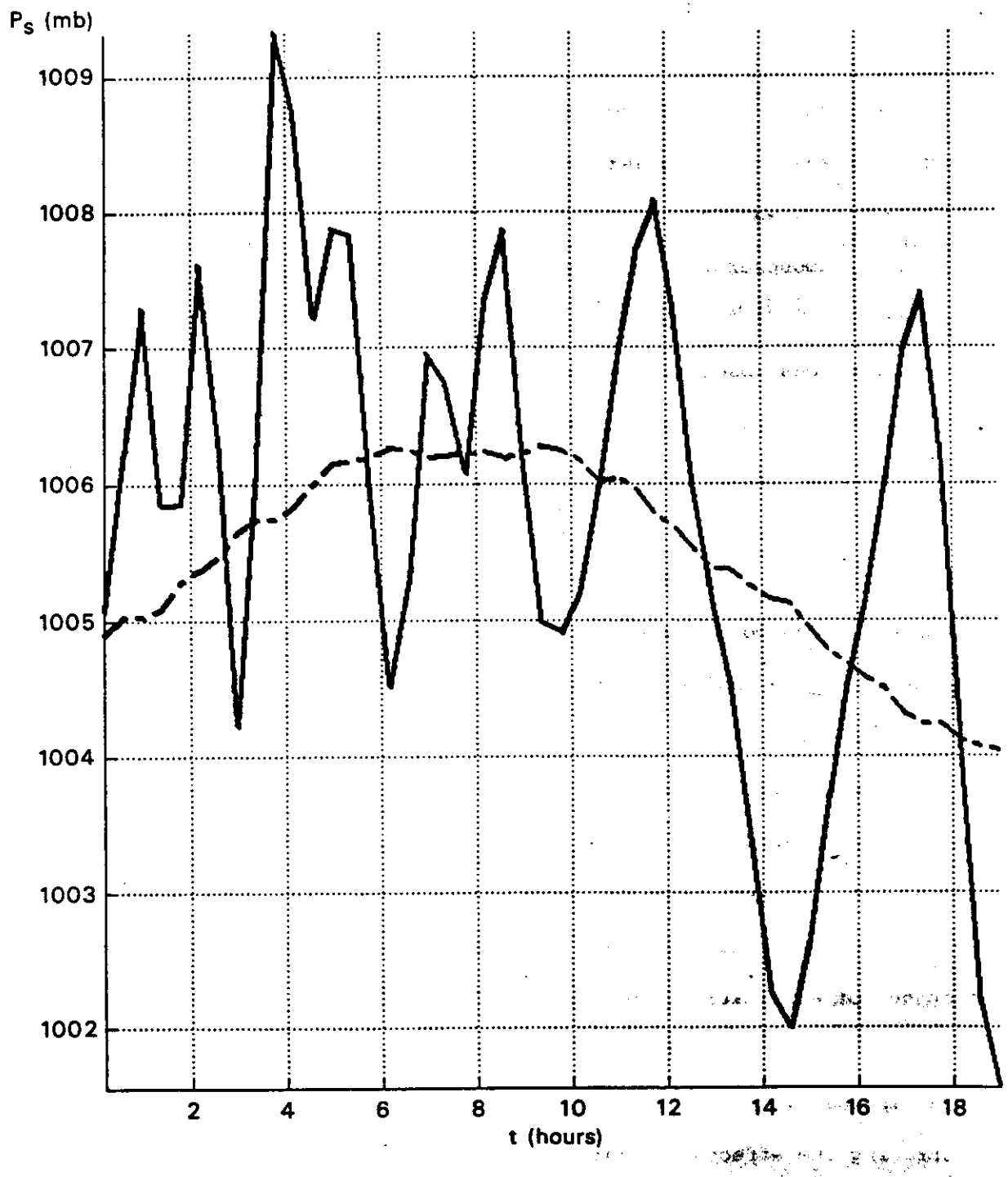


Fig. 1 Surface pressure evolution at gridpoint 0°E, 52°N for forecast started from uninitialised (full) and initialised (dashed) analysis for 12 GMT 6 September 1982.

oscillations with considerable amplitude. By contrast, if a forecast is started from an initialised analysis, the behaviour is much better.

As explained earlier, spurious oscillations in the first guess can affect whether observations are accepted or rejected for use in the analysis. To illustrate this Fig. 2 shows the 850 hPa observations superimposed on the first guess (a 6 hour forecast from the previous initialised analysis). When these are used in the analysis the two encircled wind observations are rejected. However when a first guess derived from an uninitialised analysis was used, these two observations were accepted though the winds are clearly in error.

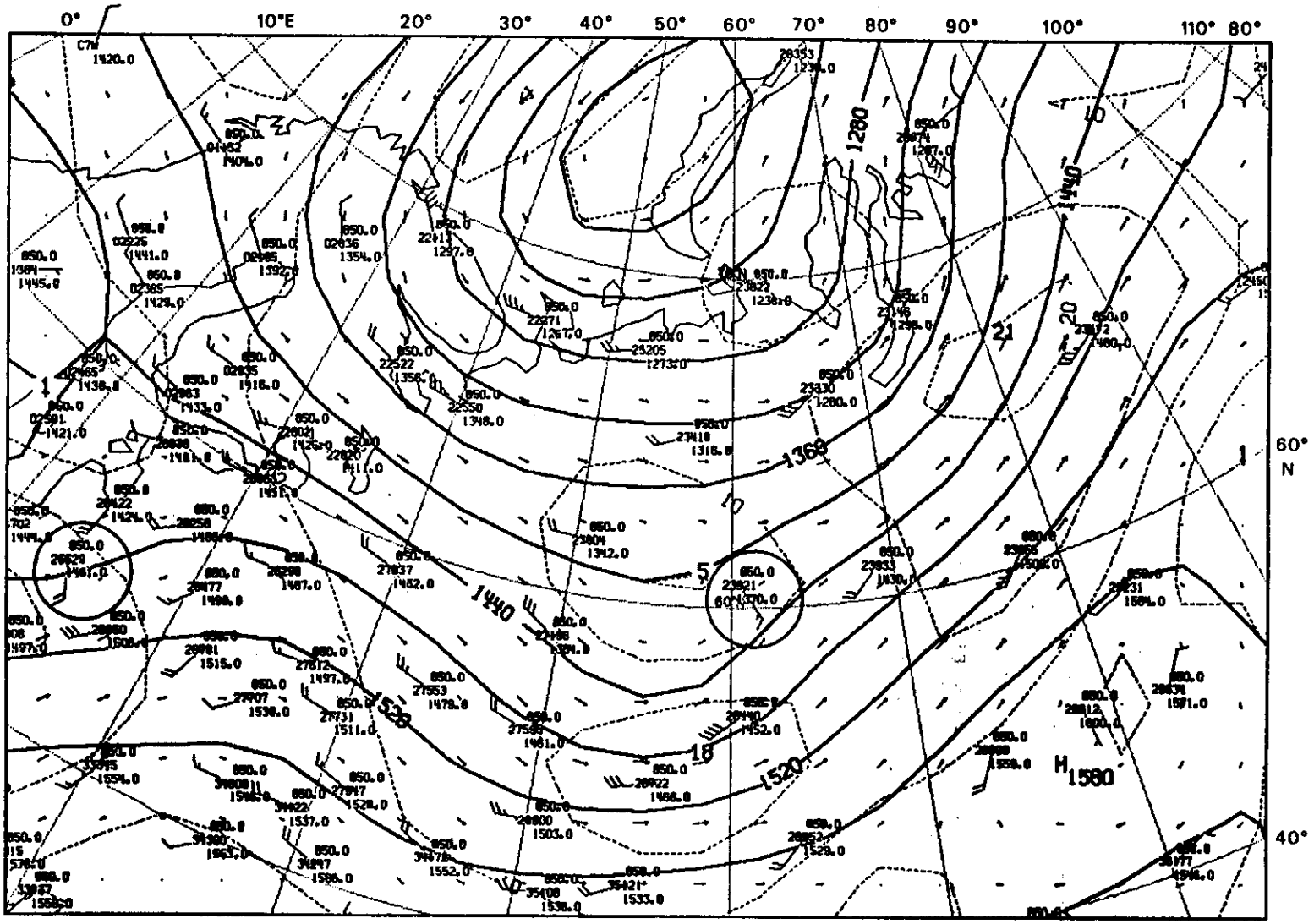


Fig. 2 850 mb observations and first guess height (full) and wind field. Contour interval for isotachs (dashed) is 10 m/s. Circles indicate differently treated observations.

2. INITIALISATION OF A SIMPLE SHALLOW WATER MODEL

In this Section the bi-periodic shallow water model on an f-plane is used in order to introduce the concepts of non-linear normal mode initialisation. This type of model allows a straightforward analytical treatment and is at the same time still relevant for more complex models.

The model equations are:

$$\frac{\partial u}{\partial t} - f_o v + \frac{\partial \phi'}{\partial x} = - u\frac{\partial u}{\partial x} - v\frac{\partial u}{\partial y} + F_x = N_u \qquad (2.1)$$

$$\frac{\partial v}{\partial t} + f_o u + \frac{\partial \phi'}{\partial y} = - u\frac{\partial v}{\partial x} - v\frac{\partial v}{\partial y} + F_y = N_v \qquad (2.2)$$

$$\frac{\partial \phi'}{\partial t} + \bar{\phi}\,(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}) = - \phi'\,(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}) - u\frac{\partial \phi'}{\partial x} - v\frac{\partial \phi'}{\partial y} + Q = N_\phi \qquad (2.3)$$

These equations describe the motions of an incompressible fluid with a free surface. They can be derived from the primitive equations by assuming that the density ρ depends only upon the vertical coordinate and that the initial wind is constant with height. ϕ, the geopotential of the free surface, has been split into a constant part $\bar{\phi}$ and a deviation ϕ'. The velocity components in the x and y directions are u and v, and f_o is the (constant) Coriolis parameter. Frictional effects are given by F_x and F_y, and Q is a heating term. The non-linear terms on the right hand side of (2.1), (2.2) and (2.3) have been combined and are represented by N_u, N_v and N_ϕ.

2.1 DERIVATION OF THE NORMAL MODES

Normal modes are the free motions of a system which is capable of vibrating. A typical example is a guitar string; once it has been excited it starts to vibrate in a way characteristic of that particular string. These vibrations

are called "own" or "eigen" vibrations, or in mathematical terms "normal modes".

In the following Section the normal modes of (2.1) - (2.3) are derived and discussed.

As a bi-periodic domain has been assumed, one may use double Fourier series to expand the dependent variable u, v, ϕ' and (symbolically) the non-linear terms N.

$$\begin{bmatrix} u \\ v \\ \phi' \\ N \end{bmatrix} = \sum_{n=-M}^{M} \sum_{m=-M}^{M} \begin{bmatrix} u_n^m \\ iv_n^m \\ \sqrt{\bar{\phi}}\ \phi_n^m \\ N_n^m \end{bmatrix} e^{i\left(\frac{2\pi mx}{L_x} + \frac{2\pi ny}{L_y}\right)} \qquad (2.4)$$

M is the maximum zonal and meridional number of waves, with L_x and L_y the zonal and meridional extent of the domain. The factor $i = \sqrt{-1}$ in front of the Fourier coefficient v_n^m results in a phase shift of a quarter of a wavelength; also the factor $\sqrt{\bar{\phi}}$ in front ϕ_n^m is required later on for consistency of dimensions. Both factors simplify the subsequent analyses.

Inserting (2.4) into (2.1) - (2.3) yields

$$\frac{\partial}{\partial t} u_n^m = f_o iv_n^m - ik\sqrt{\bar{\phi}}\,\phi_n^m + N_{un}^m \qquad (2.5)$$

$$\frac{\partial}{\partial t} iv_n^m = -f_o u_n^m - il\sqrt{\bar{\phi}}\,\phi_n^m + N_{vn}^m \qquad (2.6)$$

$$\frac{\partial}{\partial t} \sqrt{\bar{\phi}}\,\phi_n^m = -\bar{\phi}\, iku_n^m - \bar{\phi} i^2 lv_n^m + N_{\phi n}^m \qquad (2.7)$$

Here, $k=2\pi m/L_x$ and $l=2\pi n/L_y$ are the zonal and meridional wavenumbers. Now define a vector $\underline{X}_n^m$ as

$$\underline{X}_n^m = \begin{bmatrix} u_n^m \\ v_n^m \\ \phi_n^m \end{bmatrix} \tag{2.8}$$

Multiplying (2.6) by -i and dividing (2.7) by $\sqrt{\bar{\phi}}$ gives a system of equations which may be written in matrix form as

$$\frac{\partial}{\partial t}\underline{X}_n^m = i\underline{\underline{A}}_n^m \underline{X}_n^m + \underline{N}_n^m \tag{2.9}$$

where $\underline{N}_n^m$ is the vector of the non-linear terms

$$\underline{N}_n^m = \begin{bmatrix} N_{un}^m \\ -iN_{vn}^m \\ (\bar{\phi})^{-\frac{1}{2}} N_{\phi n}^m \end{bmatrix} \tag{2.10}$$

In the following, the reference to the zonal and meridional components m and n will be dropped. However, one has to keep in mind that there are different matrices and vectors for every m and n. The matrix $\underline{\underline{A}}$ is given by:

$$\underline{\underline{A}} = \begin{bmatrix} 0 & f & -k\sqrt{\bar{\phi}} \\ f_o & 0 & il\sqrt{\bar{\phi}} \\ -k\sqrt{\bar{\phi}} & -il\sqrt{\bar{\phi}} & 0 \end{bmatrix} \tag{2.11}$$

It is now possible to find the normal modes (eigensolutions of the linearised version ($\underline{N}_n^m = 0$) of (2.9)

First one has to find to the eigenvalues of $\underline{\underline{A}}$. As $\underline{\underline{A}}$ is a Hermitian matrix (rows identical to complex conjugate columns) only real eigenvalues are expected λ_j. They can easily be computed.

$$\lambda_1 = 0 \tag{2.12}$$

$$\lambda_{2,3} = \pm[f_o^2 + \bar{\phi}\,(k^2+\ell^2)]^{1/2} = \pm\sigma \tag{2.13}$$

$\lambda_{2,3}$ are the well known frequencies for inertia-gravity waves; λ_1 is the frequency of the Rossby waves which is stationary in this case because a constant f_o has been assumed. The modulus of the phase velocity for the inertia gravity waves is given by

$$|c| = \frac{\sigma}{\sqrt{k^2+\ell^2}} = \left[\bar{\phi} + \frac{f_o^2}{k^2 + \ell^2}\right]^{1/2} . \tag{2.14}$$

The phase velocity for inertia-gravity waves depends upon the wavenumbers k and l (i.e. the waves are dispersive). Thus the waves can distribute an initially locally confined quantity over a larger area.

Now the eigenvectors (normal modes) of $\underline{\underline{A}}$ are derived by solving

$$\underline{\underline{A}}\,\underline{v}_j = \lambda_j \underline{v}_j \quad j = 1,2,3 \tag{2.15}$$

for the eigenvectors $\underline{v}_j$ (these are column matrices).

Using λ_1 one may find the corresponding eigenvector $\underline{v}_1$ (now denoted by $\underline{R}$),

$$\underline{R} = \begin{bmatrix} -il\sqrt{\bar{\phi}}/f_o \\ k\sqrt{\bar{\phi}}/f_o \\ 1 \end{bmatrix} \tag{2.16}$$

This is the so called Rossby mode. It is immediately evident that it establishes, at least in this model, a geostrophic coupling between the geopotential (here arbitrarily scaled to 1) and the wind components.

For the two gravity modes (j=2,3) the eigenvectors may be written as

$$\underline{G}_{E,W} = \begin{bmatrix} if_o l \mp \sigma k \\ -f_o k \pm i\sigma l \\ \sqrt{\bar{\phi}}(k^2+l^2) \end{bmatrix} \qquad (2.17)$$

The upper sign in front of σ gives the eastward travelling gravity mode and the lower sign the westward one. As for the Rossby mode, the gravity modes are also characterised by a specific scale dependent coupling between the mass and wind fields.

Problem 1

Make a schematic sketch of

(i) a Rossby mode for m=n=1

(ii) a gravity mode for m=o, n=1

2.2 PROJECTION ON TO NORMAL MODES

An important property of normal modes is their orthogonality; it can also be shown that they form a complete set. Therefore any vector $\underline{X}$ may be expanded in terms of a series of normal modes in the same way as a field can be expanded as a Fourier series. For this purpose a modal matrix $\underline{\underline{E}}$ is defined, which has the 3 eigenvectors as its columns

$$\underline{\underline{E}} = \frac{1}{\sigma} \begin{bmatrix} -il\sqrt{\bar{\phi}} & \dfrac{if_o l - \sigma k}{\sqrt{2(k^2+l^2)}} & \dfrac{if_o l + \sigma k}{\sqrt{2(k^2+l^2)}} \\ k\sqrt{\bar{\phi}} & \dfrac{-f_o k + i\sigma l}{\sqrt{2(k^2+l^2)}} & \dfrac{-f_o k - i\sigma l}{\sqrt{2(k^2+l^2)}} \\ f_o & \dfrac{\sqrt{\bar{\phi}}(k^2+l^2)}{\sqrt{2(k^2+l^2)}} & \dfrac{\sqrt{\bar{\phi}}(k^2+l^2)}{\sqrt{2(k^2+l^2)}} \end{bmatrix} \qquad (2.18)$$

In (2.18) the modes have been normalised to length unity. To project an arbitrary vector $\underline{X}$ onto the normal modes, one writes

$$\underline{X} = \underline{\underline{E}}\,\underline{y} \tag{2.19}$$

$\underline{X}$ is the known vector to be protected and $\underline{y}$ holds its normal mode components. To get $\underline{y}$, multiply (2.19) from the left by $\underline{\underline{E}}^{-1}$. This yields

$$\underline{y} = \underline{\underline{E}}^{-1}\,\underline{X} \tag{2.20}$$

As $\underline{\underline{E}}$ is orthogonal, $\underline{\underline{E}}^{-1}$ can simply be computed by making a conjugate transposition. The component form of (2.20) may then be written as

$$y_R = \frac{1}{\sigma}\left[+il\sqrt{\bar{\phi}}\,u + k\sqrt{\bar{\phi}}\,v + f_o\phi\right] \tag{2.21}$$

$$y_{GE} = \frac{1}{\sigma\sqrt{2(k^2+l^2)}}\left[-(if_o l + \sigma k)u - (f_o k + i\sigma l)v + \sqrt{\bar{\phi}(k^2+l^2)}\,\phi\right] \tag{2.22}$$

$$y_{GW} = \frac{1}{\sigma\sqrt{2(k^2+l^2)}}\left[-(if_o l - \sigma k)u - (f_o k - i\sigma l)v + \sqrt{\bar{\phi}(k^2+l^2)}\,\phi\right] \tag{2.23}$$

For given values of u, v and ϕ (for example analysed values) (2.21) - (2.23) shows how these fields project onto Rossby (2.21) and gravity modes (2.22,2.23). If the mass and wind fields are in geostrophic balance, they do not excite gravity waves, i.e. $y_{GE} = y_{GW} = 0$. On the other hand, a geopotential amplitude ϕ alone (without wind amplitudes, u,v) projects onto both Rossby and gravity modes. All the projections are scale dependent; for instance small scale wind fields project more on Rossby modes than large scale ones. In the atmosphere, gravity mode amplitudes are usually much smaller than Rossby mode amplitudes.

Problem 2

(i) Show that the normal modes are othogonal.

(ii) Show that geostrophic winds do not project on gravity modes for this model.

(iii) Derive mass and wind fields which do not project on Rossby waves.

2.3 MODEL EQUATIONS IN NORMAL MODE FORM

Inserting (2.19) (and a corresponding transformation for the vector of non-linear terms $\underline{N}$) into (2.9) yields:

$$\frac{\partial}{\partial t}\,\underline{\underline{E}}\underline{y} = i\underline{\underline{A}}\underline{\underline{E}}\underline{y} + \underline{\underline{E}}\underline{q} \tag{2.24}$$

$\underline{q}$ is the vector of normal mode amplitudes of $\underline{N}$.

Multiplication of (2.24) by $\underline{\underline{E}}^{-1}$ from the left results in:

$$\frac{\partial \underline{y}}{\partial t} = i\underline{\underline{E}}^{-1}\,\underline{\underline{A}}\underline{\underline{E}}\underline{y} + \underline{q} \tag{2.25}$$

The similarity transform $\underline{\underline{E}}^{-1}\,\underline{\underline{A}}\,\underline{\underline{E}}$ reduced $\underline{\underline{A}}$ to a diagonal matrix $\underline{\underline{D}}$ which holds the eigenvalues λ_j as its diagonal elements and zero entries elsewhere. Therefore, a decoupled system is obtained which can be written in component form.

$$\frac{dy_R}{dt} = i\lambda_1 y_R + q_R \tag{2.26}$$

$$\frac{dy_{GE}}{dt} = i\lambda_2 y_{GE} + q_{GE} \tag{2.27}$$

$$\frac{dy_{GW}}{dt} = i\lambda_3 y_{GW} + q_{GW} \tag{2.28}$$

These are the model equations in normal mode form. They form a set of decoupled, ordinary differential equations equivalent to the original system of coupled partial differential equations (2.1 to 2.3). Neglecting the non-linear terms, (2.26) - (2.28) can be integrated to give

$$Y_R(t) = y_R(t=o)\ , \tag{2.29}$$

$$Y_{GE}(t) = y_{GE}(t=o)e^{i\sigma t}\ , \tag{2.30}$$

$$Y_{GW}(t) = y_{GW}(t=o)e^{-i\sigma t} . \qquad (2.31)$$

(2.29) - (2.31) are the analytical solutions of the linearised version of the set of Equations (2.1)-(2.3). Once an analysis has provided the initial values of u, v and ϕ, (2.21) - (2.23) can be used to get $y_R(t=o)$, $y_{GE}(t=o)$ and $y_{GW}(t=o)$ and then (2.29) - (2.31) used to compute the value of the normal mode coefficients at a subsequent time t. Using (2.19), (2.8) and the inverse of (2.4) then allows the computation of the mass and wind fields in physical space. From (2.29) it can be seen that the Rossby mode coefficient y_R remains constant in time. However the gravity mode coefficients will oscillate with their (high) frequency σ unless their initial amplitudes $y_{GE}(t=o)$ and $y_{GW}(t=o)$ are zero.

At this point it is worth recalling that inertia-gravity waves are dispersive. Suppose that there is a zero wind field and a localised mass field disturbance at only a single point. This state will excite both types of waves within a wide spectrum of horizontal scales. Each inertia-gravity wave will travel away from the disturbance with its own characteristic phase speed thus leading to a broadening of the shape of the gravity mode part of the initial state. This will finally result in a state where only the Rossby wave components remain in the region of the initial disturbance. The transition process is known as geostrophic adjustment.

2.4 LINEAR NORMAL MODE INITIALISATION

In order to suppress unwanted oscillations in the linear model, it is clear from (2.30) and (2.31) that one must set $y_{GE}(t=o)$ and $y_{GW}(t=o)$ for the gravity modes to zero. Using (2.22) and (2.23) one obtains

$$\frac{1}{\sigma\sqrt{2(k^2+l^2)}}\left[(if_o l - \sigma k)u(t=o) - (f_o k - i\sigma l)v(t=o) + \sqrt{\bar{\phi}(k^2+l^2)}\,\phi(t=o)\right] = 0 \tag{2.32}$$

$$\frac{1}{\sigma\sqrt{2(k^2+l^2)}}\left[(if_o l + \sigma k)u(t=o) - (f_o k + i\sigma l)v(t=o) + \sqrt{\bar{\phi}(k^2+l^2)}\,\phi(t=o)\right] = 0 \tag{2.33}$$

There are various ways of satisfying (2.32) and (2.33). Usually one requires that the Rossby mode component y_R is not changed by the initialisation. This can be achieved simply by subtracting the gravity wave components from the initial field.

$$u_{LI} = u(t=o) - \frac{1}{\sigma\sqrt{2(k^2+l^2)}}\left[(if_o l-\sigma k)y_{GE}(t=o)+(if_o l+\sigma k)y_{GW}(t=o)\right] \tag{2.34}$$

$$v_{LI} = v(t=o) - \frac{1}{\sigma\sqrt{2(k^2+l^2)}}\left[(-f_o k+i\sigma l)y_{GE}(t=o)-(f_o k+i\sigma l)y_{GW}(t=o)\right] \tag{2.35}$$

$$\phi_{LI} = \phi(t=o) - \frac{\sqrt{\bar{\phi}(k^2+l^2)}}{\sigma\sqrt{2(k^2+l^2)}}\left[y_{GE}(t=o) + y_{GW}(t=o)\right] \tag{2.36}$$

By modifying the analysed initial state according to (2.34) - (2.36) it is ensured that the gravity wave amplitudes vanish initially and as a consequence of (2.30) and (2.31) - for all time if integration is performed with the linearised version of (2.1)-(2.3).

It is interesting to look at the relative changes to the mass ($\Delta\phi_{LI}$) and wind fields (Δu_{LI}, Δv_{LI}) due to linear normal mode initialisation. If for convenience the gravity mode components y_{GE} and y_{GW} are treated as identical, one obtains

$$\frac{\Delta\phi_{LI}}{\Delta u_{LI}} = \frac{-i\sqrt{\bar{\phi}\,(k^2+l^2)}}{f_o l} \tag{2.37}$$

$$\frac{\Delta\phi_{LI}}{\Delta r_{LI}} = \frac{-i\sqrt{\bar{\phi}\,(k^2+l^2)}}{f_o k} \tag{2.38}$$

From (2.37), (2.38) it is seen that the wind field changes decrease with increasing value of the Coriolis parameter. In other words, when approaching the tropics, the wind field becomes more dominant. Furthermore, the mass field changes increase with increasing mean depth ϕ of the fluid. Finally, the larger the horizontal scale, the more the wind field changes when compared to the mass field. All these results are consistent with geostrophic adjustment theory.

Problem 3

(i) Show that the initialised fields defined by (2.34)-(2.36) yield vanishing gravity mode projections.

(ii) Show that (2.34)-(2.36) do not change the Rossby wave projections.

(iii) Using (2.32) and (2.33) derive equations which do not change the mass field. Do these relations change the Rossby mode projections?

2.5 NON-LINEAR NORMAL MODE INITIALISATION

The relations (2.34)-(2.36) guarantee that, for a linear model, the gravity modes will vanish initially and stay zero for all time during the integration. However, when the same initial conditions are used to integrate a non-linear model, (2.26) - (2.28) show that the non-linear terms q_{GE} and q_{GW} will give rise to time tendencies, although y_{GE} and y_{GW} are initially zero. In other words, the gravity wave modes will not remain zero for a non-linear model. Therefore, one must use a different initialisation procedure. Machenhauer (1977) proposed that the initial tendencies should be set to zero. From (2.26)-(2.28) one then obtains

$$y_{GE} = \frac{iq_{GE}}{\sigma} \qquad (2.39)$$

$$y_{GW} = -\frac{iq_{GW}}{\sigma} \qquad (2.40)$$

(2.39)-(2.40) result in an initial state where the linear tendencies for the gravity waves are exactly balanced by the corresponding projections of the non-linear terms, thus yielding zero time tendencies for the gravity modes. In contrast to the linear normal mode initialisation, the initial amplitudes do not vanish in the non-linear case.

The mechanism for non-linear normal mode initialisation can be graphically displayed on a slow manifold diagram (Fig.3). It shows the space spanned by the Rossby and gravity modes. For the sake of simplicity, only one gravity mode is shown. The analysis can be represented by one particular point A in the diagram. Initialisation then modifies the analysis without changing the Rossby mode projection y_R. The initialised analysis will lie on the curve labelled M which is the locus of all model states where the initialisation equations (2.39)-(2.40) are fulfilled. For every analysed Rossby mode component a balancing gravity mode component y_{GB} is required. Note, that for $y_R=o$ there is generally a balancing gravity component because of the physical processes (such as convection, radiation and diffusion) in the model.

(2.39) and (2.40) form a set of non-linear algebraic equations as the non-linear terms lead to products of gravity modes with both gravity modes and Rossby modes. A convenient way to solve the system is to iterate

$$y_{GE}^{n+1} = \frac{iq_{GE}^{n}}{\sigma} \qquad (2.41)$$

$$y_{GE}^{n+1} = -\frac{iq_{GW}^{n}}{\sigma} . \qquad (2.42)$$

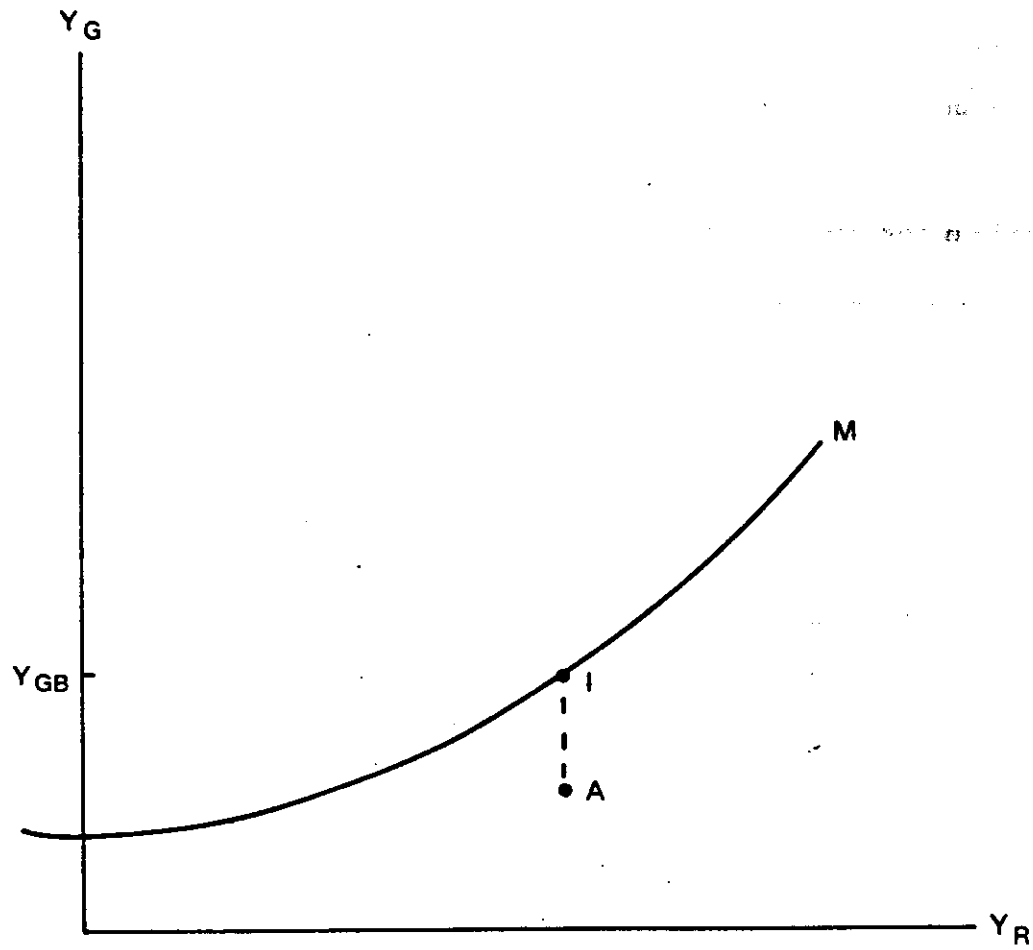


Fig. 3 Slow manifold diagram

Starting from the analysed fields, one first computes the non-linear terms. From (2.41) and (2.42) a first approximation is obtained to the gravity mode components which are then used for the renewed computation of the non-linear terms. This process is repeated and usually 2 or 3 iterations are sufficient.

At this point it should be mentioned that initialisation conditions other than (2.41) and (2.42) have been suggested by Baer and Tibbia. They use scaled versions of the normal mode form of the model equations. Then using various assumptions they derive relations which guarantee that the model shows no fast oscillations during the integration. Their scheme is somewhat more general than the scheme discussed above. However, it is more complex and difficult to use in practice. For a description see the review article by Daley and Machenhauer quoted in the references.

Problem 4

Show that for the non-linear normal mode initialisation the relative changes to mass and wind field are consistent with geostrophic adjustment theory.

2.6 Bounded derivative method

Recently, a second general concept of initialisation has been introduced by Kreiss, the bounded derivative method (BDM). Its basic idea is extremely simple: if a solution is to vary slowly in time, then the time derivative of the prognostic variables must be small in some sense.

In this Section we shall use the model described by (2.1)-(2.3) to show how this concept can be used to derive diagnostic equations (of differing degrees of approximation) which the initial conditions have to fulfill.

(a) Scaling the model equations

The first step in BDM is an appropriate scaling of the equations. We assume the following scales:

Horizontal length scale: $L = 10^6$ m

Horizontal particle speed: $U = 10$ m/sec

Time scale: $L/U = 10^5$ sec

The independent variables x,y,t are non-dimensionalised as follows

$$x = Lx' \qquad y = Ly' \qquad t = \frac{Lt'}{U} \tag{2.43}$$

Primes denote dimensionless quantities. For the dependent variables we get:

$$u = Uu' \qquad v = Uv' \qquad \phi = \bar{\phi} + f_o UL\phi'' \tag{2.44}$$

The coefficients $f_o UL$ of the dimensionless geopotential deviation ϕ'' is implied by a quasi-geostraophic relation between mass and wind.

For the physical parameters we assume

$$\begin{aligned} f_o &= 10^{-4} \text{ sec}^{-1} \\ \bar{\phi} &= 10^5 \text{ m}^2 \text{ sec}^{-2} \end{aligned} \tag{2.45}$$

The scaled version of (2.1)-(2.3) then becomes

$$\frac{\partial u'}{\partial t'} + \frac{f_o L}{U}\left(-v' + \frac{\partial \phi''}{\partial x'}\right) = -u'\frac{\partial u'}{\partial x'} - v'\frac{\partial u'}{\partial y'} + \frac{F_x L}{U^2} = N'_u \tag{2.46}$$

$$\frac{\partial v'}{\partial t'} + \frac{f_o L}{U}\left(u' + \frac{\partial \phi''}{\partial y'}\right) = -u'\frac{\partial v'}{\partial x'} - v'\frac{\partial v'}{\partial y'} + \frac{F_y L}{U^2} = N'_v \tag{2.47}$$

$$\frac{\partial \phi''}{\partial t'} + \frac{\bar{\phi}}{LUf_o}\left(\frac{\partial u'}{\partial x'} + \frac{\partial v'}{\partial y'}\right) = -\phi''\left(\frac{\partial u'}{\partial x'} + \frac{\partial v'}{\partial y'}\right) - u'\frac{\partial \phi''}{\partial x'} - v'\frac{\partial \phi''}{\partial y'} + \frac{Q}{U^2 f_o} = N'_\phi \tag{2.48}$$

In the following, the primes will be omitted. With $\varepsilon = U/f_o L$ being a parameter of the order of 10^{-1}, (2.46) - (2.48) become

$$\frac{\partial u}{\partial t} + \varepsilon^{-1} \left(-v + \frac{\partial \phi}{\partial x}\right) = N_u \tag{2.49}$$

$$\frac{\partial v}{\partial t} + \varepsilon^{-1} \left(u + \frac{\partial \phi}{\partial y}\right) = N_v \tag{2.50}$$

$$\frac{\partial \phi''}{\partial t} + \varepsilon^{-2} \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial x}\right) = N_\phi \tag{2.51}$$

Now we will assume that all fields are smooth; this guarantees that horizontal derivatives are of order unity (note however that this assumption will not always be fulfilled in practice).

(b) Bounding the time derivatives

As N_u, N_v and N_ϕ are of order of unity, it is clear from (2.49)-(2.51) that the time derivatives will be of order unity if and only if

$$-v + \frac{\partial \phi}{\partial x} = \varepsilon a(x,y,t) \tag{2.52}$$

$$u + \frac{\partial \phi}{\partial y} = \varepsilon b(x,y,t) \tag{2.53}$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = \varepsilon^2 c(x,y,t) \tag{2.54}$$

Here a,b and c are assumed to be of order unity and have yet to be determined. An important conclusion from (2.54) is that the initial divergence has to be of order ε^2 in order to ensure smooth integrations.

From the left hand sides of (2.52)-(2.54) we can infer that the following relation has to hold

$$\frac{\partial b}{\partial x} - \frac{\partial a}{\partial y} = \varepsilon c \tag{2.55}$$

At this stage we could choose a,b and c to be smooth arbitrary functions of order unity which satisfy (2.55). One choice would be to drop terms of order ε. This would result in geostrophic winds as initial conditions, which bound the first order time derivative. If one wants to go to a higher approximation, additional equations can be derived requiring the second order time derivatives to be of order unity.

Differentiating (2.46)-(2.48) with respect to t and using (2.52)-(2.54) we get

$$\frac{\partial^2 u}{\partial t^2} + \frac{\partial a}{\partial t} = \frac{\partial N_u}{\partial t} \tag{2.56}$$

$$\frac{\partial^2 v}{\partial t^2} + \frac{\partial b}{\partial t} = \frac{\partial N_v}{\partial t} \tag{2.57}$$

$$\frac{\partial^2 \phi}{\partial t^2} + \frac{\partial c}{\partial t} = \frac{\partial N_\phi}{\partial t} \tag{2.58}$$

By making use of (2.52)-(2.54) we have made sure that the time derivatives are bounded. Therefore, terms of the form $\frac{\partial^2}{\partial x \partial t}$ which occur on the right hand side of (2.56)-(2.58) are of order unity. Hence, the second order time derivatives are of order unity if and only if $\frac{\partial a}{\partial t}$, $\frac{\partial b}{\partial t}$ and $\frac{\partial c}{\partial t}$ are of order unity.

We can easily derive equations for these variables by differentiating (2.52)-(2.54).

$$\varepsilon \frac{\partial a}{\partial t} = -\frac{\partial v}{\partial t} + \frac{\partial^2 \phi}{\partial x \partial t} = b - N_v + \frac{\partial}{\partial x}(-c + N_\phi) \tag{2.59}$$

$$\varepsilon \frac{\partial b}{\partial t} = \frac{\partial u}{\partial t} + \frac{\partial^2 \phi}{\partial y \partial t} = -a + N_u + \frac{\partial}{\partial y}(-c + N_\phi) \tag{2.60}$$

$$\varepsilon^2 \frac{\partial c}{\partial t} = \frac{\partial}{\partial t}\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) = \frac{\partial}{\partial x}(-a + N_u) + \frac{\partial}{\partial y}(-b + N_v) \tag{2.61}$$

Note that (2.61) is a linear combination of (2.59) and (2.60). In order to solve (2.59) and (2.60) - together with (2.55) instead of (2.61) - for a, b and c we require initial values of these functions. Unfortunately these are just the variables we are trying to determine. However, if we determine a, b and c only to order unity (i.e. if we neglect terms of order ε), we can uniquely solve the system for a, b and c. Once we have computed these variables, we an use (2.52)-(2.54) to determine u, v and ϕ. As (2.54) is a linear combination of (2.52) and (2.53) we can arbitrarily specify one variable and compute the others accordingly. For a mid-latitude model one would, for instance, specify the mass field and compute the wind field from (2.52) and (2.53).

For this type of model it is easy to establish the relationship with the quasi-geostrophic theory. Neglecting terms of order ε^2 in (2.54) and (2.61), and replacing a and b in (2.61) by (2.52) and (2.53), one gets after some manipulation

$$\nabla^2\phi - \left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right) = \varepsilon\left[2\left(\frac{\partial u_R}{\partial x}\frac{\partial v_R}{\partial y} - \frac{\partial u_R}{\partial y}\frac{\partial v_R}{\partial x}\right) - \frac{L}{U^2}\left(\frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y}\right)\right] \qquad (2.62)$$

Here u_R and v_R are the rotational wind components which are solutions of (2.54) with c=o. Equations (2.62) is the well-known balance equation which has been used for initialisation for a long time. As we saw, it bounds the model variables by requiring the divergence and its first derivative to vanish initially.

(c) Discussion

To the degree of accuracy used so far, the bounded derivative method yields results identical to the quasi-geostrophic theory. Therefore, there appears

to be no advantage in the procedure. However, BDM is more general as we can easily derive higher order constraints by bounding higher time derivatives of a, b and c. Furthermore, the method can also be applied in the tropics where the quasi-geostrophic approach breaks down. An important aspect with respect to limited area modelling is the fact that the influence of the boundary conditions can be incorporated.

A difficulty with BDM is the solution of the diagnostic equations which get more and more complex for higher order constraints. Already (2.62) can lead to problems in anticyclonic areas where an ellipticity criterion must be satisfied in order to solve (2.62) as a boundary value problem; in practice this criterion may not be fulfilled. A further problem is the separation of the free and diagnostic variables. Using geostrophic adjustment theory, one usually wants to specify the mass field in higher latitudes and the wind field in the tropics. However, as explained in Section 2.1.d, these relationships critically depend upon the horizontal and vertical scales. Overall, the BDM is of most use for limited area modelling whereas normal mode initialisation is particularly suitable for global or hemispheric models.

Problem 5

Show that replacing u_R and v_R in (2.62) by geostrophic winds u_g and v_g bounds the first and second order time derivatives to order ϵ.

3. INITIALISATION OF A GLOBAL BAROCLINIC MODEL

3.1 COMPUTATION OF THE NORMAL MODES

As in Section 2 the model equations (divergence, vorticity and thermodynamic equations) are linearized about a basic state at rest, with a temperature profile $\bar{T}(\eta)$ a function of height only. The model equations can be written in matrix form:

$$\begin{aligned} \frac{\partial \underline{D}}{\partial t} - f\underline{\xi} + \beta\underline{u} + \nabla^2 \underline{P} &= \underline{R}_D = 0 \\ \frac{\partial \underline{\xi}}{\partial t} + f\underline{D} + \beta\underline{v} &= \underline{R}_\xi = 0 \\ \frac{\partial \underline{P}}{\partial t} + \underline{\underline{B}}\,\underline{D} &= \underline{R}_P = 0 \end{aligned} \qquad 3.1$$

The terms on the right hand side contain all the nonlinear tendencies and are here set to zero. The vector notation is used in 3.1 to represent the values at all the model levels. The vertical structure matrix $\underline{\underline{B}}$ depends on the basic state chosen and on the numerical technique used in the vertical discretization. The auxiliary potential P is defined as $P = \bar{\phi} + R\bar{T} \ln p_s$. In the definition of the geopotential of the mean state $\bar{\phi}$ a mean surface pressure $\bar{p}_s$ is assumed.

In order to separate the vertical dependence from the horizontal in 3.1 the model variables $\underline{D}$, $\underline{\xi}$ and $\underline{P}$ are expressed in terms of the eigenvectors $\underline{\phi}_m$ of matrix $\underline{\underline{B}}$. For example:

$$\underline{D} = \sum_{m=1}^{M} D_m \underline{\phi}_m \qquad 3.2$$

The equations obtained after substitution of 3.2 into 3.1 have the form of M independent systems of shallow-water equations with equivalent geopotential depth ϕ_m, equal to the eigenvalue corresponding to $\underline{\phi}_m$. Formally the

linearised baroclinic problem can be treated as a sequence of M barotropic problems in the vertical mode representation. Therefore, the methods of Section 2 can be used for baroclinic models. Once the matrix B is defined and its eigenvectors and eigenvalues found. The initialisation procedure for barotropic models can be used if ϕ is replaced and the model variables by the m'th amplitude of the vertical mode expansions. The eigenvalues, ϕ_m, are called 'equivalent depths'.

After performing the vertical separation, the M two-dimensional systems may be separated in the zonal direction by Fourier transforming the variables; thus we write e.g.

$$D_m(\lambda,\theta,t) = \sum_{k=0}^{N-1} D_{m,k}(\theta,t) \exp(ik\lambda) \qquad 3.3$$

If we now call $\underline{x}_{m,k}$ the vector which contains $D_{m,k}$, $\xi_{m,k}$ and $P_{m,k}$ (scaled to be non-dimensional), the system of linear equations becomes formally:

$$\frac{d\underline{x}_{m,k}}{dt} = i\,\underline{\underline{A}}_{m,k}\,\underline{x}_{m,k} \qquad 3.4$$

The matrix $\underline{\underline{A}}_{m,k}$ is real and symmetric. Hence its eigenvectors are orthogonal. They form a set of horizontal normal modes which can be used to express $\underline{x}$ (dropping the indices m,k for simplicity of notation):

$$\underline{x} = \sum_{\ell=1}^{3L} c_\ell\,\underline{\xi}_\ell \qquad 3.5$$

In fact these modes naturally divide into two classes: symmetric and antisymmetric with respect to the equator. This property is used to reduce the dimension of the matrix $\underline{\underline{A}}$ when finding its eigenvectors which are the normal modes required.

It is illuminating to look at the structure of the modes. Every mode shows a specific 3-dimensional structure of the mass and wind fields. The vertical structures are identical for Rossby and for gravity waves, but there are considerable differences in their horizontal structure and in the relation between mass and wind fields. Fig. 4 gives some examples of vertical structures. The first vertical mode (the external mode) represents fields which are nearly constant throughout the atmosphere. It therefore describes the barotropic component of the mass and wind fields. The second mode (also called the first internal) changes sign near the tropopause, thus accounting for the differences between stratosphere and troposphere. In general, vertical model m has m-1 sign changes. With increasing order m, the region of maximum amplitudes moves towards lower levels.

The horizontal structure of an external Rossby mode is shown in Fig. 5. The prominent feature is the approximately geostrophic relation between geopotential and wind in the extra-tropics. In the tropics, where the geostrophic relation is not applicable, normal modes still define a relation between the mass and wind fields. In fact, they are the only means to establish globally valid coupling. The Rossby mode shown in Fig. 5 can be characterised by 3 indices: the vertical mode number, the zonal wavenumber and the meridional index. As Fig. 5 shows an external mode, it means that at every point the wind components and the geopotential all have the same vertical structure as the curve labelled 1 in Fig. 4. The meridional index in Fig. 5 is 1, which means the largest meridionally symmetric scale. Obviously, the zonal wavenumber is 1.

When looking at a gravity mode (Fig. 6) for the same indices (vertical mode 1, zonal wavenumber 1, gravest symmetric), the ageostrophic structure of these modes is immediately evident. In the extra-tropics, strong winds are not

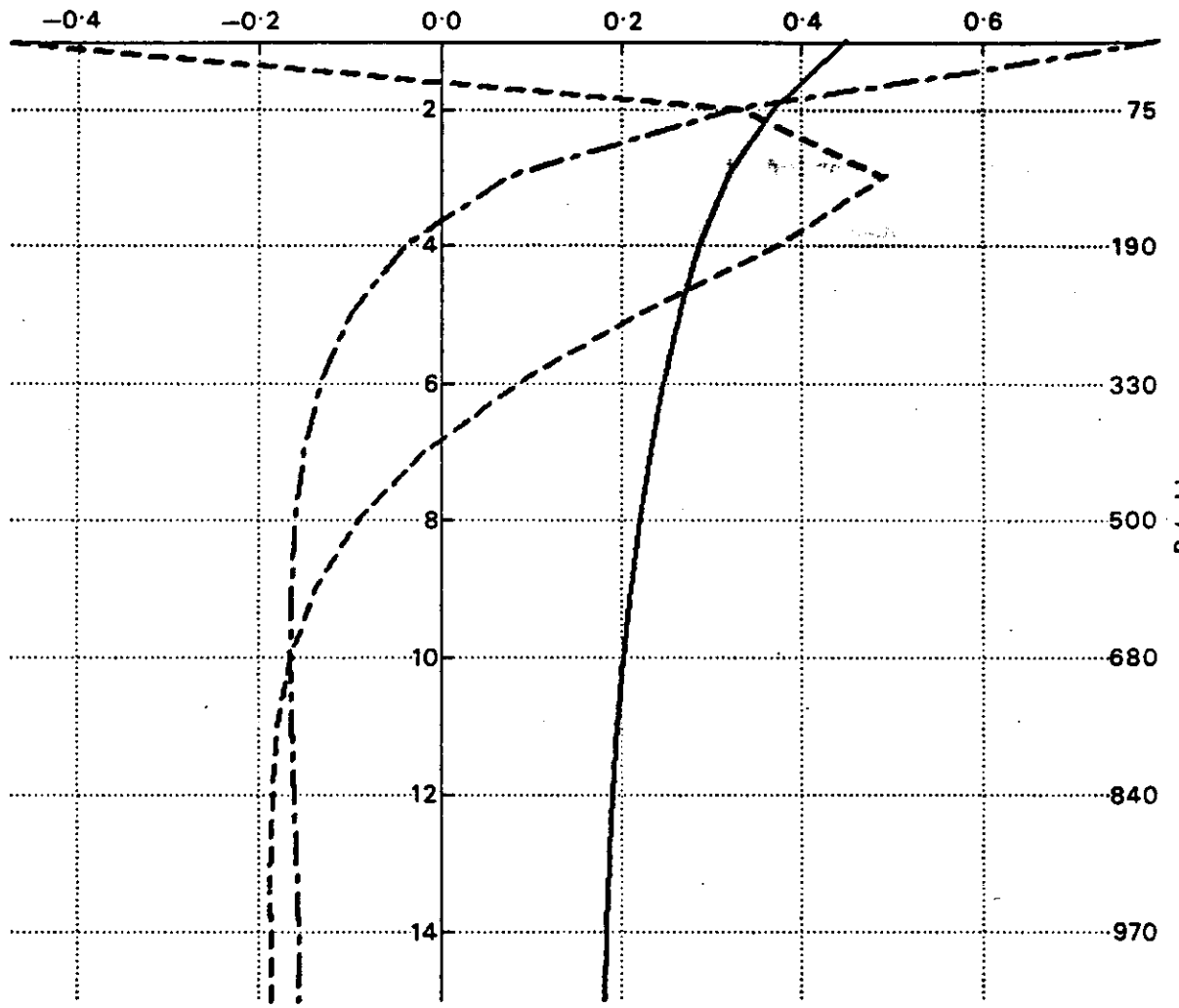


Fig. 4 First 3 vertical modes for isothermal (300 K) basic state, Rounded pressure values in mb are valid for p_s = 1000 mb.

supported by a height gradient. In the tropics the flow is highly divergent. Gravity waves can further be separated into two groups: the eastward and westward travelling waves. Fig. 6 shows a westward travelling gravity wave.

For the same combination of indices, Rossby and gravity waves have very different phase speeds (and periods). For example, the free period of the Rossby wave in Fig. 5 is 4.7 days, whereas the westward gravity wave in Fig. 6 has a free period of only 12.7 hours. Apart from a few exceptions gravity wave periods decrease with increasing zonal wavenumber and meridional index. In contrast, large scale Rossby waves have shorter periods than small scale ones. For both types, the periods increase with increasing vertical mode number. Therefore, a large scale internal gravity mode may have the same period as an external small scale Rossby mode. That is why the period is not a useful tool to distinguish between Rossby and gravity modes; nor is the wavelength, as apparent from Figs. 5 and 6. The only way to make a proper distinction is to look at the 3-dimensional structure of both the mass and the wind fields.

Note that the derivation of the structure of the modes only has to be done once; the results are stored and used in all subsequent initialisations.

Table 2.1 gives the numerical values of the first 5 eigenvalues (divided by g) as given by the ECMWF model for $\overline{T}$ = 300K. Also shown is the square root of ϕ_m which corresponds to the phase speed of gravity waves for a non-rotating earth (f_o=0).

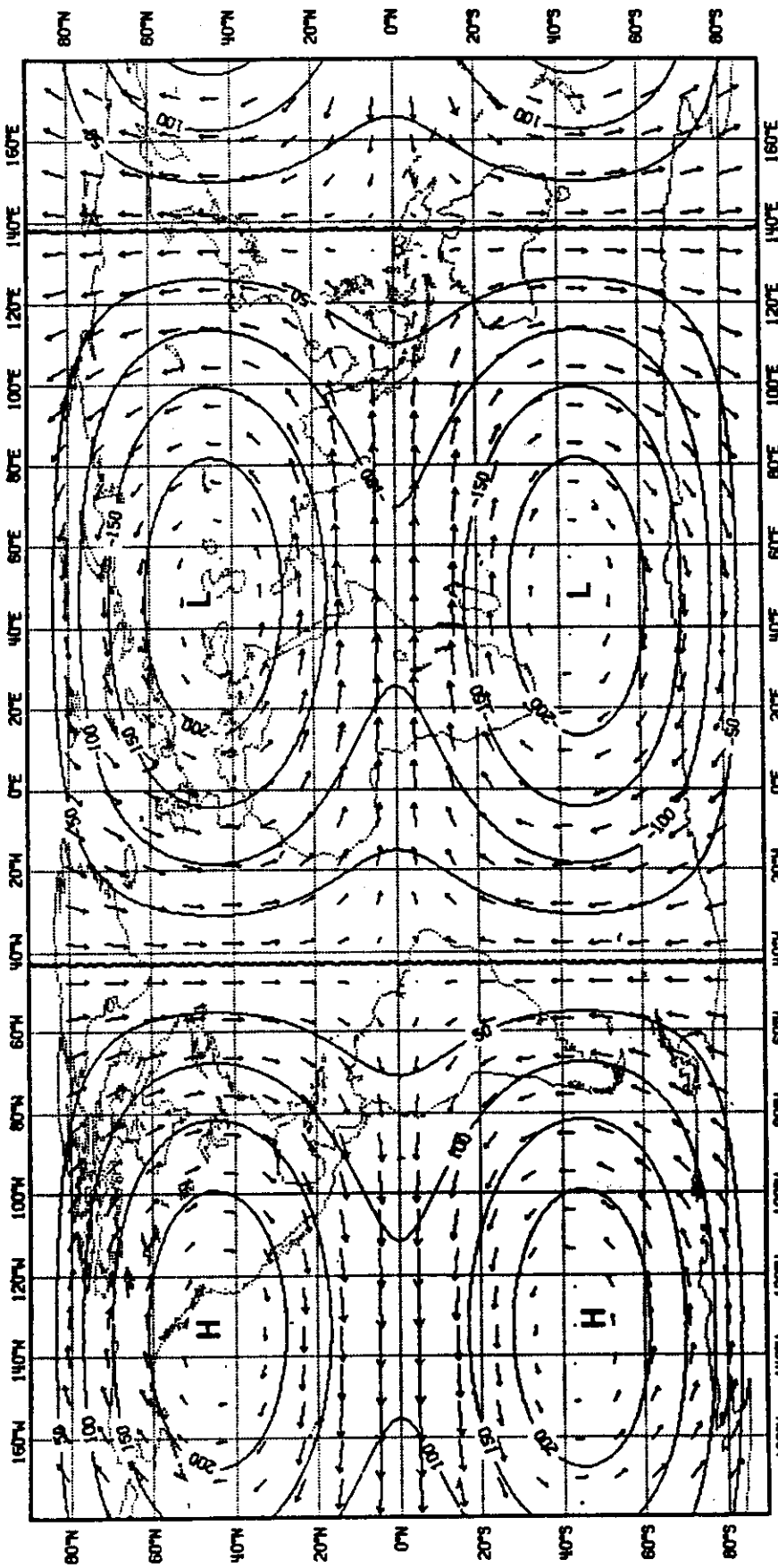


Fig. 5 Gravest symmetric, zonal wavenumber 1, external Rossby mode.

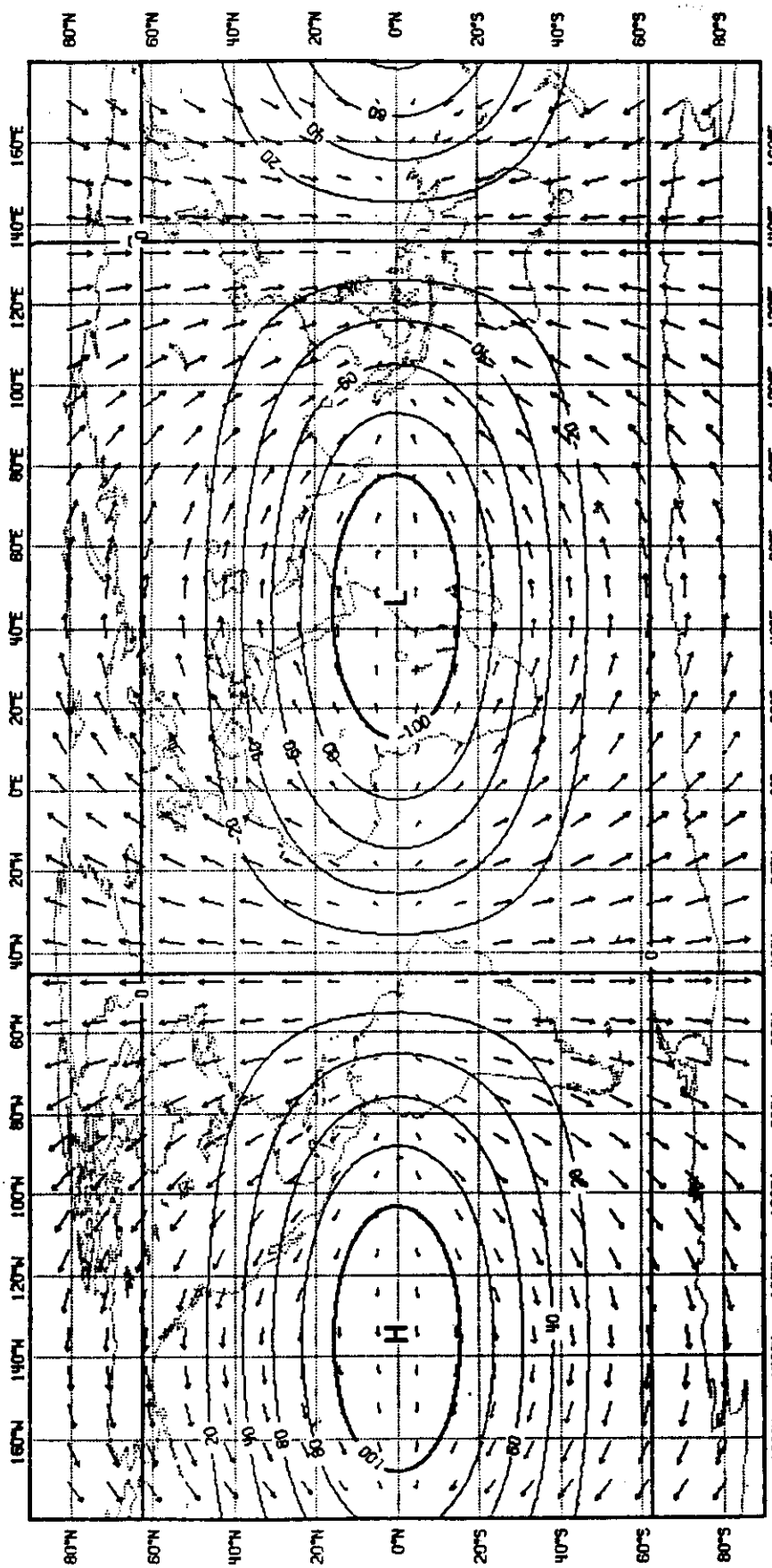


Fig. 6 Gravest symmetric, zonal wavenumber 1, westwards travelling, external gravity mode.

Table 2.1 Equivalent depths and phase speeds of pure gravity waves if the first 5 vertical modes for the ECMWF model with $\bar{T}$=300K.

k	ϕ_m/g in m	$\sqrt{\phi_m}$ in m/sec
1	11.89×10^3	341
2	3.37×10^3	182
3	947	96
4	365	60
5	170	41

With increasing vertical mode number m, the equivalent depth decreases. Therefore, higher vertical modes have low frequencies for inertia-gravity waves, and must not be included in the initialisation process as the intention is only to remove only unwanted high frequency oscillations.

3.2 THE INITIALISATION PROCESS

Using 3.5, the equation 3.4 can now be written

$$\frac{dc_\ell}{dt} = i\ \nu_\ell\ c_\ell \qquad 3.6$$

for each ℓ, with ν_ℓ being the eigenvalues of $\underline{\underline{A}}$.

Hence

$$\underline{x}(t) = \sum_\ell c_\ell(0)\ \exp\ (i\nu_\ell t)\ \underline{\xi}_\ell \qquad 3.7$$

where the amplitudes $c_\ell(0)$ are determined by the values of D, ξ, P at t = 0. At least for the first few vertical modes (with large equivalent depths ϕ_m) there is a clear distinction between low-frequency Rossby wave solutions (small ν_ℓ) and high frequency gravity wave solutions (large ν_ℓ). Only solutions of the former type are observed in the atmosphere with significant amplitude.

If the real model equations were linear, it would be easy to ensure that high frequency gravity waves do not exist by simply reducing to zero the corresponding normal mode coefficient $c_\ell(0)$ of the analysis. But this method does not work for the full nonlinear model.

The equivalent of 3.6 for the nonlinear equations is

$$\frac{dc_\ell}{dt} = i\ \nu_\ell\ c_\ell + r_\ell(t) \qquad 3.8$$

The term r_ℓ is the projection of the nonlinear terms of the model equations (computed by running one time step of the model) onto the normal modes. If one were to simply set $c_\ell = 0$ for t=0, very soon this mode will reappear, forced by r_ℓ. This was shown by Williamson (1976).

Machenhauer (1977) has proposed an iterative scheme for removing the gravity-mode oscillations by setting the initial time-derivatives of the gravity-mode coefficients to zero. From 3.8,

$$\left(\frac{dc_\ell}{dt}\right)_{t=0} = 0 \quad \text{if } c_\ell(0) = -\frac{r_\ell(0)}{i\ \nu_\ell}\ . \qquad 3.9$$

Since the nonlinear term $r_\ell(0)$ depends partly on the gravity-mode coefficients themselves, it is necessary to iterate the procedure; but for a barotropic model (or for the first few vertical modes of a multi-level model) the scheme converges rapidly, and two iterations are perfectly adequate.

Since the initialisation condition 3.9 requires stationarity for the initialisation of inertia-gravity waves, it clearly mishandles the tidal component of the atmospheric circulation. It should be allowed to propagate westwards and therefore be excluded from the initialisation process. This is achieved by performing a time series analysis of the total dynamical and physical tendencies for the 10 days preceding the actual analysis time. The

westward propagating component with a 24 hour period for zonal wavenumbers are are and a 12 hour period for zonal wavenumber two are excluded from 3.9 for all five vertical modes and for the eight gravest meridional modes. With t_ℓ being the tidal component of the tendencies, the initialisation condition becomes:

$$c_\ell(0) = - \frac{r_\ell(0) + d_\ell - t_\ell}{i\nu_\ell} \qquad 3.10$$

Further discussion of the tidal problem is given in Section 4.5.

In the current version of the analysis cycle two iterations of Machenhauer's procedure are carried out, initializing just the first five vertical modes. (The higher internal modes have very low frequencies and thus do not contribute to the problem of spurious high-frequency oscillations). The nonlinear forcing terms are computed by running the model itself for one timestep at each iteration. Although in principle the non-linear forcing can include the "physics" package as well as the dynamics, in practice this leads to the immediate divergence of the iteration process. Therefore, an estimate d_ℓ of the quasi-stationary part of the physical forcing is used, which is kept constant for all iterations. This estimate is computed by time-averaging the physical tendencies during a 2 hour forecast starting from an uninitialized analysis. Only those components which force inertia-gravity waves with periods longer than a certain cut-off period are retained, thus discarding less reliable small-scale structures. Operationally this cut-off period is 11 hours.

The steps of the initialization procedure can be summarized as follows:

(a) Run model for 2 hours from the uninitialized analysis to compute time-averaged physical forcing without diurnal cycle.

(b) Filter physical forcing field.

(c) Run adiabatic model for one timestep to compute non-linear terms.

(d) Extract tidal signal from the total tendencies.

(e) Compute new gravity mode coefficients according to (10).

(f) Restore analysed surface pressure after first iteration.

(g) As (c) but starting from results of first iteration step.

(h) As (c) to start the second iteration.

Changes brought about by initialisation do not affect the Rossby waves. In principle both the mass and wind field can be changed, but in reality initialisation mainly affects the divergent wind component and the surface pressure. Typical values of initialisation changes are 1-2 ms^{-1} for divergent winds in the upper troposphere and 1 hPa for the surface pressure. From time to time much larger changes can be found, but these invariably point to erroneous data which have found their way into the analysis despite the many data checks. After initialisation all fields are in a non-linear global three-dimensional balance which allows noise free forecasts.

It is worth noting that uninitialised analyses usually fit the observations better than the initialised analyses. However it is important that only initialised analyses are used for calculations of divergence or derived quantities such as vertical velocity or surface pressure tendency.

The effect of the analysis-initialisation cycle on the vertical velocity is illustrated in Fig. 7. A 500 hPa analysis is shown in Fig. 7a and the corresponding vertical velocity field from the first guess given in Fig. 7b clearly shows the large-scale ascent ahead of the trough in northerly latitudes. After initialisation, Fig. 7c, the ascent has been maintained, though there are significant changes over the Mediterranean and North Africa.

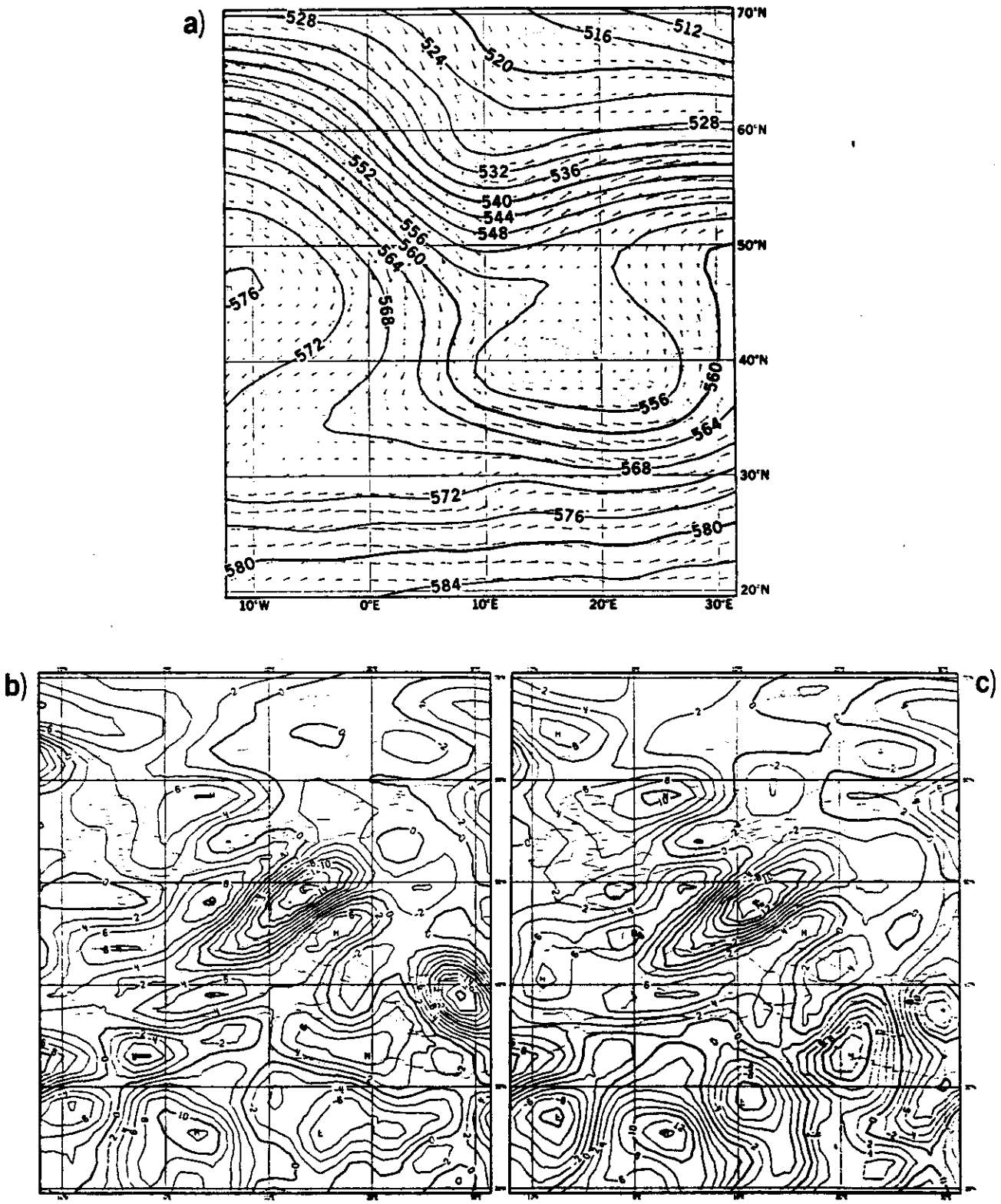


Fig. 7 500 hPa analysis of wind and height at 12 GMT, 6 December 1983 and the corresponding ω field (mb/h) from (b) the first guess and (c) after initialisation.

4. THE PERFORMANCE OF THE INITIALISATION SCHEME IN DATA ASSIMILATION

In the following the performance of the ECMWF initialisation scheme will be discussed under three aspects:

- the control of noise
- the changes made to the analysed fields
- the quality of forecasts.
- the spin-up problem

4.1 NOISE CONTROL

Fig. 8a shows the evolution of surface pressure during a 24 hour forecast started from the uninitialised (full line) and diabatically initialised (dashed line) analysis for a gridpoint in the northern Rocky Mountains. The uninitialised forecast is contaminated with high frequency noise, which leads to unrealistic surface pressure variations. The initialised forecast predicts the slow rise in pressure without the superimposed high frequency oscillations. The orography does not seem to cause any particular problems. Fig. 8b gives similar curves for a point off the Norwegian coast. The forecast range is extended to 5 days. Again, the uninitialised forecast is noisy. It takes about 3 days into the forecast to dampen the gravity wave activity. Note, that both forecasts are very close to each other in describing the synoptic development. Finally, Fig. 8c shows the vertical velocity at model level 8 for the same point during a 1 day forecast. While surface pressure is dominated by the external modes, the vertical velocity reflects noise in internal modes. There are large oscillation of ω in the forecast started from the uninitialised analysis. By contrast, the evolution of ω in the initialised forecast is smoother, thus showing that the initialisation scheme has some skill in controlling internal gravity wave activity.

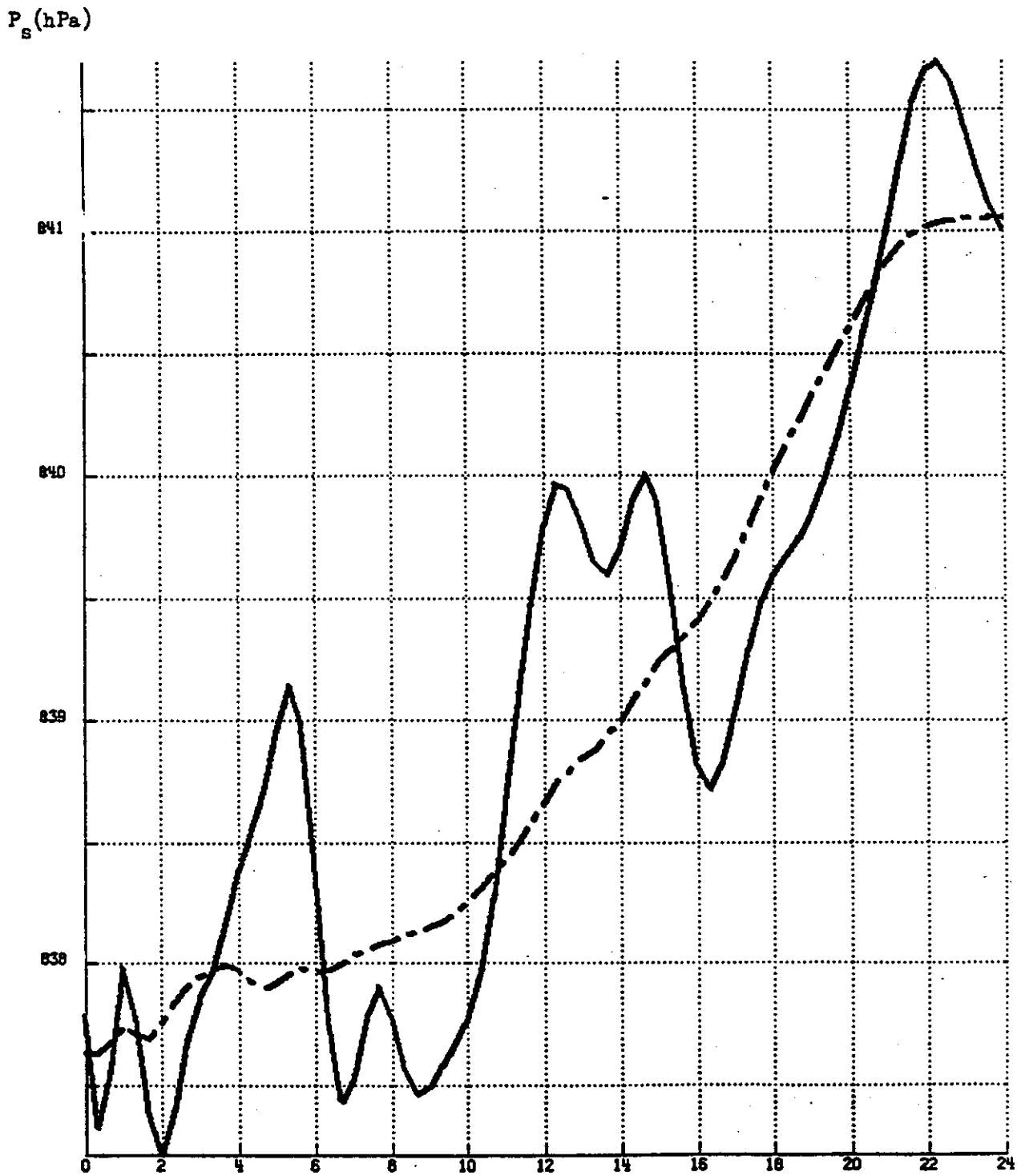


Fig. 8(a) Surface pressure evolution at a grid point in the northern Rocky Mountains. 24 hour forecast started from uninitialised (full line) and initialised (dashed line) analysis.

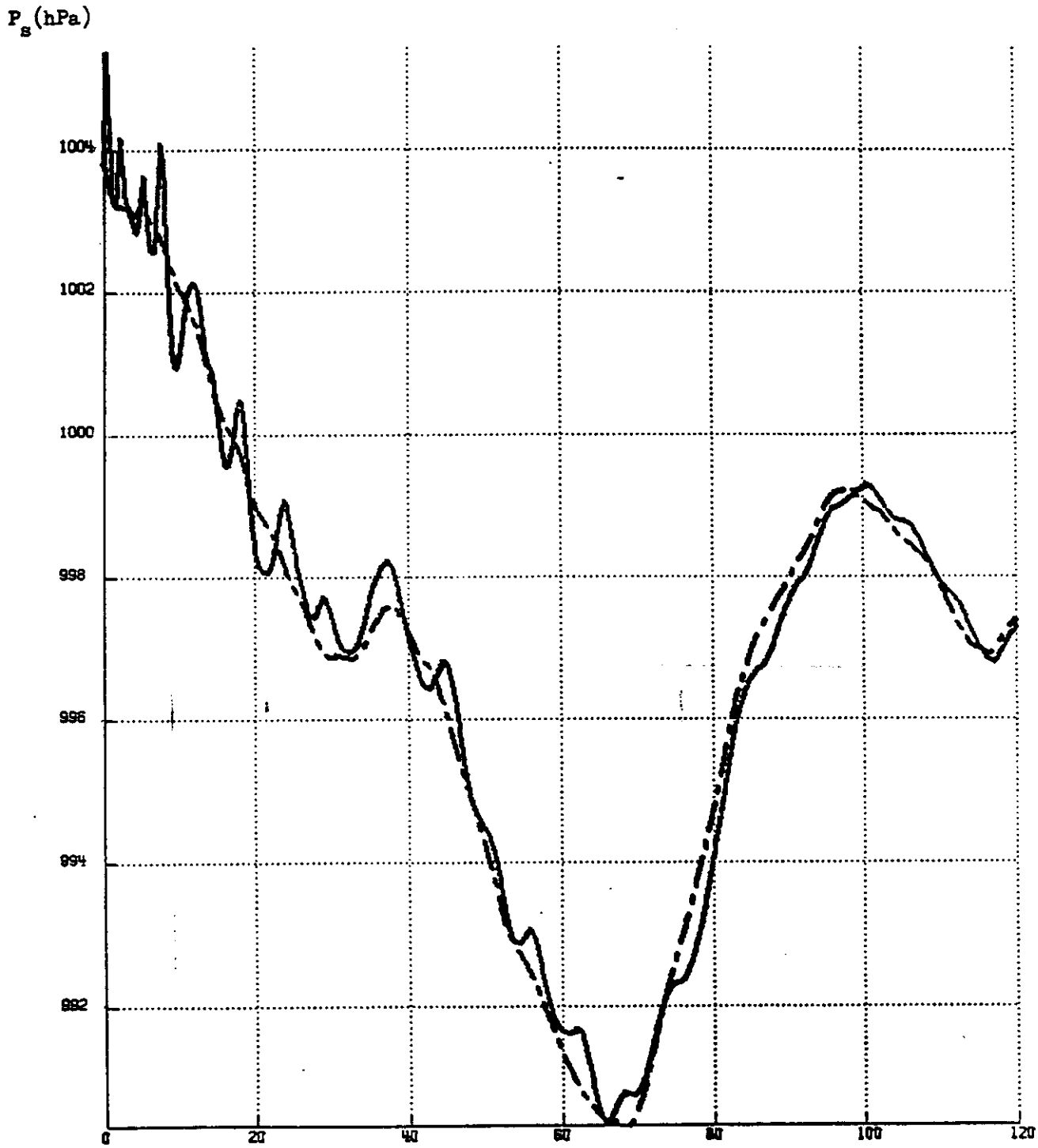


Fig. 8(b) As Fig. 8(a), but for a point off the Norwegian coast (64°N, 131°W).

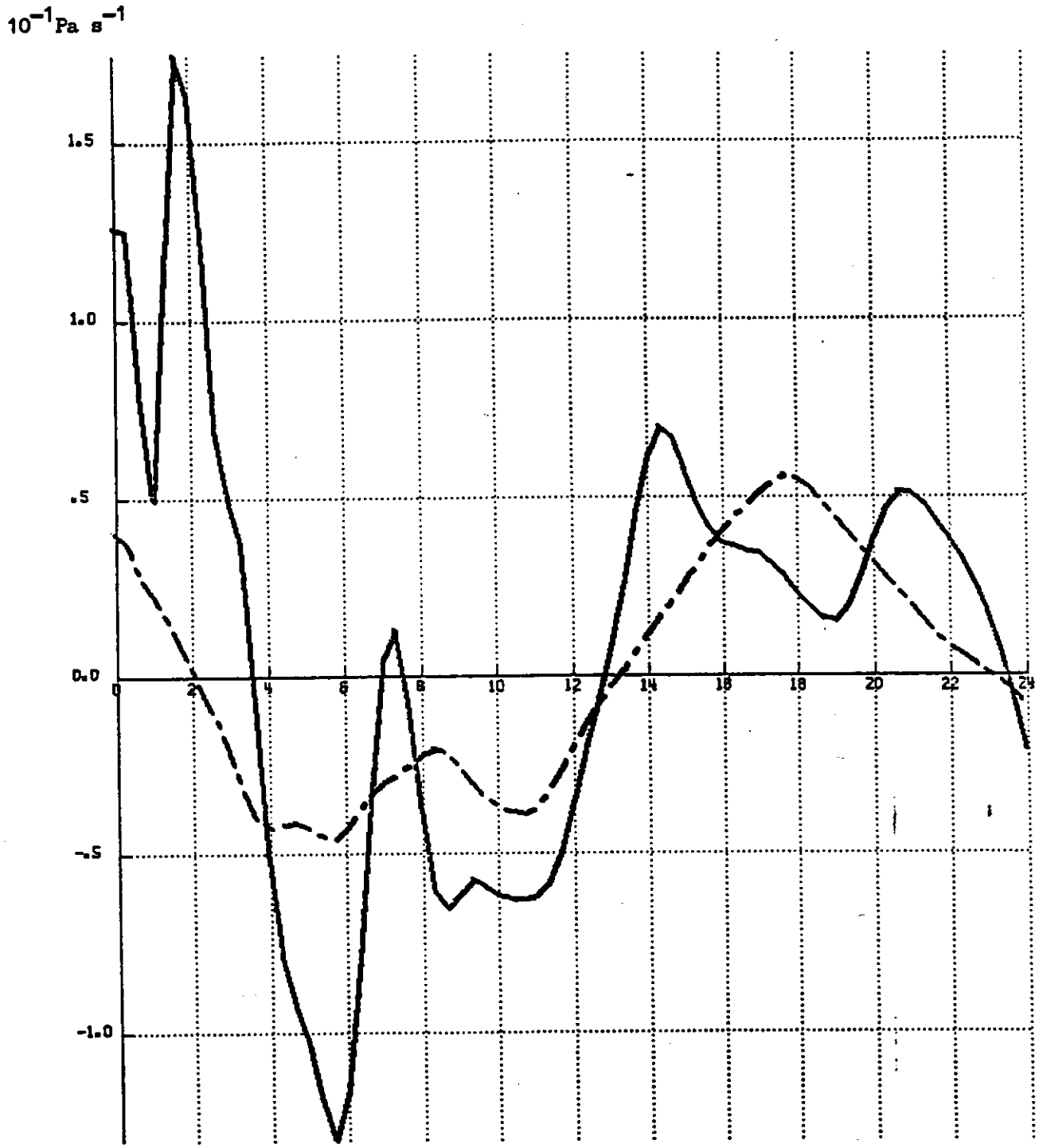


Fig. 8(c) As Fig. 8(b), but for vertical velocity at model level 8 (about 250 hPa).

A further indication of the effective noise control brought about by the diabatic initialisation is given in Fig. 9. It shows the field of instantaneous surface pressure tendencies, together with the reported values as registered during the 3 hours proceeding the analysis time. Note that the observed pressure tendencies are not used in the ECMWF analysis system. They are diagnostically derived using the analysed wind and surface pressure fields. Allowing for the small time difference, the tendencies computed from the initialised analysis (Fig. 9a) fit the reported value quite well. The area of falling pressure in front of a cyclone approaching the British Isles is verifying in amplitude and phase by the reports. Also, the area of rising pressure in the Atlantic is supported by the few observations in this area. When calculating the pressure tendencies from the uninitialised analysis, there is little agreement between this field and the observations (Fig. 9b). Note that the contour interval has been increased from 1mb/3h in Fig. 9a to 5mb/3h in Fig. 9b. There are many small scale structures with excessively high tendency values. It is almost impossible to make synoptic sense of this field. A typical horizontal scale of the noise seems to be around 700 km. The level of noise in the uninitialised analysis is hardly surprising as the analysis scheme approximates the global nonlinear three-dimensional constraints between the variables by simple linear local relations. A further source of imbalance is the difference in data selection between neighbouring analysis boxes and in different vertical slabs within the same analysis box.

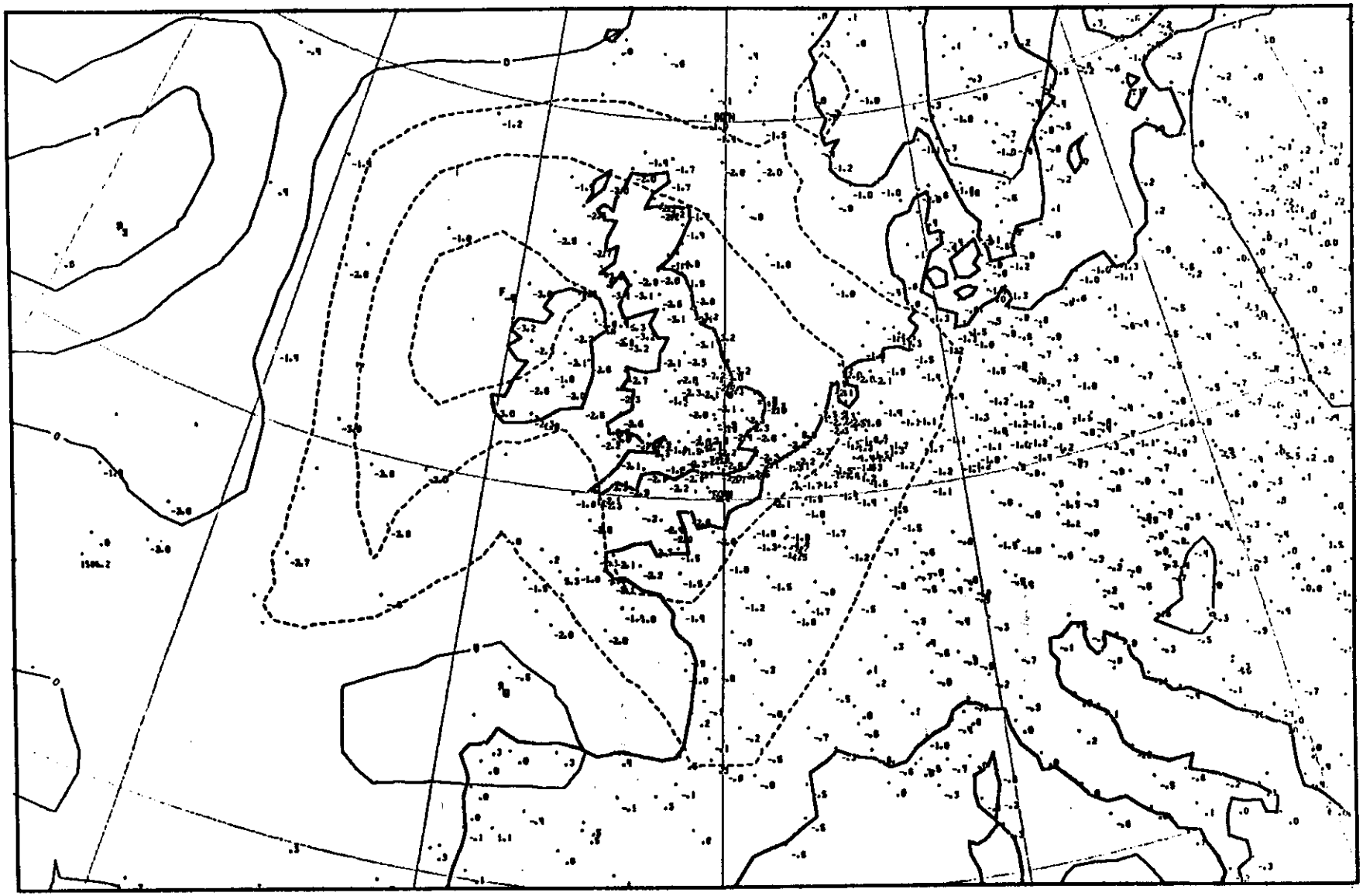


Fig. 9(a) Instantaneous surface pressure tendencies computed from an initialised analysis (contours), together with reported values as registered during the 3 hours preceeding the analysis time (spot values).

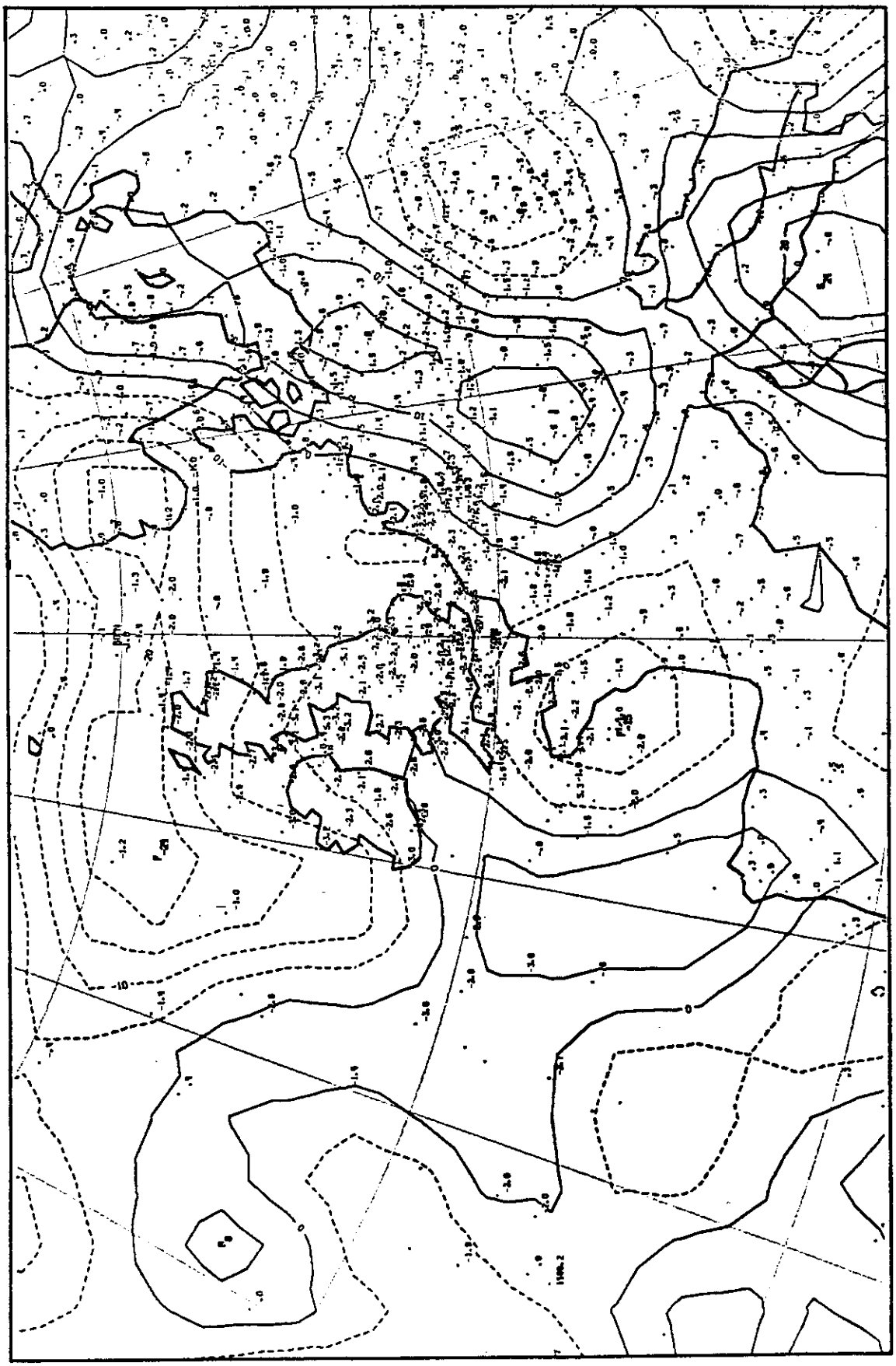


Fig. 9(b) As Fig. 9(a), but for the unintialised analysis.

4.2 INITIALISATION CHANGES TO THE ANALYSIS

In general, initialisation changes to the analysed fields can have their origin in three areas:

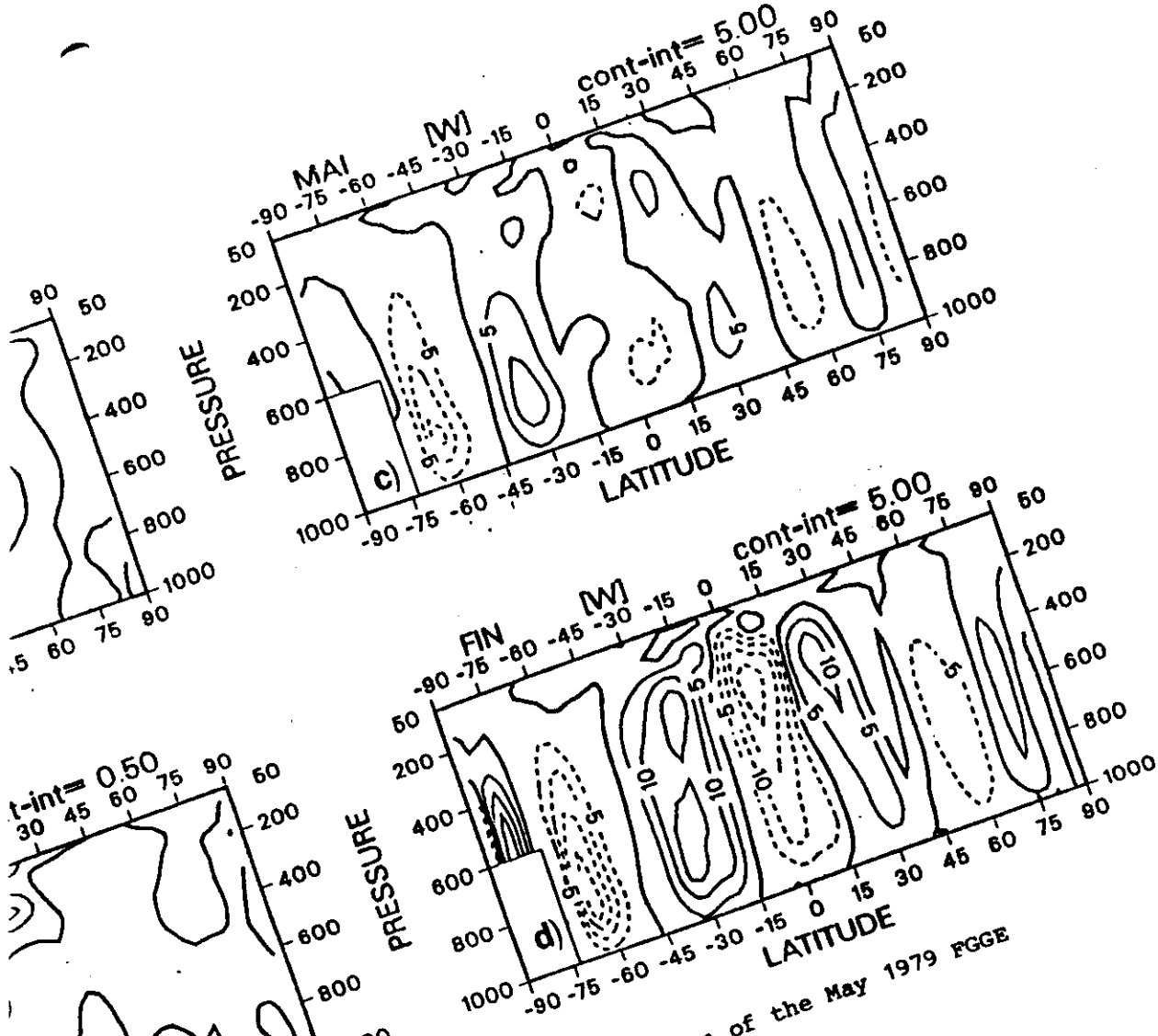
i) Failure of the analysis scheme to get the analysis close to the slow manifold. This can either be caused by erroneous data or by the numerous simplifications in the analysis method.

ii) Weaknesses in the model formulation, which would result in initialisation changes even if point i) above would not apply.

iii) Errors in the initialisation scheme itself.

Usually, all three areas contribute to initialisation changes and it often is not easy to isolate a particular process.

A typical example for point iii) is the suppression of diabatically driven circulations when using an adiabatic initialisation scheme. Many early implementations of the normal mode method suffered from this problem. As an example, Fig. 10a shows the monthly mean, initialised, zonally averaged, meridional wind component for 12Z, May 1979 as calculated from the "Main" FGGE analyses produced by ECMWF. The adiabatic initialisation, which was used for the "Main" production, almost entirely wiped out the upper tropospheric return flow of the Hadley cell. Also, the tropical vertical velocities are very small (Fig. 10c) when using an adiabatic initialisation scheme. By contrast, the "Final" assimilation produces a much more vigorous Hadley cell (Figs. 10b,d). The upper tropospheric return flow has increased from 0.5 to 2m/sec and the vertical velocity from .005 to .02 pa/sec in the tropical troposphere. Although the assimilation system used for the "Final" runs differs in many respects from the early version used for the "Main"



10 Zonally averaged meridional flow from the mean of the May 1979 FGGE analyses produced by ECMWF.
(a) "Main" ECMWF FGGE analyses, meridional wind;
(b) "Final" ECMWF FGGE analyses, meridional wind;
(c) as (a), but for vertical velocity;
(d) as (b), but for vertical velocity;

production, (Uppala, 1986) the inclusion of diabatic tendencies is mainly responsible for the strengthening of the Hadley cell. This was confirmed by independent tests (Wergen, 1983).

Another indication of the intensification of the divergent circulation is shown in Fig. 11, which gives the monthly mean velocity potential field at 200 mb for May 1979. The first two panels are taken from Lau, 1984. They show the uninitialised fields as produced by GFDL (top) and ECMWF (middle) for the "Main" runs. Generally, the ECMWF divergent flow is smoother but weaker than the GFDL field. On the other hand, the ECWMF "Final" field (Fig. 11c) shows a considerably increased divergent circulation; especially over Indonesia and South-America. Note that the "Final" ECMWF field (Fig. 11c) is initialised. Thus the initialised "Final" velocity potential shows stronger gradients than both the "GFDL" and "ECMWF" "Main" uninitialised fields.

It is interesting to compare the changes the analysis makes to the first guess to the changes made to the analysis by the initialisation. Ideally, the initialisation increment should be smaller than the analysis increment. Otherwise, either the analysis increment is unbalanced or the initialisation scheme is defective. For the northern hemisphere 500 mb height field, Hollingsworth et al. (1986) demonstrated that the major part of the evolution is described by the forecast model within the ECMWF data assimilation system. The analysis increments are generally smaller and the initialisation changes are smaller still. Fig. 12 gives the global seasonal mean analysis (top) and initialisation (bottom) increment for the 200 mb wind averaged for 00Z of June, July and August 1984. The mean analysis increments are generally small in the northern extra-tropics. However, in the tropics and in the southern hemisphere values around 5m/sec can be found. The initialisation increment is much smaller, generally of the order 1m/sec. This suggests, that most of the

information contained in the data has been accepted by the initialisation. This also applies for the tropics, where the initialisation increment is frequently a mirror image of the analysis increment if an adiabatic initialisation is used. The inclusion of diabatic tendencies in the initialisation is essential for retaining the analysed information in the tropics.

The most critical variable with respect to initialisation changes is the divergence. Fig. 13 shows the June, July, August 1984 00Z mean divergence at 200 mb for the uninitialised (top) and initialised (bottom) fields over Indonesia. The patterns are very similar. Extremes are slightly reduced in the initialised fields. There are also some areas where the initialisation results in a modest increase in amplitude. From this it would seem that neither the analysis nor the initialisation any longer suffer from systematic deficiencies, at least not in the upper tropospheric wind field.

Next we shall look at the vertical structure of the initialised fields. For this, we show zonal cross-sections rather than zonal averages. While zonal averages are useful to highlight gross features, they are prone to sampling errors for variables such as the meridional and vertical velocity, which frequently change sign in the zonal direction. In a zonal average day and night time as well as continental and maritime regimes of circulation are all lumped together.

Fig. 14 shows a longitude-pressure cross section of the meridional average (20°N-5°N) of the initialised divergence (14a) and vertical velocity (14b) for 00Z, Summer 1984 between 65 and 180E. The most prominent feature in the divergence field is the strong maximum associated with convective outflow concentrated around 150 hPa. In the boundary layer there is strong

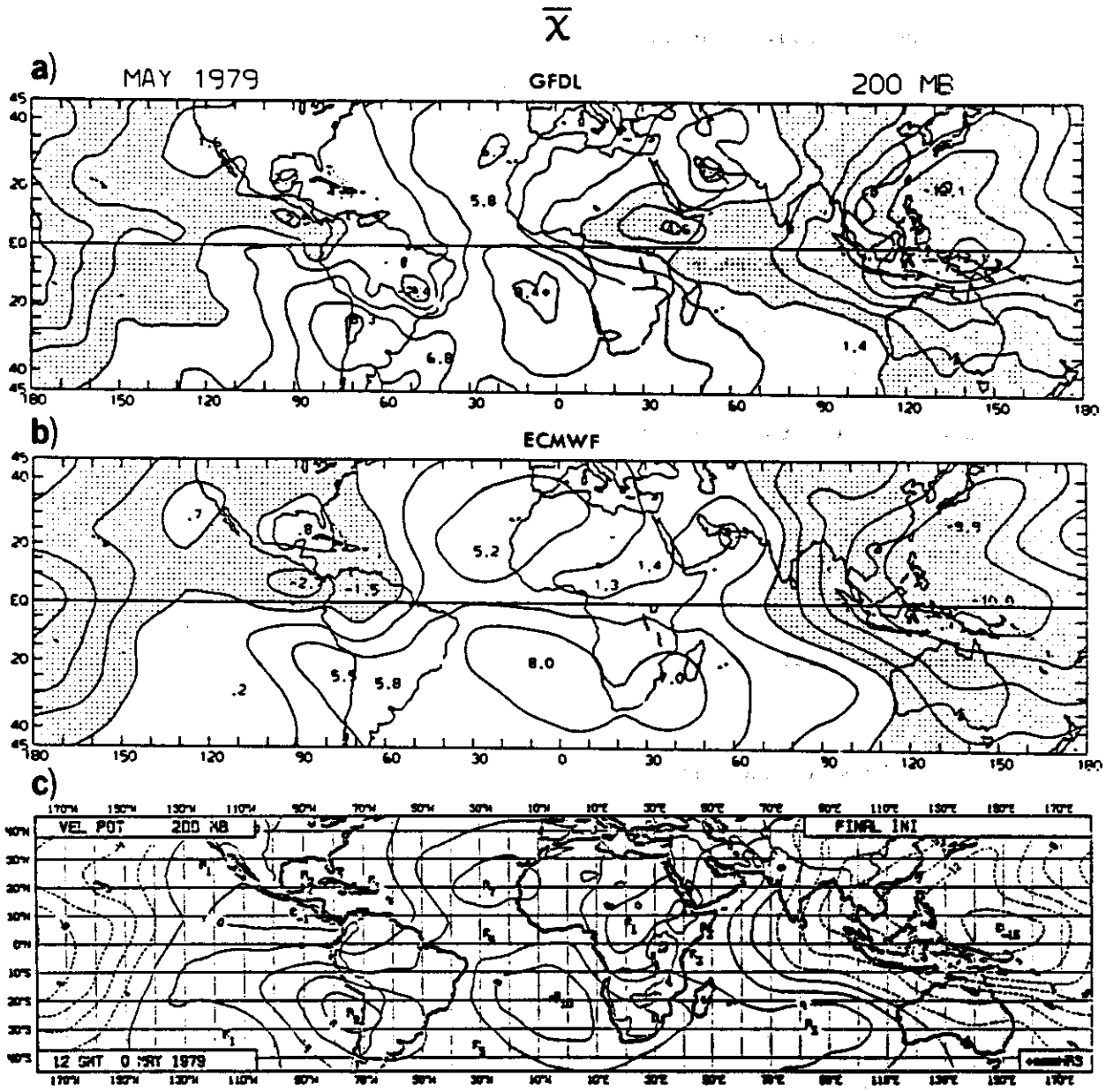


Fig. 11 Monthly mean velocity potential at 200 hPa for May 1979.
(a) Uninitialised fields as produced by GFDL (From Lau, 1984)
(b) Uninitialised fields as produced by ECMWF ("Main", also from Lau, 1984).
(c) Uninitialised fields as produced by ECMWF ("Final").

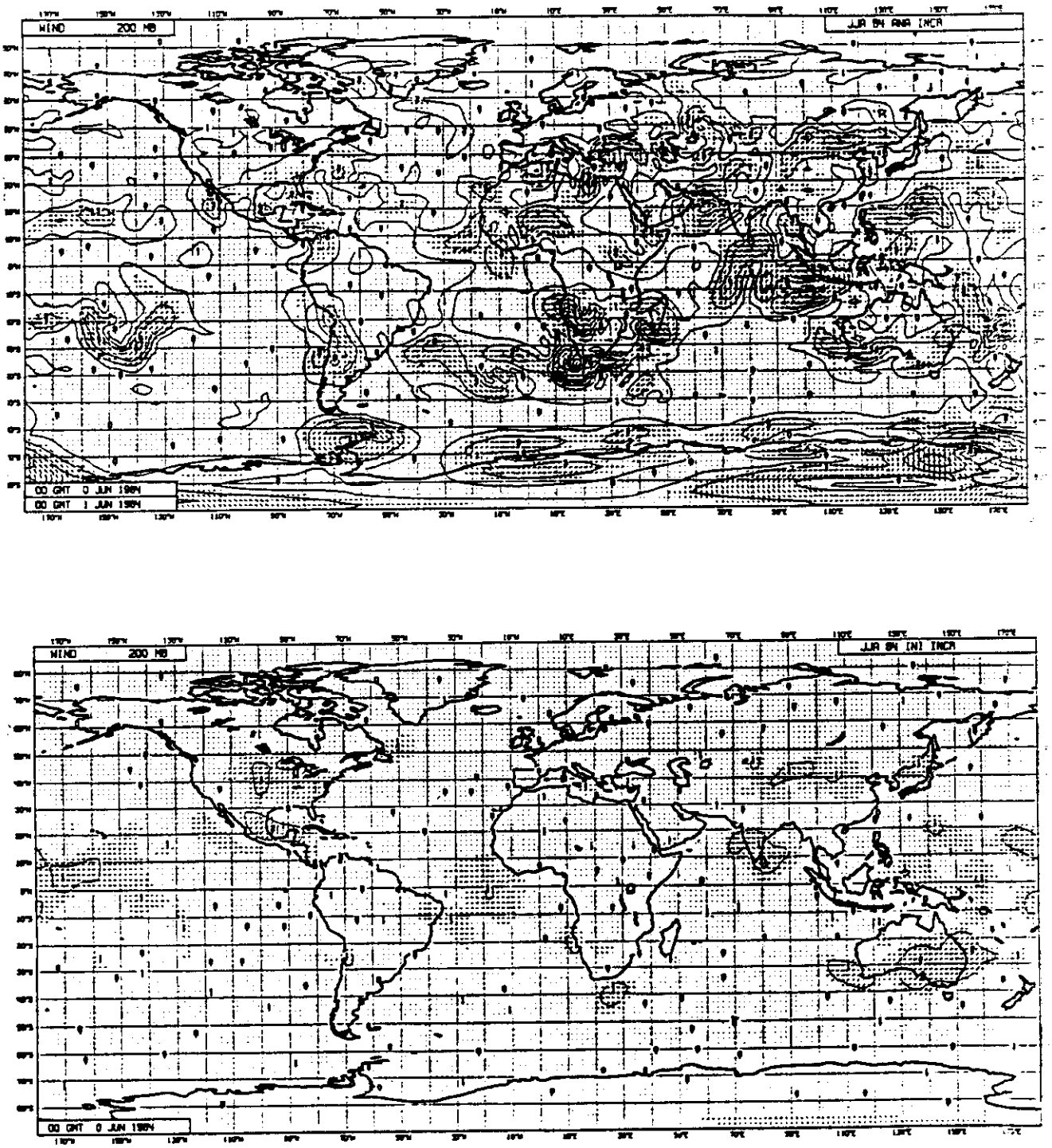
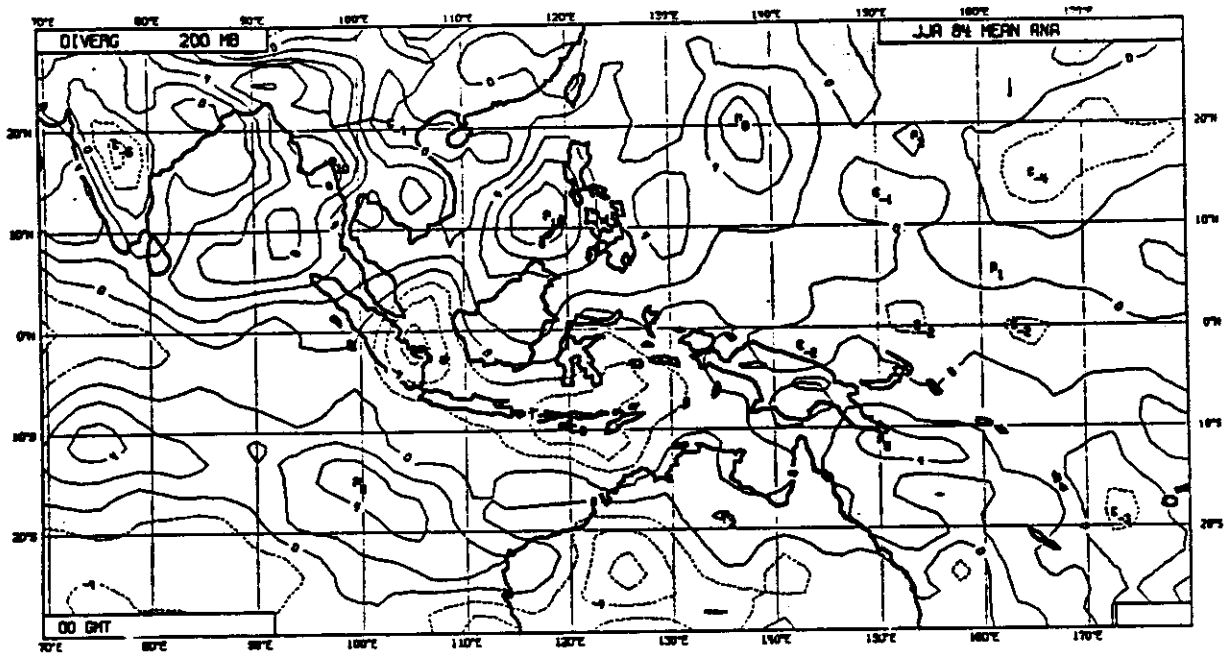


Fig. 12 Seasonal mean (June, July, August) analysis (top) and initialisation (bottom) increments for the 200 hPa wind field at 00Z.



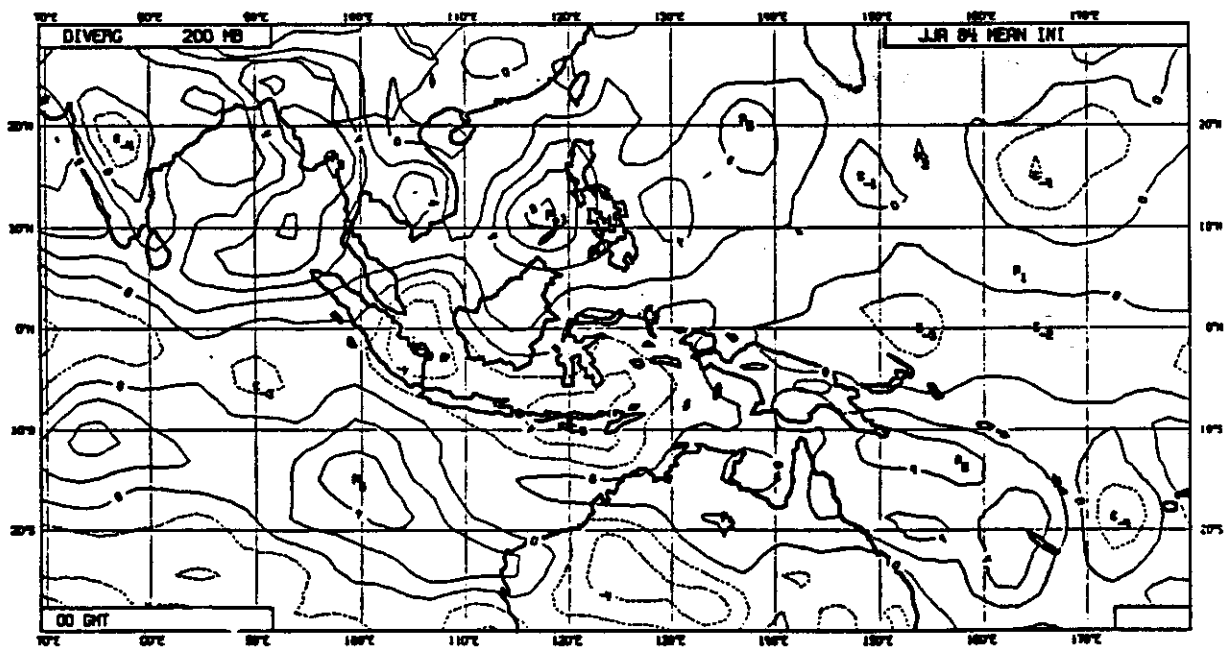


Fig. 13 Seasonal mean divergence field (June, July, August, 00Z) at 200 hPa for the uninitialised (top) and initialised (bottom) analyses.

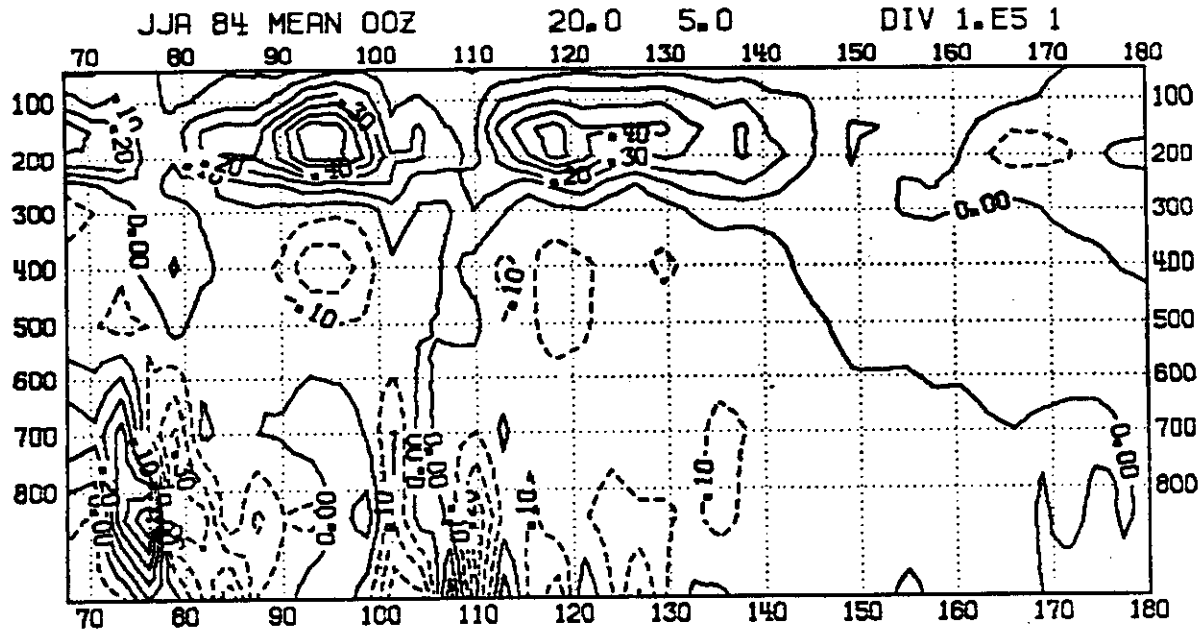


Fig. 14 (a)

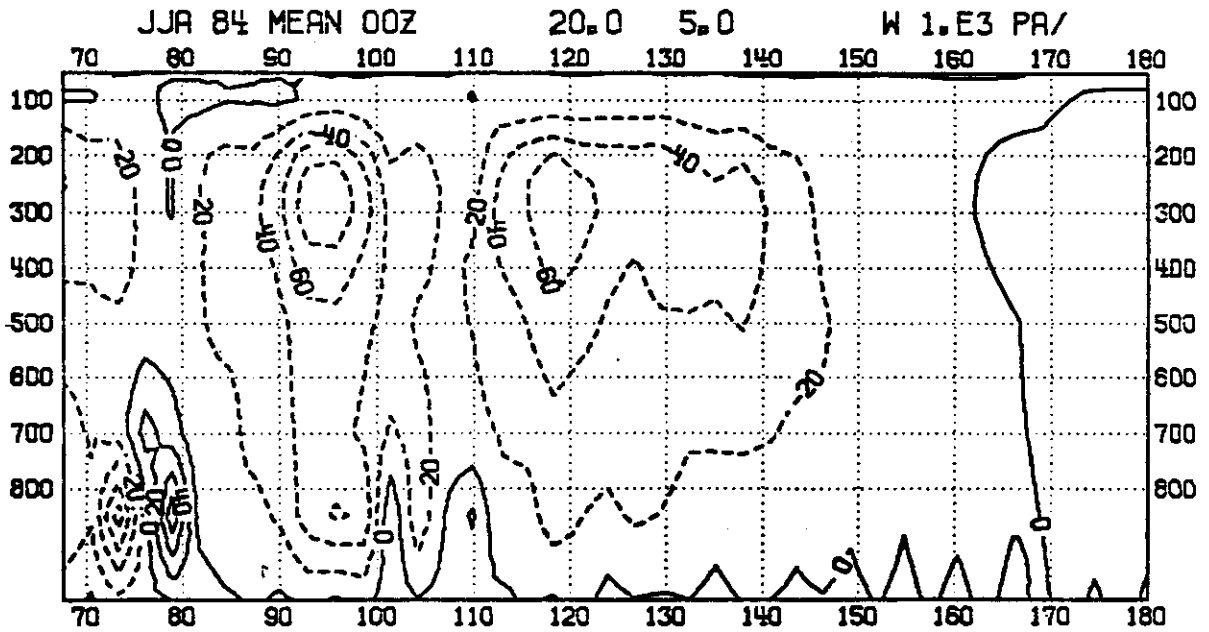


Fig. 14 Longitude-pressure cross section of the meridional average (20°N-5°N) of the initialised divergence (14a) and vertical velocity (14b), seasonal mean (June, July, August, 1984, 00Z), 65°E to 180°E.

convergence. A secondary maximum of convergence can be found around the 400 hPa level between 90°E and 100°E and 110°E and 125°E. This vertical structure implies a maximum vertical velocity at 300 hPa (Fig. 14b). Between 90°E and 100°E there is a secondary maximum in ω around 850 hPa. Although this complicated vertical structure disagrees with some long-held views of the tropical atmosphere (Newell et al., 1974) there is observational evidence from special observing campaigns which support the results in Fig. 14. For instance, Thompson et al. (1979) reported a secondary maximum of convergence around the 400 hPa level for the GATE dataset. Johnson and Young (1982) (using winter MONEX data) computed a ω profile with a maximum at 450 hPa for 08 local time and at 300 hPa for 14 local. Both profiles show a small secondary maximum around 900 hPa. The observed amplitudes agree very well with the values given in Fig. 14.

Despite the agreement between the initialised fields and the calculations based directly on observations, there remain some reservations about the detailed structure of the initialised tropical fields. One is the quality of the diabatic heating field used in the initialisation. Any deficiency in the parameterisation package will be reflected in the initialised fields to some extent. Experiments with a different convective parameterisation (Betts and Miller, 1984) indicate, however, that the differences are small. This is partly because of the horizontal filtering of the physical tendencies, which retains only the more reliable large scale components. Another problem concerns the uncertainties in using single level data in the analysis scheme. For example, it is not clear how best to spread the information from cloud track winds vertically. There is a problem with the height assignment for this data. These reports also produce systematic under-estimates of wind speed (Kallberg 1985). Nevertheless, they are almost the only source of information about the wind field in large areas of the tropics.

As only a limited number of vertical modes are initialised, spurious structures in the vertical can also be introduced by the initialisation scheme. Operationally only the first 5 vertical modes are initialised at ECMWF. Therefore, "noise" contained in the higher modes remains uncontrolled. When transforming from the vertical normal mode space back to physical space this can lead to some arbitrary structures.

4.3 FORECAST IMPACT

While closeness of the initialised and uninitialised analysis and the effective noise control are desirable features of an initialisation scheme, they do not necessarily guarantee satisfactory performance of the scheme within a data assimilation system. For instance, initialisation changes can be substantially reduced by initialising fewer horizontal and/or vertical modes. This does, however, not necessarily mean that the ensuing forecasts are better, as they might be contaminated with noise from the uninitialised modes. Similarly, the forecasts can also suffer from excessive noise control; e.g. by not making a proper distinction between Rossby and gravity modes when applying a scheme which is only based on frequency.

Therefore, a thorough testing of any modification can only be done within a data assimilation scheme. The most critical parameter is the quality of the short range forecast used as a first-guess for the next analysis. Ideally, any improvement in the initialisation should be reflected in the ensuing forecast. The improvement in the first guess will be directly reflected in the quality of the next analysis through better quality control decisions. The easiest way to test the quality of initialisation and first guess fields is to compare them to the observations after erroneous reports have been eliminated. Fig. 15 uses the fit of the winds to tropical radiosonde data to compare the RMS fit of the initialised analysis (left) and of the 6 hour first

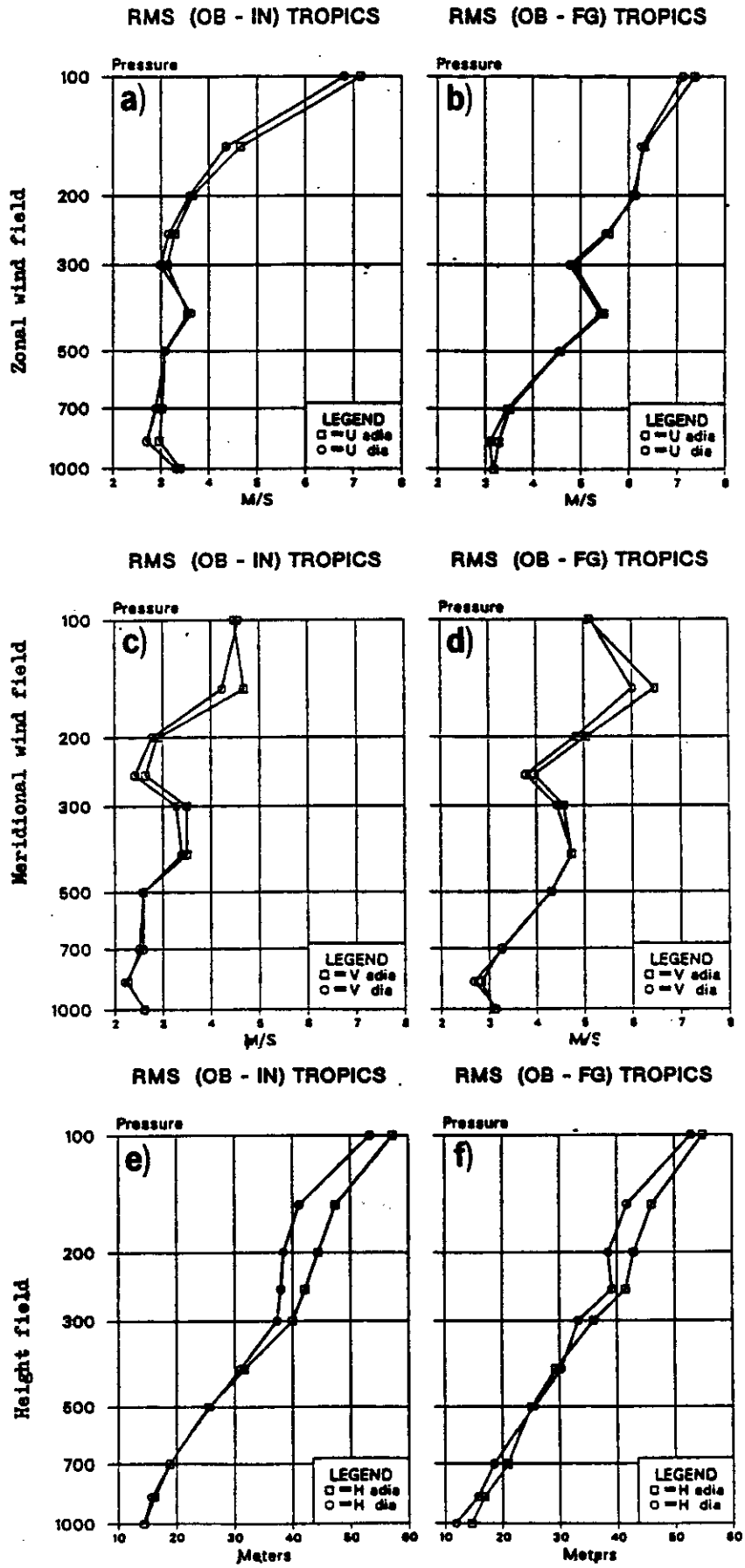


Fig. 15 Fit to tropical radiosonde data of the initialised analysis (left) and the 6 hour first guess field (right).

guess (right) for the zonal (a,b) and meridional (c,d) wind components and for the height field (e,f). On average, about 120 reports were available at each analysis time. The curves with squares are for the adiabatically initialised analysis, those with circles are for the diabatic initialisation. The results are valid for the 00Z fields, averaged between 26.12.78 and 28.12.78. Quite clearly, the diabatically initialised fields are closer to the observed values than the adiabatic fields. More importantly, the improvement is maintained through the 6 hour forecast. This is demonstrated by the closer fit of the forecast started from the diabatically initialised analysis. The signal in the u-component is somewhat obscured by the rapid growth of forecast errors both in Rossby and gravity modes, in the middle atmosphere. Nevertheless, a positive impact can be found in the boundary layer and close to the tropopause. For the v-component and for the height field most of the improvement in the initialisation is reflected in the forecast.

There is also a positive impact on the forecasts beyond 6 hours. This is demonstrated in Fig. 16 which shows monthly mean operational forecast scores before and after the introduction of the diabatic initialisation on 21.9.82. The light full and heavy lines gives the absolute correlation of vector wind at 200 hPa in the tropics for day 1 and day 2. During September 1982 the scores improved by about 3%. The dashed curve gives the northern hemisphere anomaly correlation of 500 hPa height for day 2. It also shows a slight improvement during September 1982. As the forecasts are verified against the initialised analyses, part of the improvement is due to a more realistic verifying analysis.

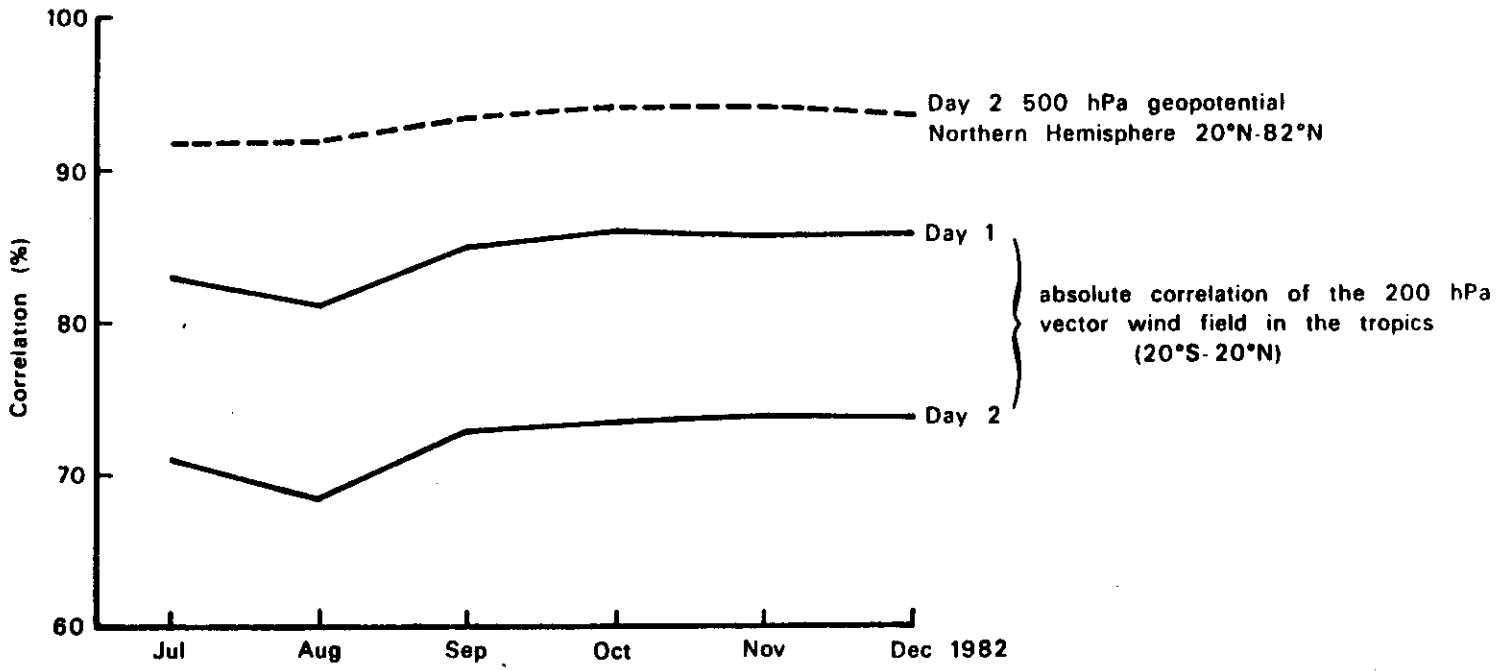


Fig. 16 Monthly mean operational forecast scores before and after the introduction of the diabatic initialisation on 21.9.82.

4.4 THE SPIN-UP PROBLEM

A matter of ongoing concern is the so-called "spin-up" problem arising from the length of time it takes the model to fully develop a divergent circulation and to reach a balanced hydrological budget, this problem and some of its implications for the tropical forecasts is discussed in Heckley (1985). It is sometimes argued that the initialisation negatively influences the spin-up. The following two figures are meant to shed some light into this problem. They show the convective rain accumulated between 00 and 12 hours and between 24 and 36 hours into the forecast. The forecasts were started from the initialised and from the uninitialised analyses. Fig. 17 is for the Indonesian area. During the first 12 hours the precipitation patterns are already well established for the initialised forecast (Fig.17a). The corresponding forecast started from the uninitialised analysis (Fig. 17b) shows a similar pattern, although the rain-band extending northeast from the Philippines is less well organised. Later in the forecast, the precipitation pattern is similar to the earlier pattern in each forecast respectively (Fig. 17c,d), thus indicating that the precipitation was already well spun-up during the first 12 hours of the forecast. The picture is fundamentally different for the east Pacific. Initially, there is little precipitation in the ITCZ area, both with initialised (Fig. 18a) and in the uninitialised (Fig. 18b) forecast. However, 24 hours later, the convective precipitation along the ITCZ is well established. Qualitatively, there is little difference between uninitialised and initialised forecasts.

From this it would seem that the spin-up problem is more severe in some areas than in others. Compared to this, the impact of the initialisation is small. At present, it is not clear what causes the large variability. Apart from differences in flow type, differences in data coverage could be responsible.

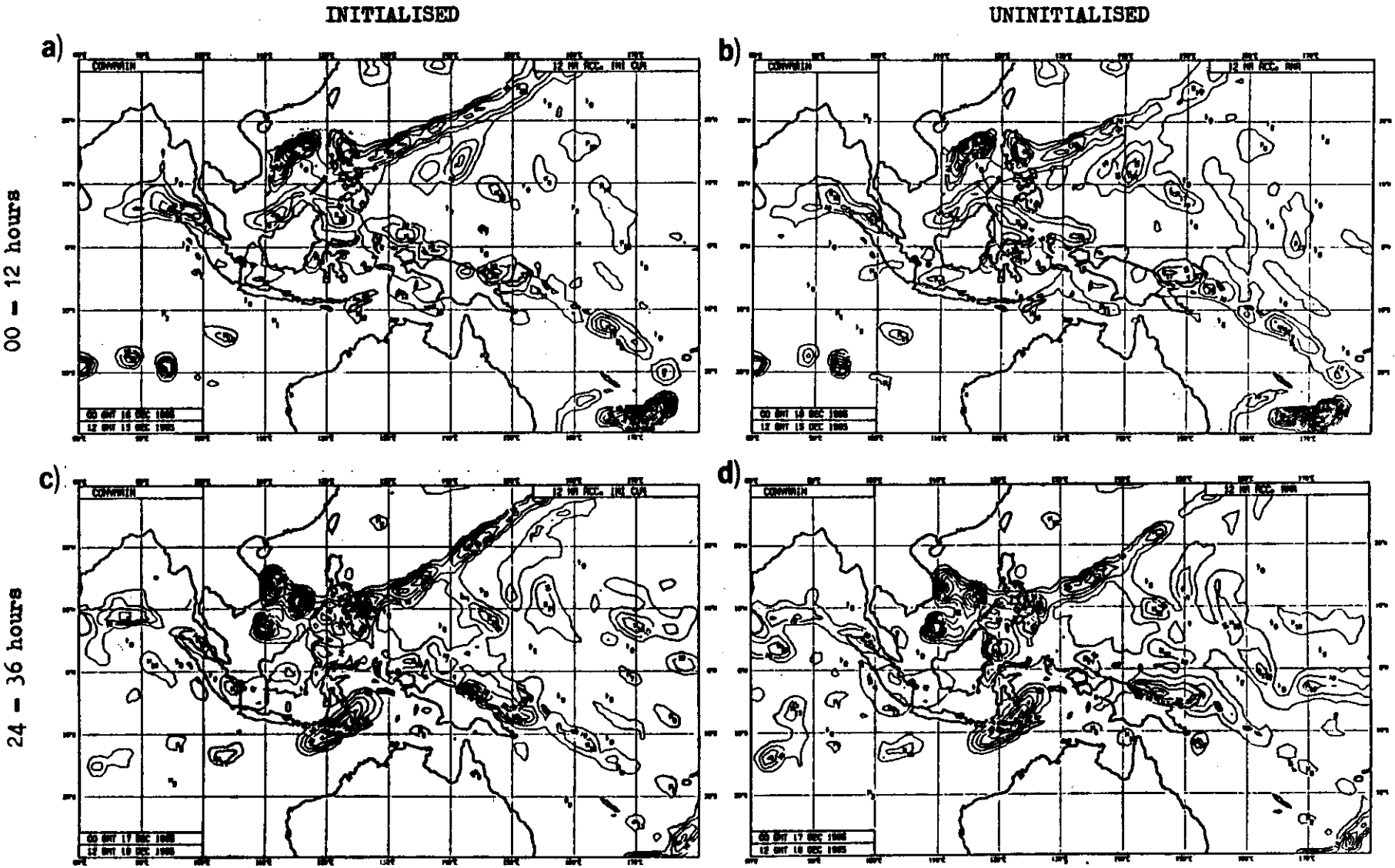


Fig. 17 Convective rain accumulated between 00 and 12 hours (a,b) and between 24 and 36 hours (c,d) for forecasts started from initialised (a,c) and uninitialised (b,d) analyses.

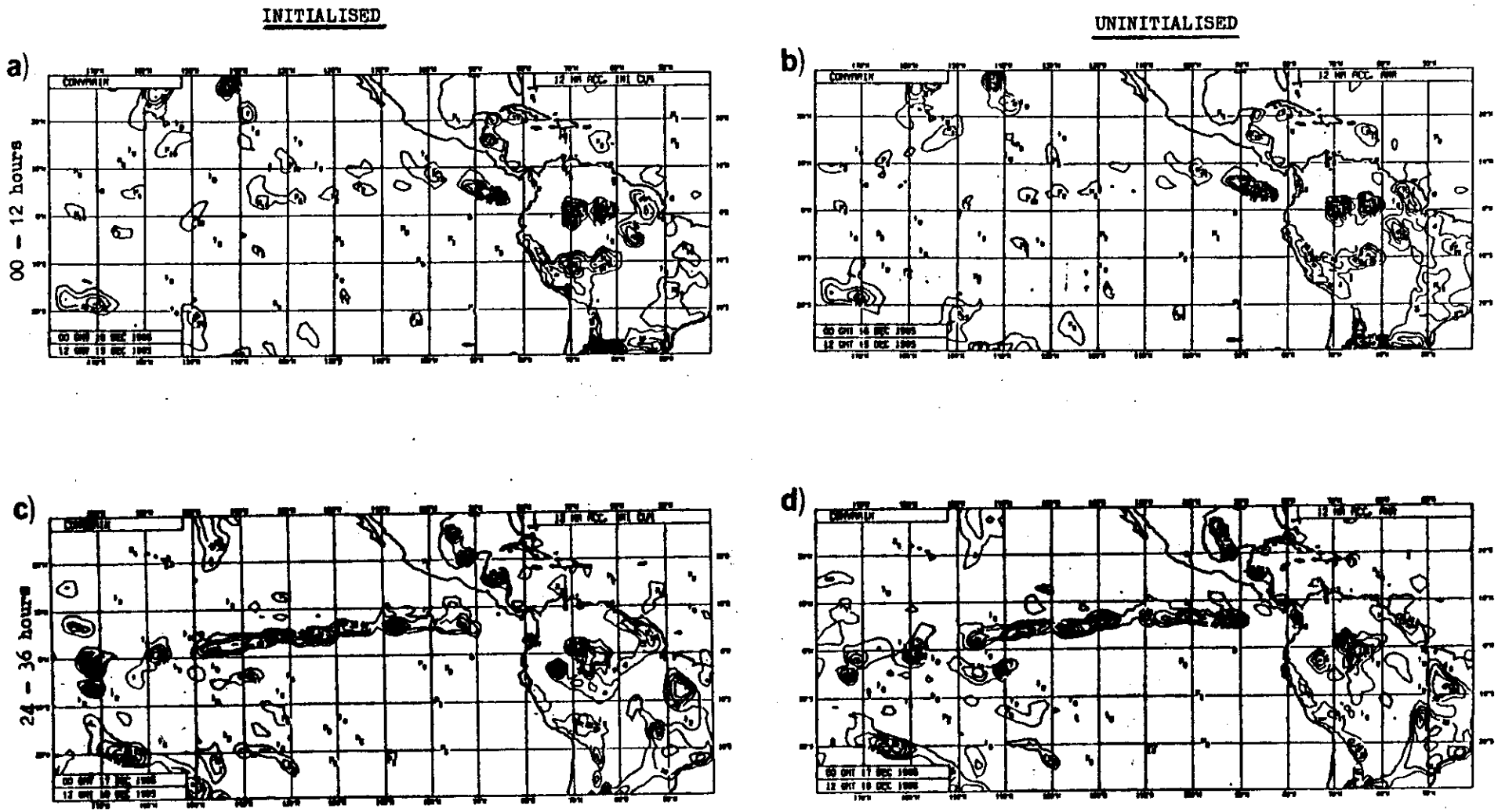


Fig. 18 Convective rain accumulated between 00 and 12 hours (a,b) and between 24 and 36 hours (c,d) for forecasts started from initialised (a,c) and uninitialised (b,d) analyses.

In the east Pacific area the analysis mainly relies on satellite data. Cloud track winds are very difficult to use because they are single level data and they suffer from bias and height assignment problems (Kallberg, 1985). Furthermore, it is not clear whether the temperature retrieval from satellites are accurate enough and have sufficient vertical resolution to consistently correct the first guess. On the other hand, the data coverage in the Indonesian area is quite different as there are a number of radiosondes and numerous surface reports which are likely to lead to a better definition of the three-dimensional correction field.

Quite generally, the demands on an analysis scheme are very subtle in the context of data assimilation and the spin-up problem in particular. Apart from the requirements to fit the basic fields such as mass and wind, the parameterisation packages expect a consistent correction of a combination of derived quantities. For instance, the Kuo convection scheme depends on a precise specification of the initial moisture convergence together with the matching static stability. The boundary layer scheme requires a consistent definition of vertical wind shear, and stability. Without a dense, three-dimensional, high quality data coverage, analysis schemes are likely to upset some of these internal model balances. This is not to mention weaknesses in the parameterisation schemes themselves.

4.5 THE TIDAL PROBLEM

As discussed in Section 3.2 the initialisation condition, $\dot{y}=0$, is not appropriate for the atmospheric tides. Instead requiring stationarity for the gravity mode projection of the tidal signal, it should be allowed to propagate westward with the movement of the sun. The Machenhauer condition is only appropriate if the forcing is quasi-stationary. This is clearly not the case if the radiation scheme simulates the diurnal cycle of the sun. Also, the observational data reflect the presence of tidal activity in the analysis and thus imply transient tidal components in the model tendencies.

Atmospheric tides are mostly thermally excited by the absorption of solar radiation by ozone and water vapour. Mainly zonal wavenumber 1 and 2 modes are excited. Due to the vertical distribution of the absorbers, the diurnal zonal wavenumber 1 pressure wave is largely trapped, whereas the semi-diurnal wave can propagate vertically (Chapman and Lindzen, 1970). Therefore, the semi-diurnal wave dominates in the surface pressure field. Chapman and Lindzen (1970) quote amplitudes of 1.22 hPa and .22 hPa for the dominant, first two gravest symmetric spherical harmonics of the semi-diurnal surface pressure oscillation. For the corresponding harmonics of the diurnal surface wave they give .46 and .21 hPa.

The problem is overcome in the ECMWF system by excluding the tidal signal from the initialisation. The tidal signal is estimated from a time series analysis of the previous ten days. The gravity model projection of the tidal tendencies is then excluded from the initialisation.

The scheme has been evaluated in a 5 day data assimilation. Fig. 19 shows the RMS fit of the initialised (left) and of the first guess (right) height field averaged over the last 3 days of the assimilation. Squares depict the 'standard' scheme, circles are for the test with the special handling of the tides in the initialisation. Excluding the tides from the initialisation improves the initialisation height field RMS fit the data by about 5m. The improvement becomes larger in the upper troposphere. In the six hour forecasts, most of the improvement in the lower troposphere is maintained. Around the 300 hPa level the positive impact becomes smaller, but on the whole the exclusion of the tides from the initialisation results in a better tropical surface pressure forecast.

The impact of the modification on the evolution of the predicted tropical surface pressure is shown in Fig. 20. It gives traces for 24 hours forecasts started from the uninitialised analysis (full), from the standard diabatic initialisation (dashed-dotted) and from the initialisation with the special handling of the tides (dashed). At .93°N and 11°W (fig. 20a) the tidal initialisation is initially closer to the uninitialised analysis and captures the semi-diurnal pressure variation better than the standard initialisation; especially in the early stages of the forecast. It does not, however, suffer from the noise problems of the uninitialised analysis. Similar conclusions hold for a point in the tropical Pacific (Fig. 20b) which again shows a more pronounced semi-diurnal pressure wave for the "tidal" initialisation. It is instructive to compare the predicted surface pressure evolution to the observed values. Fig. 21 shows the corresponding curves for the station Niamey in Niger (13°N, 2°E) (Humphreys, Pers. comm). The model captures most of the semi-diurnal oscillations and of the superimposed longer term trend.

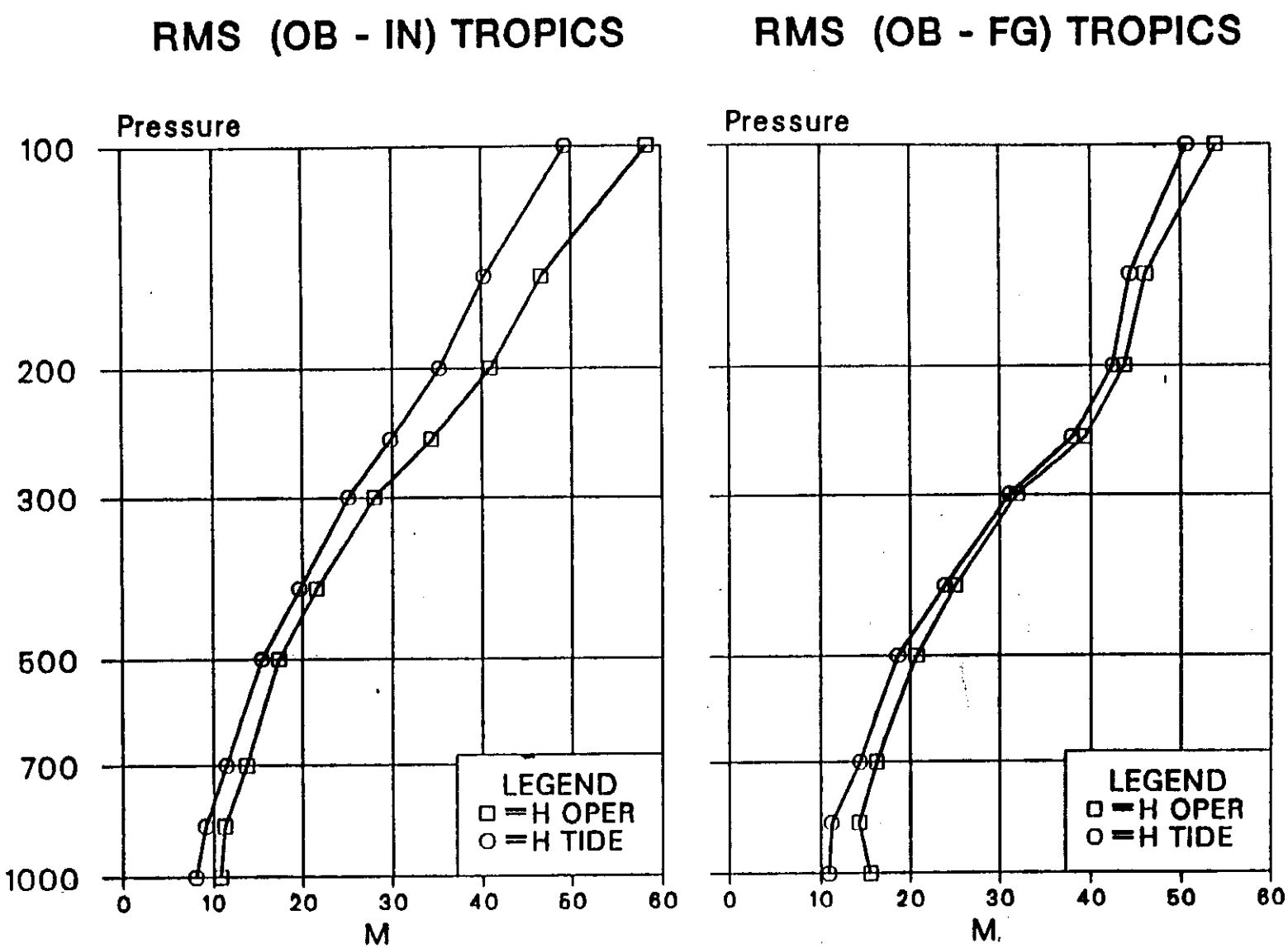


Fig. 19 RMS fit of initialised (left) and first guess (right) height fields to radiosonde observations averaged over three days of assimilation. Square symbols: tides not excluded. Round symbols: tides excluded.

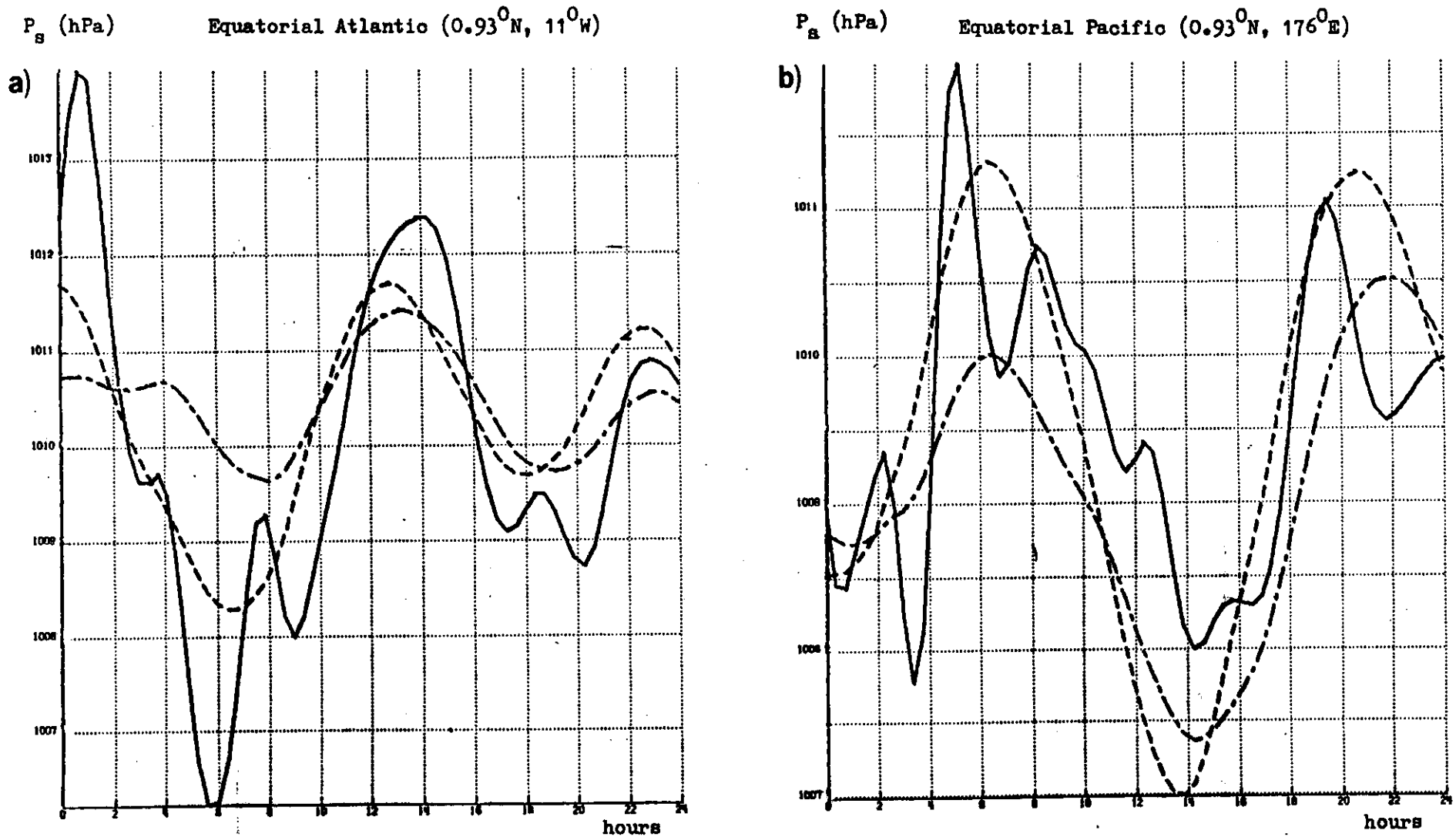


Fig. 20 Surface pressure trace for 24 hour forecasts started from uninitialised analysis - full line, diabatic initialisation, tides included - dashed-dotted line, diabatic initialisation, tides excluded - dashed line

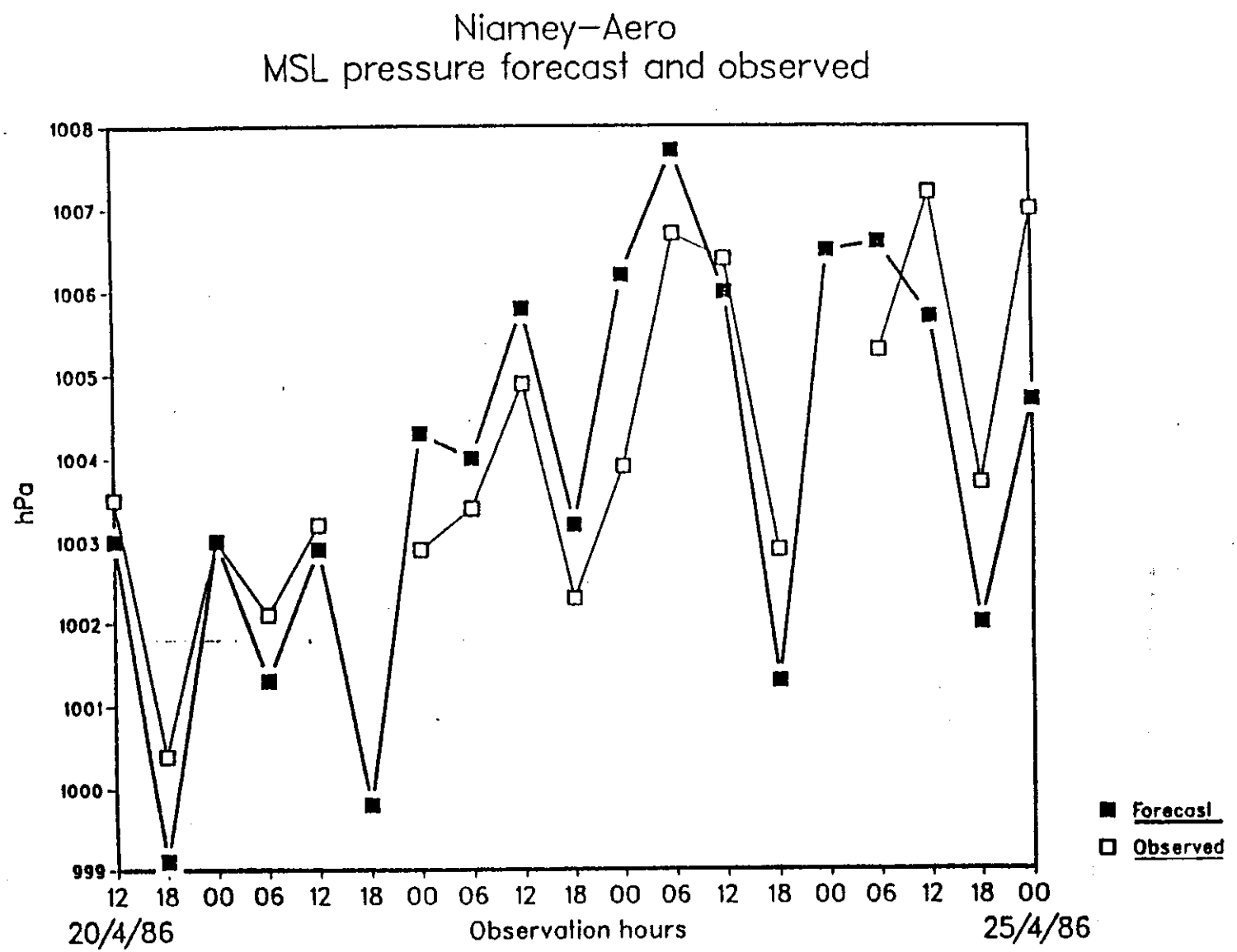


Fig. 21 Surface pressure trace for 24 hour forecast started from initialised fields (diabatic, tides excluded), together with observed values for Niamey in Niger (13°N, 2°E).

The assimilation system benefits from the improved surface pressure forecasts mainly through a reduced need for corrections to the mass field. Most of the mass information over the tropical continents is from SYNOP reports. While it is easy to use them to correct the surface pressure, there are large uncertainties as to how to spread the information vertically. By improving the first guess this problem can be alleviated.

5. INITIALISATION IN A WIDER CONTEXT

Some of the problems discussed in Section 4.4 in the content of model 'spin-up' suggest the need to achieve an initial state which is not only compatible with observational data and in dynamical balance; but which at the same time is balanced with the models convective and radiative parameterisation. Thus requiring the model to have a reasonable hydrological and energy balance right at the start of the forecast.

The normal mode initialisation, the subject of earlier sections, imposes a dynamical balance between the mass and wind fields. The humidity field only plays a very indirect role, appearing through the diabatic forcing in the non-linear terms. The humidity field is unchanged by the initialisation. Yet subtle changes in the field have an effect on the hydrological and energy balances through its role in the parameterisation of precipitation and radiative effects.

Krishnamurti et al. (1983) describe one possible approach to this problem. Their efforts were simulated by prediction experiments which yielded unsatisfactory forecasts of the divergent wind field after only two days. Their "physical initialisation" process takes as its starting point the uninitialised mass, wind and humidity analysis. A precipitation intensity analysis is then produced, based on a combination of raingauge information and satellite (TIROS-N) infrared radiance data (these analyses were performed on FGGE/MONEX data bases, for which substantial raingauge returns were available). A distinction is then made between the precipitating and non-precipitating regions. Where rain is falling (ie in areas of marked convergence/divergence) it is judged desirable to enhance the divergent flow of the basic wind analysis. This is done by adjusting the divergence (and

hence vertical velocity) profiles so that the convective parameterisation scheme yields equivalent rainfall rates. The divergence fields in the non-precipitating areas remain unchanged, as does the rotational flow over the entire domain. At the same time, in the non-precipitating areas, the humidity profiles are adjusted to yield a balance between advective and radiative forcing above the PBL. This was considered necessary because the deterioration in the forecast divergences, evident in rain-free as well as convectively active regions, was judged to stem in part from errors in the specified humidities in these rainfree areas. The adjustments of the humidity profiles are constrained to preserve the total precipitation water content within a vertical column. Dynamic initialisation is then applied to these fields. This will produce some changes in the dynamical structures, particularly in sensitive indicators such as vertical velocity. Having performed this dynamical initialisation, it is undesirable to seek subsequent adjustment of the mass and wind fields, but the humidity field is largely passive and can be adjusted to achieve further physical balance with the model's diabatic processes. In precipitating areas the humidity is adjusted to yield rainfall rates compatible with the analysed rainfall rate, the prescribed vertical velocities and the convective parameterisation; in non-precipitating areas the humidity is again adjusted to yield the same type of balance between advection and radiative forcing as described before. These adjustment processes yield fields which are in approximate, rather than precise, physical balance. Fig. 22, after Krishnamurti et al (1983), shows an example of an 850 mb specific humidity field (a) before and (b) after the full initialisation sequence. The broad structure has been preserved, but with notable changes in some localities. Fig. 23, after the same authors, shows the rainfall rates from (a) the (raingauge plus satellite data) analysis and (b) the initialised model. A reasonable delineation of precipitating and non-precipitating regions has been achieved by the physical initialisation.

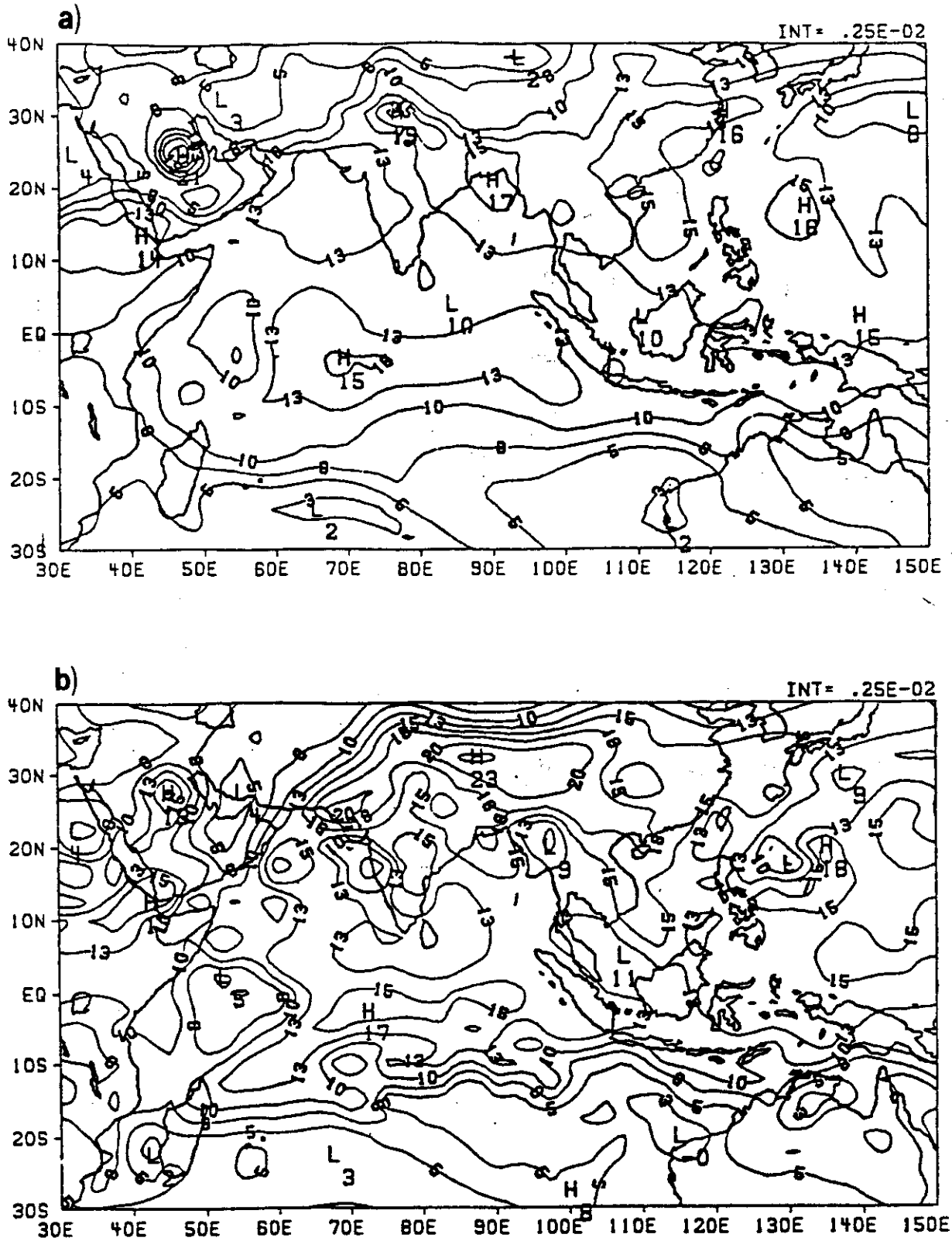


Fig. 22 850 mb Specific Humidity field (a) before (b) after physical initialisation. After Krishnamurti et al (1983).

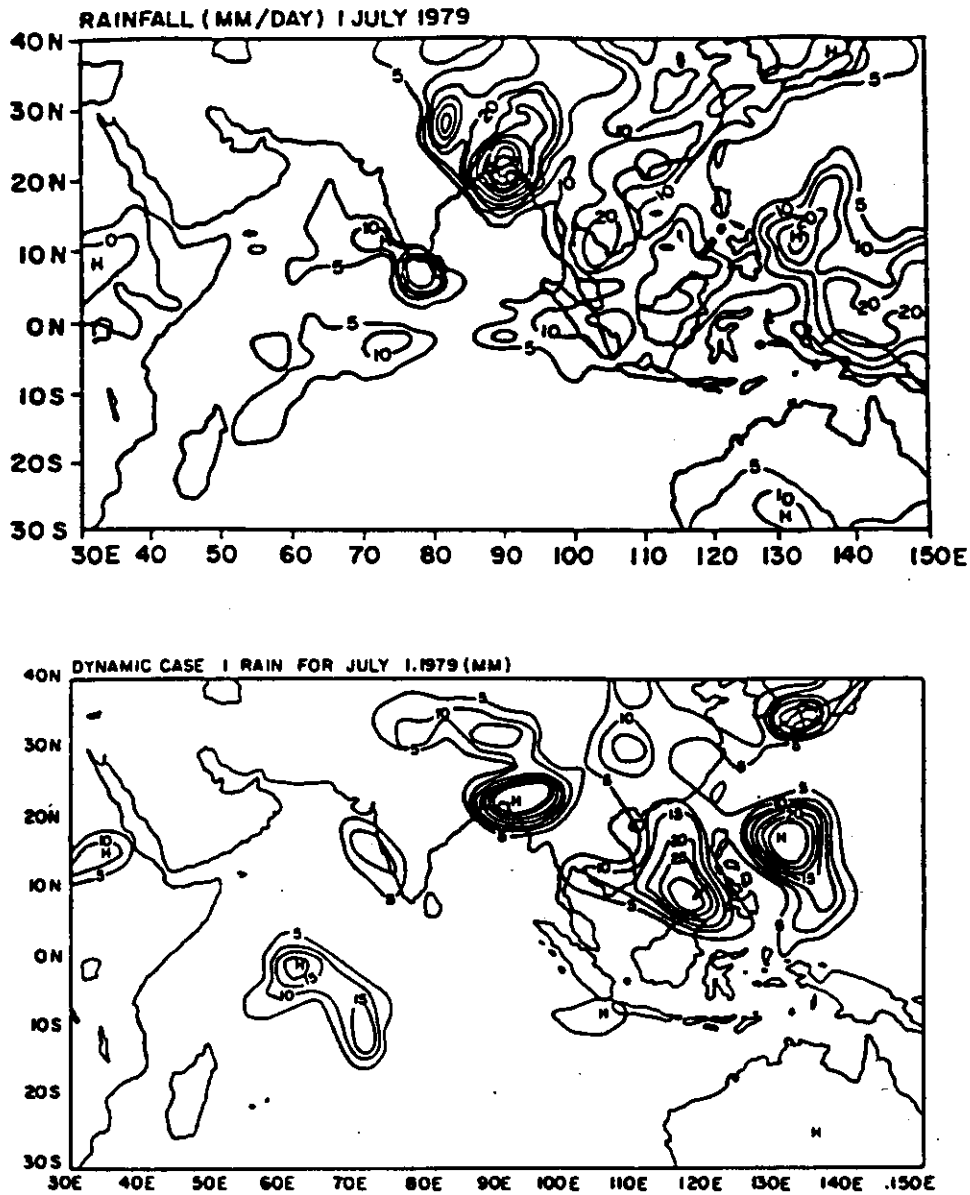


Fig 23 Rainfall totals (mm/day) for July 1 1979 above: based on satellite and rainguage data. below: recovered by model after initialisation (after Krishnamurti et al 1983).

In the ensuing forecast, which was concerned with the development and movement of a Bay of Bengal depression, it was found that the combined physical/dynamical initialisation yielded a better forecast than did a purely dynamical initialisation. The forecast from the improved initial state is judged to be reasonable out to day 6, both in its day to day movement of the feature and in its prediction of the larger scale, time-meaned, velocity potentials.

The relevance of such a procedure in an operational context is not as yet clear, since the technique requires a good estimate of the precipitation rates at analysis time and also of the initial divergence field. Both of which present considerable analysis problems. Nevertheless such techniques are being currently evaluated in many operational and research centres.

Acknowledgements

Thanks are due to Dr. R.W. Riddaway who provided many useful suggestions for revising the course.

Further Reading

A useful review of the technique of normal mode initialisation is given by Daley (1981). Further details of the normal mode initialisation as applied at ECMWF are given by Wergen (1987).

References

Betts, A.K. and M.J. Miller, 1984: A new convective adjustment scheme. ECMWF Tech.Rep.No.43, 69pp.

Daley, R., 1981: Normal mode initialisation. Reviews of geophysics and space physics. Vol.19, No.3, pp450-468.

Heckley, W.A., 1985: Systematic errors of the ECMWF operational forecasting model in tropical regions. Quart.J.R.Met.Soc., 111, pp709-738.

Hollingsworth, A., D.B. Shaw, P. Lönnberg, L. Illari, K. Arpe and A.J.Simmons, 1986: Monitoring of observation and analysis quality by a data assimilation system. Mon.Wea.Rev., 114, No.5, pp861-879.

Johnson, R.H. and G.S. Young, 1983: Heat and moisture budgets of tropical mesoscale anvil clouds. J.Atmos.Sci., 46, pp2138-2147.

Källberg, P., 1985: On the use of cloud track wind data from FGGE in the upper troposphere. ECMWF Research Dept. Tech.Memo.No.111, 39pp. Available from ECMWF.

Krishnamurti, T.N., K, Ingles, S. Cocke, T. Kitade and R. Pasch, 1983: Details of low latitude medium range NWP using a global spectral model II. Report No. 13-11. Dept. of Meteorology. Florida State University, Tallahassee, USA.

Lau, N.-C., 1984: A comparison of circulation statistics based on level III-b analyses produced by GFDL and ECMWF for the special observing periods. NOAA. Report ERL

Machenhauer, B., 1977: On the dynamics of gravity oscillations in a shallow water model with applications to normal mode initialisation. Contrib.Atmos.Phys., 50, pp253-271.

Newell, R.E., J.W. Kidson, D.G. Vincent and G.J. Boer, 1974: The general circulation of the tropical atmosphere and interactions with extratropical latitudes. Volume 2. MIT Press, Boston, USA.

Shaw, D.B., P. Lönnberg and A.H. Hollingsworth, 1984: The 1984 revision of the ECMWF analysis system. ECMWF Tech.Memo.No.92.

Thompson, R.M., S.W. Payne, E.E. Recker and R.J. Reed, 1979: Structure and properties of synoptic-scale wave disturbances in the inter-tropical convergence zone of the eastern Atlantic. J.Atmos.Sci., 36, pp53-72.

Uppala, S., 1986: The assimilation of the final level IIB data set at ECMWF, Part I. Proceedings of the National conference on the scientific results of the first GARP global experiment. January 14-17, 1986, Miami, USA. pp24-29. Published by the American Meteorological Society, Boston, Mass.

Wergen, W., 1983: Initialisation. ECMWF Seminar/Workshop on the interpretation of numerical weather prediction products. Sept. 1982, Reading, UK, pp31-58.

Wergen, W., 1987: Diabatic non-linear normal mode initialisation for a spectral model with a hybrid vertical coordinate. ECMWF Tech.Rep.No. 59.

Williamson, D., 1976: Normal mode initialisation procedure applied to forecasts with the global shallow water equations, Mon.Wea.Rev., 104, pp195-206.