Initialisation

W.Wergen

ECMWF

1. FOREWORD

This paper is not meant to be a strict mathematical review of nonlinear normal mode initialisation. This can be found elsewhere (Daley, 1980). It rather tries to explain in simple terms what initialisation is, and why it is needed. Frequent use of examples and analogies is made. This allows a straightforward physical explanation of normal mode initialisation and of its benefits compared to other methods. After these general ideas have been established, the performance of the scheme - in its recently introduced diabatic version - is discussed.

2. WHY INITIALISATION?

Once an analysis scheme has provided initial values for the prognostic variables of the model, one is in a position to start a forecast. However, as Fig. 1 shows, the result will be quite disappointing. The surface pressure at this particular point (0°E,52°N) shows unrealistic high frequency oscillations with considerable amplitude. By contrast, if the forecast is started from the initialised analysis (dashed), the behaviour is much better. When looking at the difference (initialised-uninitialised) between the two predicted surface pressure fields after 6 hours, a marked wave pattern is immediately evident (Fig. 2). In order to cause the pressure oscillations shown in Fig.1, these waves have to travel with a speed in the order of hundred meters per second, which is the characteristic speed of gravity waves.

The crucial role of initialisation becomes most evident within a data assimilation scheme. Before the actual analysis is performed, the observations are checked against the first guess (Lönnberg,1982). If the deviations exceed

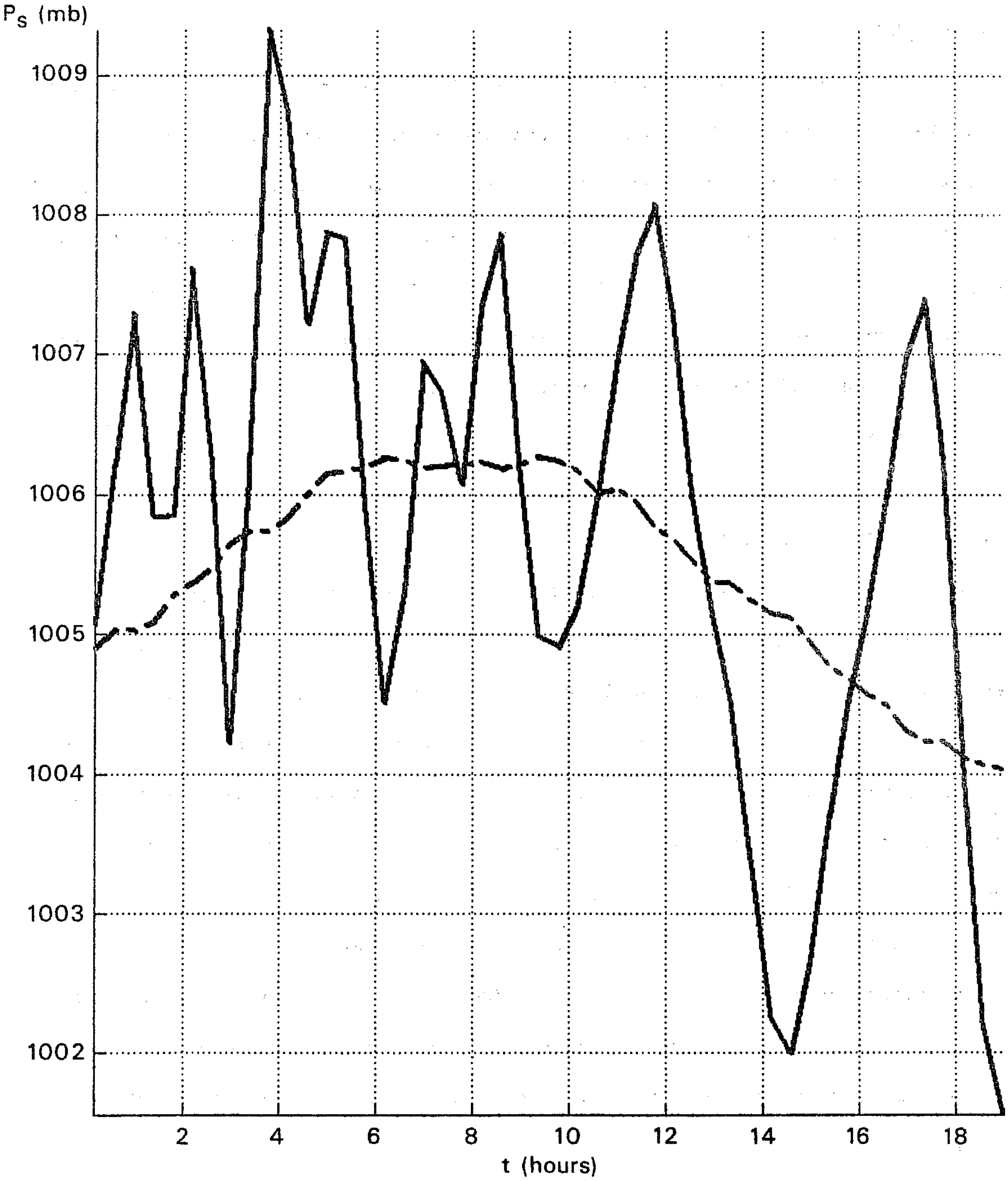

Fig. 1 Surface pressure evolution at gridpoint 0°E, 52°N for forecast started from uninitialised (full) and initialised (dashed) analysis for 12 GMT 6 September 1982.

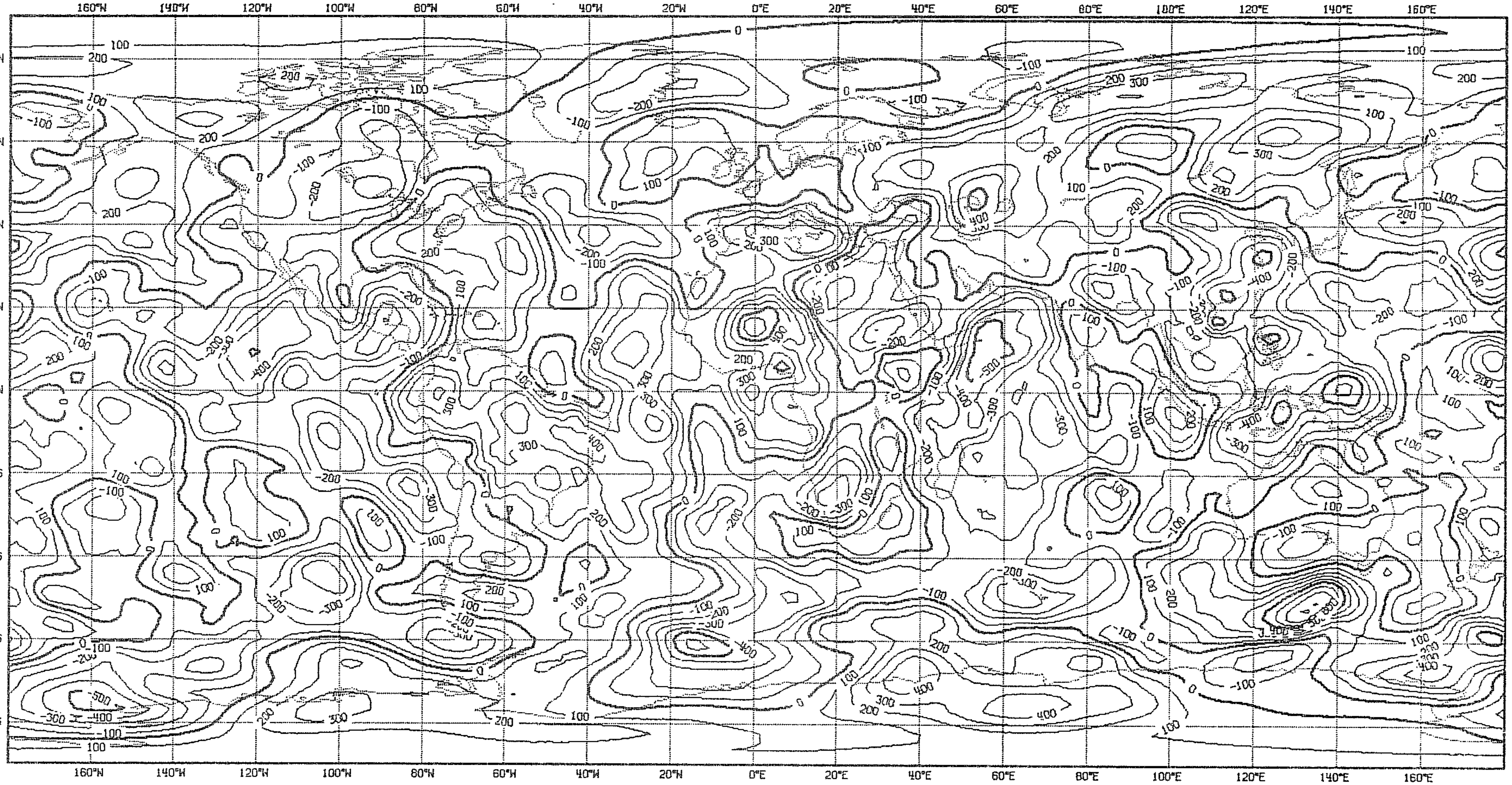

Fig. 2 Difference between 6 hour surface pressure forecasts started from uninitialised and initialised analysis for 12 GMT 6 September 1982. Contour interval is 100 Pa (=1 mb).

certain limits, the data are rejected. If the first guess is contaminated with spurious oscillations, this can easily lead to unjustified rejection or acceptance of data. Such a degraded analysis leads to a degraded forecast which will deteriorate the next analysis and so on.

To demonstrate this effect, the operational assimilation cycle for Sept.6, 1982 was repeated with the initialisation switched off. Although the absolute number of rejections did not change very much, a substantial number of data (around 30) were treated differently. Fig.3 shows the 850 mb observations superimposed on the first guess height and wind field. The two encircled wind observations near 25°E, 50°N and 60°E, 60°N were accepted in the re-run although their directions are clearly in error. The operational run with initialisation correctly rejected them. A number of other observations, which were perfectly valid, were rejected in the re-run because of spurious waves in the first guess field. The effect of assimilation without initialisation becomes more serious if one continues beyond one day.

3. WHAT ARE GRAVITY WAVES?

A simple example of gravity waves are the patterns originating from a stone thrown into a pond. The main processes leading to the propagation of these waves are sketched in Fig.4. Once the surface of an incompressible fluid has been elevated at an arbitrary point B, gravity tries to bring the particles back into their original lower level. This leads to mass convergence in regions A and C which, in turn, result in an elevation of the surface, as shown in the lower part of Fig.4. Two aspects are worth keeping in mind: the need for an initial disturbance and the importance of divergence.

For large scale gravity waves on a rotating planet, the Coriolis force cannot be neglected. Therefore, these waves are called inertia-gravity waves. Without them, the important geostrophic adjustment process cannot take place.

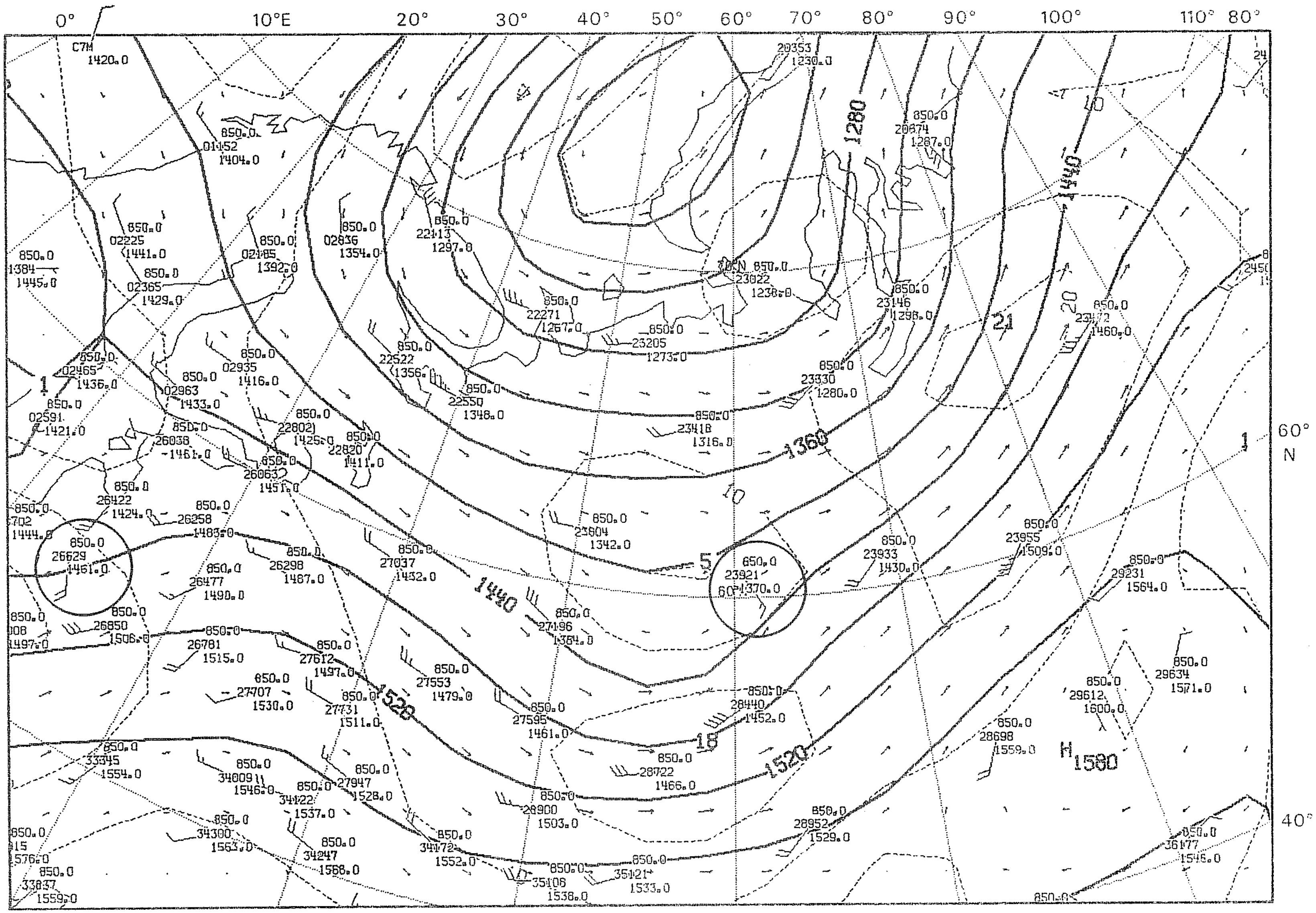

Fig. 3 850 mb observations and first guess height (full) and wind field. Contour inverval for isotachs (dashed) is 10 m/s. Circles indicate differently treated observations.

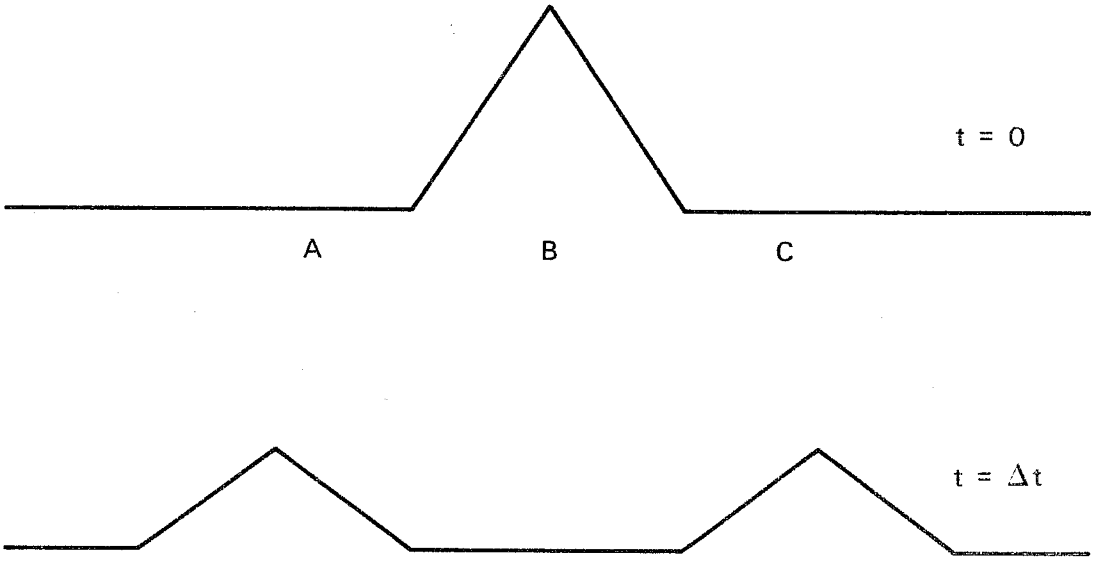

Fig. 4 Prototype gravity wave.

That is why their careful modelling has proven to be important for an atmospheric model. In a stratified fluid, gravity waves can travel not only along the surface, but also within the fluid. These internal waves are responsible for transports of momentum and energy, especially near mountains. They can sometimes be observed downstream of mountain ridges in the form of lenticular clouds which form in the wavetops, where the particles reach the condensation level.

The question now remains: why do atmospheric models show the excessive gravity wave activity as given in Figs.1 and 2. This question can best be answered by referring to Fig.4. Suppose we have a perfect model to predict the surface height of the fluid in a pond. To start the forecast, we need initial values of the surface height. Suppose, we have 3 observations in points A,B and C. If now the observation in point B has a positive error, the model will start with an initial state similar to the one sketched in Fig.4, thus immediately leading to excessive gravity wave activity. A further source for unrealistic initial amplitudes for gravity waves is the analyses system used to interpolate the irregularly spaced observations to regular grid points. Even with perfect observations an erroneous initial state might result only from errors in the analysis scheme. A third cause is the model itself, which - although presented with a perfect initial state - can generate gravity waves because it is only an approximate description of the fluid. Generally, all 3 causes contribute to unrealistic gravity wave amplitudes in the initial state.

4. WHAT IS INITIALISATION?

In quite general terms, initialisation tries to define a reasonable initial amplitude for inertia-gravity waves. (For the simple example in Fig.4 it would mean to set the amplitude at point B to zero). A number of methods have been used in the past. In the early days of numerical weather prediction, gravity waves were excluded altogether from the models by modifying the governing equations. Therefore, no initialisation was required. However, it was soon realised that these filtered models are inferior to the primitive equation models. Early initialisation methods used for these models all tried to control gravity wave activity by a suitable diagnostic definition of the initial wind field using various kinds of approximate relations. The most simple approach is to use the geostrophic relation. A higher degree of approximation can be achieved by the "balance equation" which yields a non-divergent wind field from the analysed mass field. Thus, the observed wind information is not used. Furthermore, the diagnostic relations are extremely simplified. For instance, there are considerable divergent motions in synoptic systems, which are suppressed initially when the balance equation is used. A third shortcoming is the neeed, sometimes to modify the analysed height field for mathematical reasons.

It is therefore no surprise, that alternative methods have been developed. One of them is dynamical initialisation, where the model itself is used to derive a suitable initial state. Usually, the model is integrated forward and backward in time and gravity waves are damped by heavy time and space filtering. However, the filter operators cannot damp gravity waves selectively, they also influence the meteorologically important flow. Because of the backward integration in time, irreversible physical processes can not, in general, be included. Furthermore, the method is expensive in terms of computer time. Recently, a third method, normal mode initialisation has successfully been applied. This method will be outlined in the following sections.

5. WHAT ARE NORMAL MODES?

Normal modes are the free motions of a system capable to vibrations. A typical example is a guitar string. Once it has been excited, it starts to vibrate in a way characteristic for that particular string. Therefore, these vibrations are called "own" or "eigen" vibrations, or - in mathematical terms - normal modes. Another example are the surface waves on a pond. They are the eigenvibrations of the physical system "pond". Usually, there is an infinite number of eigenvibrations, each having a particular scale. If a stone is thrown into a pond, only a subset of these modes is excited, depending on the initial shape of the disturbance caused by the stone.

The eigenvibrations of atmospheric models fall into two classes: the Rossby modes and the gravity modes. Every mode shows a specific 3-dimensional structure of mass- and windfield. The vertical structures are identical for Rossby and for gravity waves. They differ, however, considerably in their horizontal structure and in the relation between mass- and windfield. Fig.5 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 windfield. 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 mode l has l-1 sign changes. With increasing order l, the region of maximum amplitudes moves towards lower levels.

The horizontal structure of an external Rossby mode is shown in Fig. 6. 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 mass- and windfield. In fact, they are the only means to establish globally valid coupling. The Rossby mode shown in Fig. 6 can be characterised by 3 indices: - the vertical mode number - the zonal wavenumber - the meridional index. As Fig. 6 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. 5. The meridional index in Fig. 6 is 1, which means the largest meridionally symmetric scale. Obviously, the zonal wavenumber is 1.

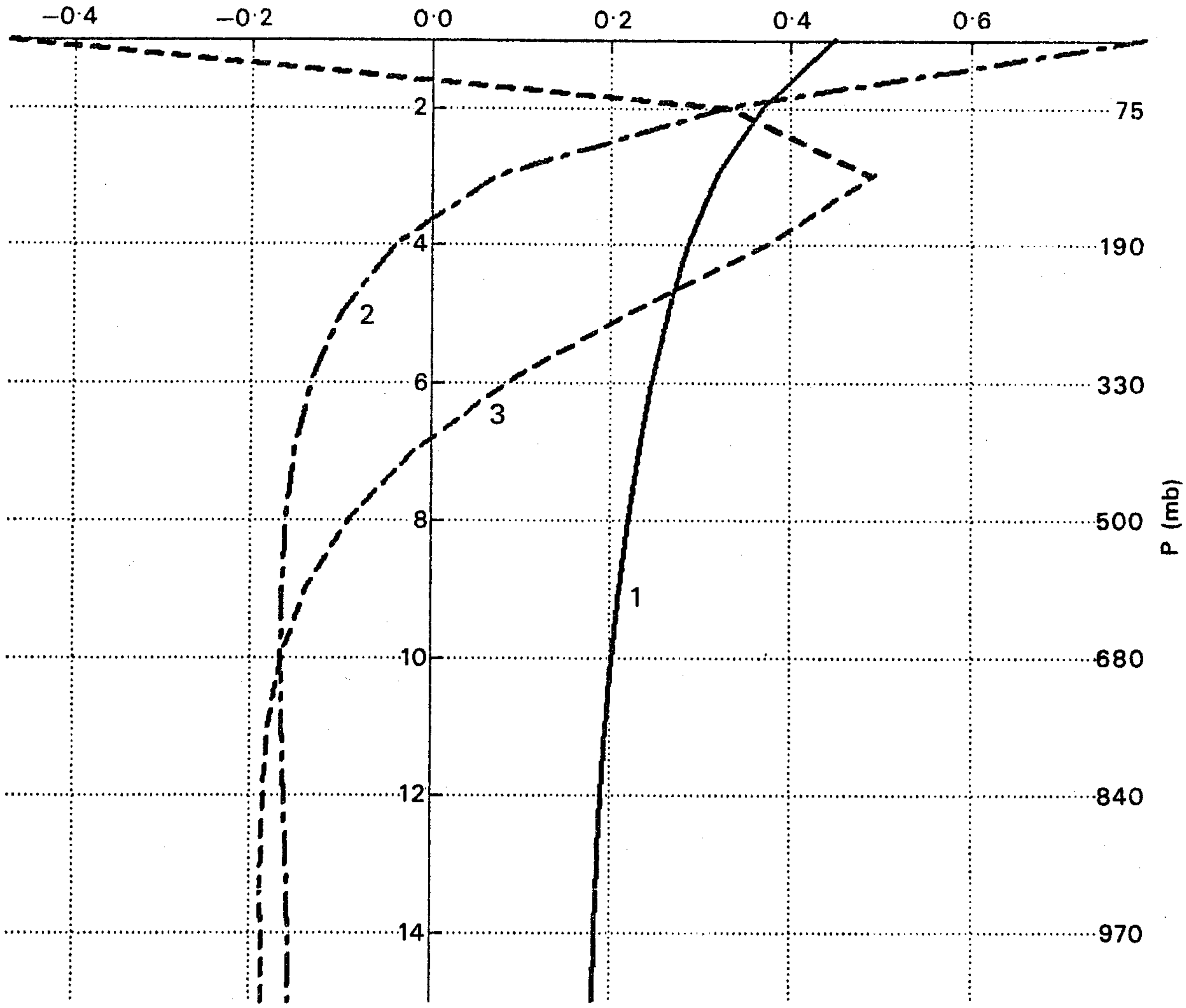

Fig. 5 First 3 vertical modes for isothermal (300 K) basic state. Rounded pressure values in mb are valid for p_s = 1000 mb.

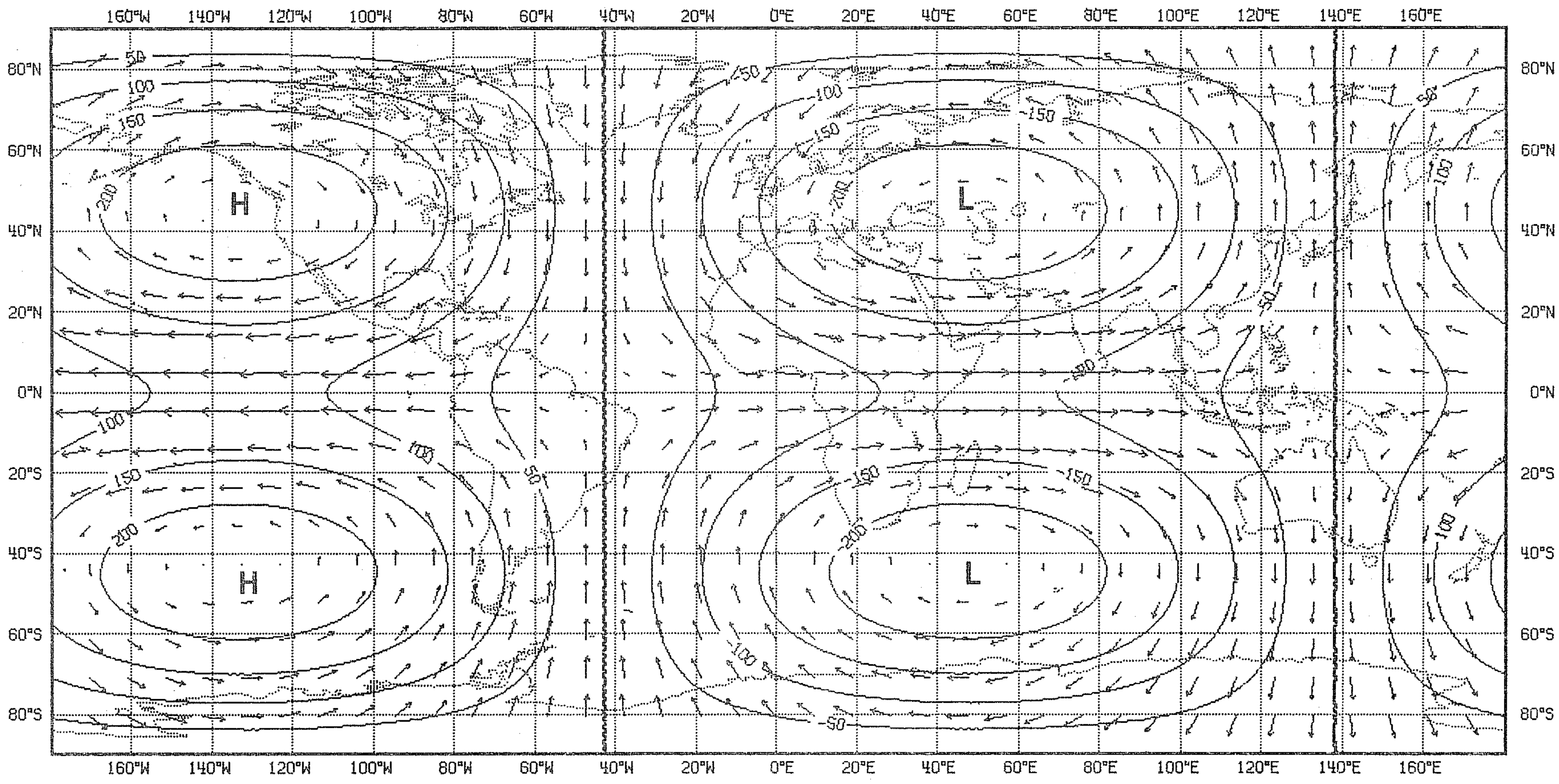

Fig. 6 Gravest symmetric, zonal wavenumber 1, external Rossby mode.

When looking at a gravity mode (Fig. 7) 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 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.7 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. 6 is 4.7 days, whereas the westward gravity wave in Fig.7 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 decrease 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. 6 and 7. The only way to make a proper distinction is to look at the 3-dimensional structure of both the mass- and the windfield.

6. WHAT IS NORMAL MODE INITIALISATION?

In normal mode initialisation, the unique features of normal modes are efficiently used to remove unwanted gravity wave amplitudes from the analysis.

First of all, the normal modes have to be computed. This is done by linearising the model equations around a simple basic state and solving the resulting set of equations. The details can be found in Temperton and Williamson (1981). The result of these computations are the Rossby and gravity modes discussed in the previous section. This computation needs to be done only once; the normal modes are then stored for subsequent use.

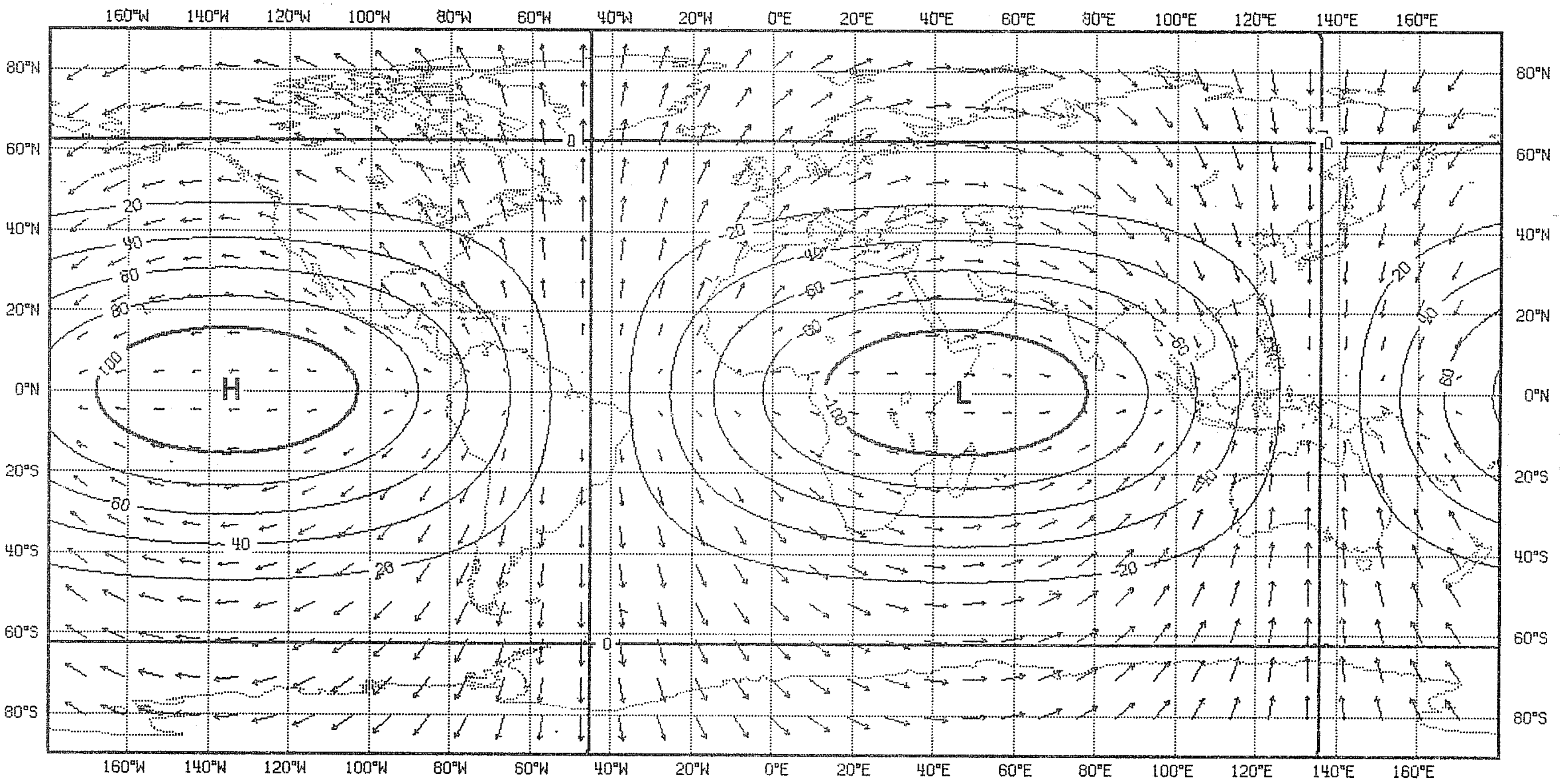

Fig. 7 Gravest symmetric, zonal wavenumber 1, westwards travelling, external gravity mode.

The first step in the initialisation procedure itself is the identification of those structures in the analysed fields which lead to the excitation of gravity waves. Speaking mathematically, the analysis is projected on the gravity modes. This guarantess, that the most important part described by the Rossby modes, is not touched at all.

Fig. 8 shows the gravity mode projection of the analysis increment (analysis - first guess) for 6.9.82, 12Z. As can be expected, the gravity mode projection of the wind field is mostly divergent. A closer inspection of the analysis off the Chilean coast (Fig. 9) reveals, that the analysis (first guess + increment) in this area is not in geostrophic balance. The isotachs (dashed) indicate a rapid decrease in wind speed towards the coast, which is not supported by a corresponding slackening of the height gradient (full lines). Similar results apply to all extra-tropical areas, where large gravity wave projections are invariably related to ageostrophic analysis increments.

In the second step, the identified gravity mode projections are modified. The simplest method would be to set them to zero (linear normal mode initialisation). However, this would not solve the problem, because the nonlinear advection processes and the parameterised physical processes would immediately start to excite new gravity waves. Machenhauer (1976) proposed to eliminate only those parts of the gravity wave components, which are not in balance with the nonlinear processes. As a result, the linear tendencies for the gravity waves will be exactly compensated by the nonlinear dynamical and physical tendencies, yielding vanishing total tendencies for the initialised gravity waves. The corresponding mathematical equations can easily be formulated in normal mode space. It is a system of nonlinear algebraic equations, which can be solved iteratively.

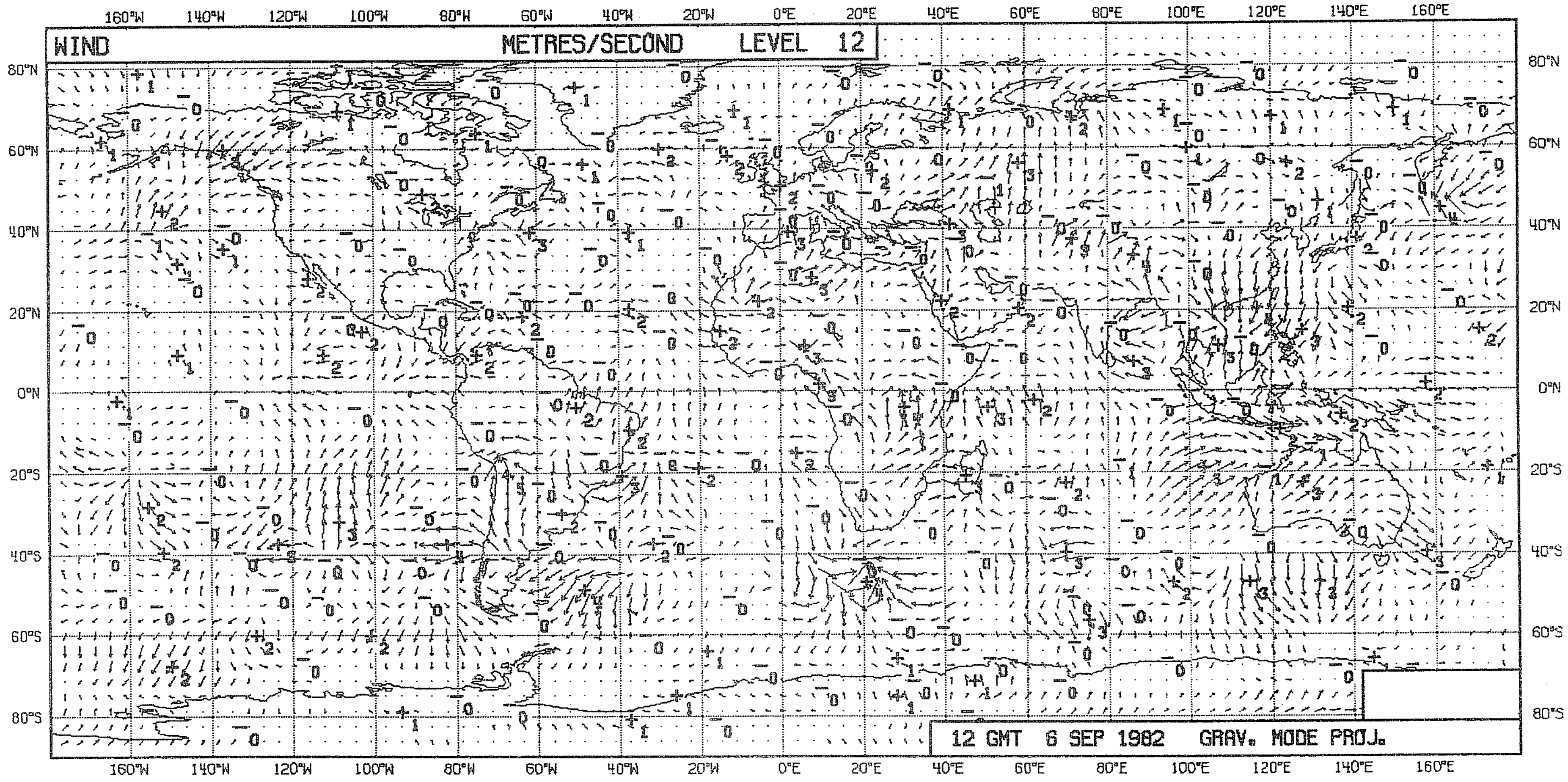

Fig. 8 Gravity mode projection of analysis wind-increments at model level 12 (around 850 mb) for 12 GMT 6 September 1982. Numbers indicate wind speeds in meters/second.

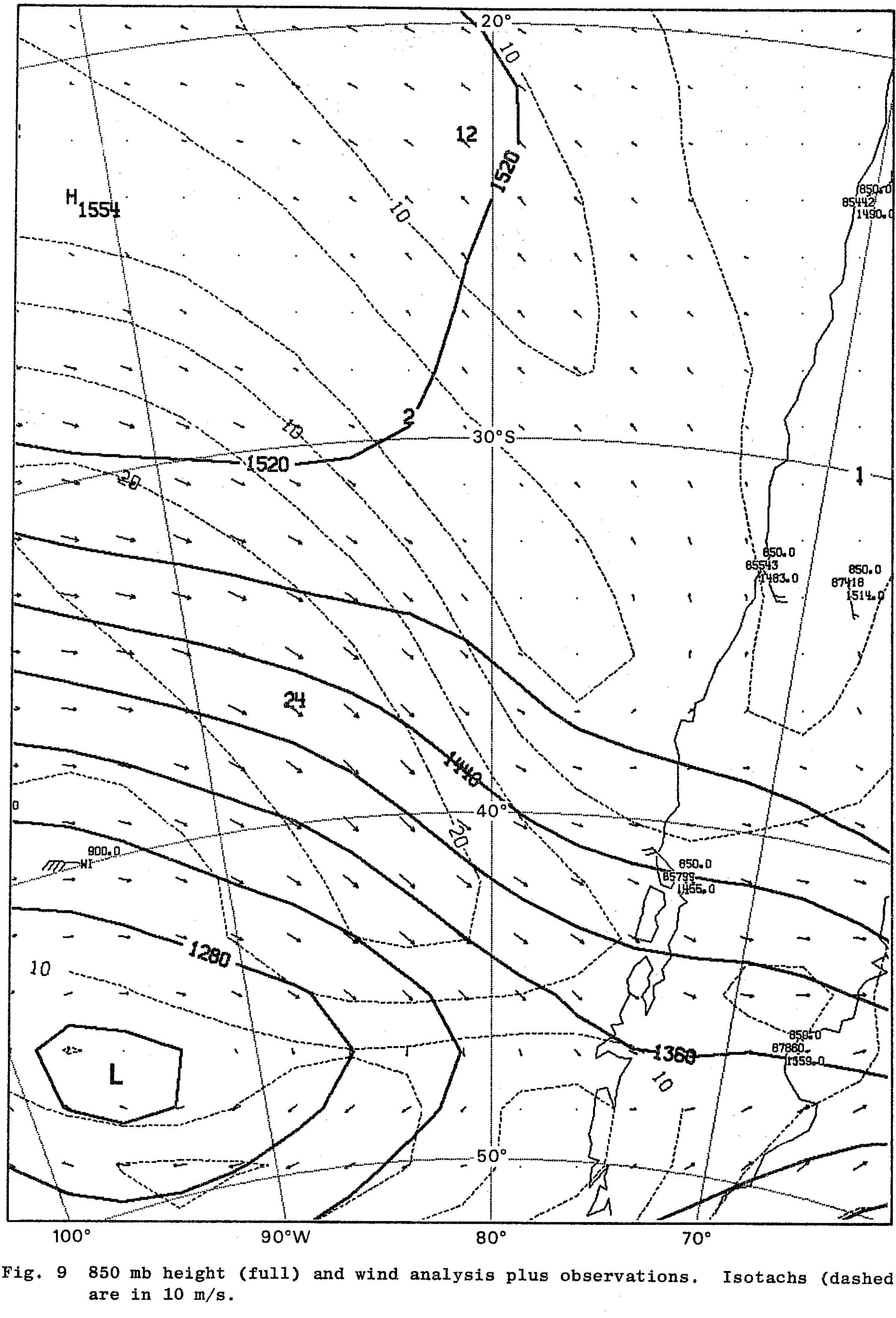

Fig. 9 850 mb height (full) and wind analysis plus observations. Isotachs (dashed) are in 10 m/s.

7. THE PERFORMANCE OF THE ECMWF SCHEME

a) Implementation

The normal modes used in the operational scheme are derived assuming an isothermal atmosphere (300K) at rest. Only the first five vertical modes are initialised. Two iterations are used to solve the nonlinear algebraic system. In an attempt to reduce surface pressure changes by initialisation, the analysed surface pressure is restored with a latitude-dependent weighting after the first iteration. An estimate of the diabatic forcing is included, but kept independent of the iteration because of convergence problems. This estimate is computed by time-averaging the diabatic tendencies during a 2-hour forecast started from the uninitialised analysis. Only those tendencies, which force inertia-gravity waves with periods longer than a certain cut-off period (11 hours) are included. In the calculation of the adiabatic tendencies, virtual temperature is used.

b) Typical initialisation changes

As neither the analysis scheme nor the data know the subtle nonlinear balance required to supress unwanted gravity wave activity, initialisation usually changes the analysed fields. Surface pressure and divergent wind are most affected, temperatue and rotational wind changes are neglegible. Fig. 10 shows the initialisation changes to surface pressure, averaged over all 12Z cases for April 1981. Mean Changes are in the order of 1 mb. The initialisation changes to the wind field in 200 mb (Fig.11) are in the order of several meters/second. They are mostly divergent. With the adiabatic scheme in use in April 81, wind changes were largest in the tropics. As will be shown in the next section, diabatic initialisation results in smaller changes. In individual cases, initialisation changes can amount to 5 mb in surface pressure and up to 10 m/sec in upper tropospheric winds. Invariably, large initialisation changes point to gross imbalances in the analysis. In the extra-tropics, this is possible, because the mass wind relation for the

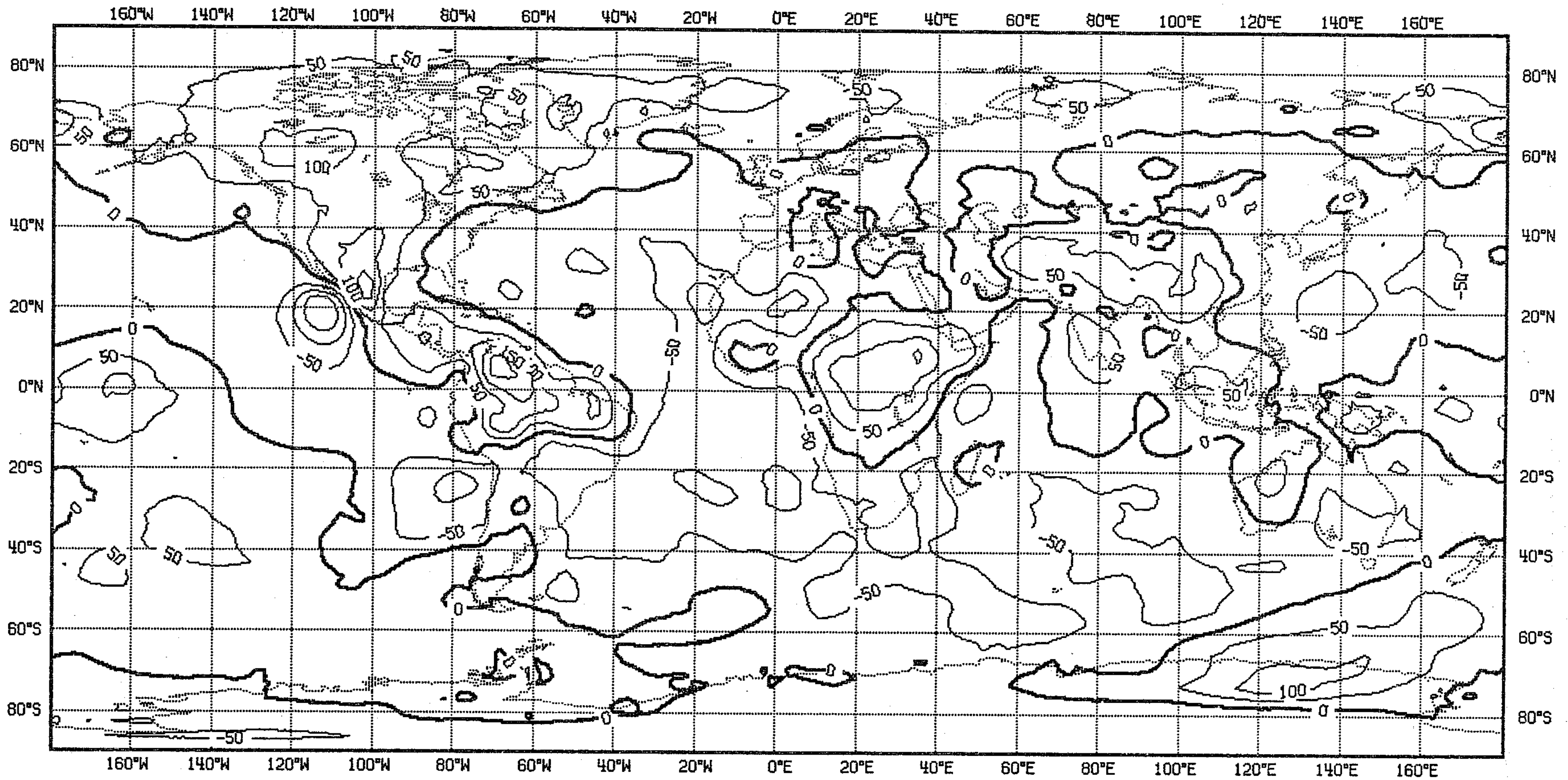

Fig. 10 Mean surface pressure initialisation change (initialised - unitialised) in 50 Pa (= 0.5 mb) for April 1981. (Averaged over 30 12 GMT cases).

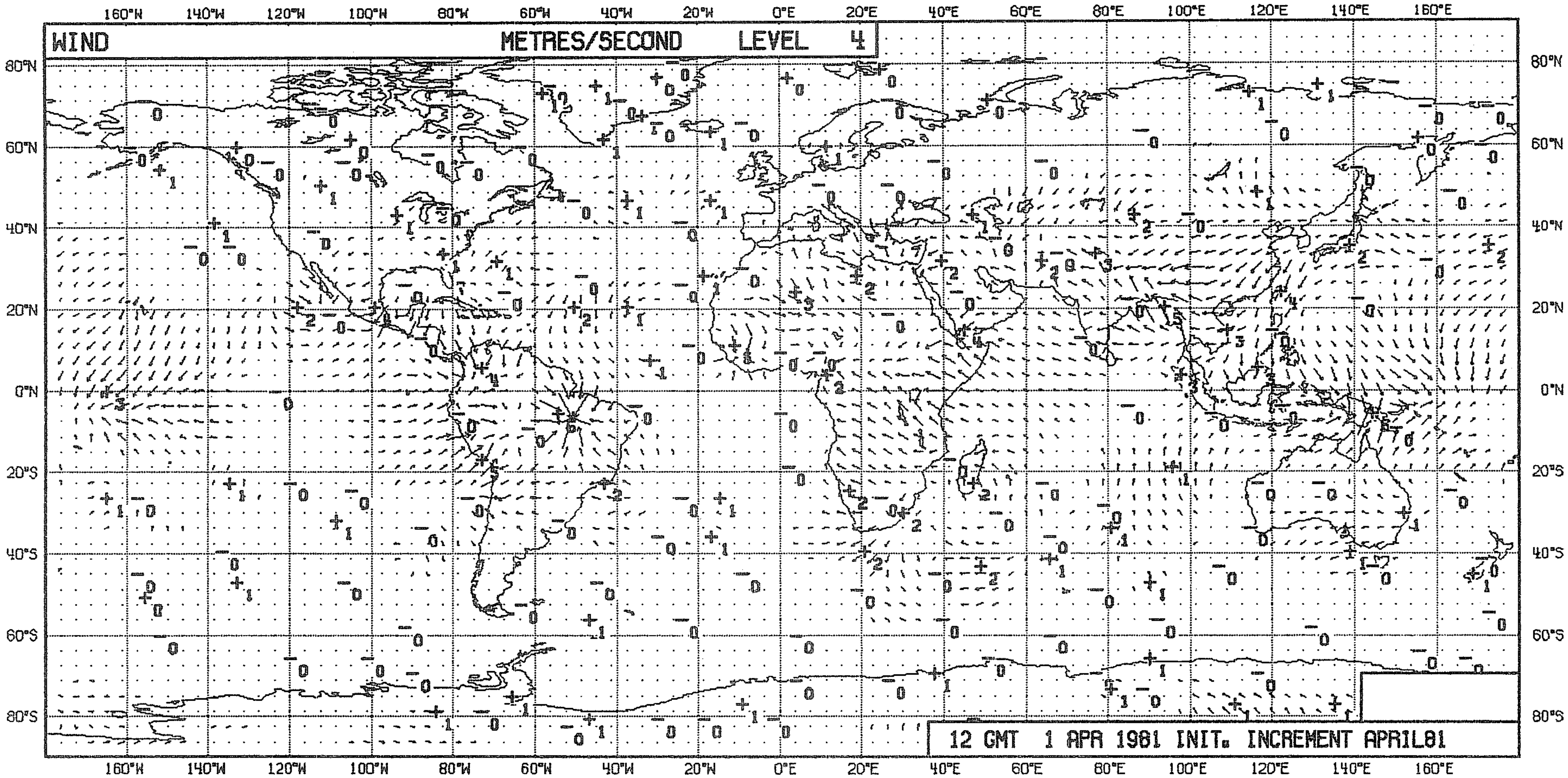

Fig. 11 Mean wind changes by initialisation at level 4 (around 200 mb) for April 1981. Numbers indicate wind speeds in m/s.

analysis increments is not exactly geostrophic (Lönnberg,1982) and because of different data selection for neighbouring analysis boxes. In the tropics, the present analysis scheme tends to alias a pure Rossby mode on a number of different Rossby- and gravity modes (Cats and Wergen, 1982).

c) The impact of diabatic processes

Adiabatic initialisation tends to supress diabatically driven circulations in the tropics (Bengtsson,1980). In order to test the impact of diabatic initialisation, 6 days of assimilation were repeated for the FGGE period 5. - 11.6.1979. Fig. 12 shows the zonally averaged, vertically integrated, initialised mass flux for 12.6.79, 12Z. This variable can be interpreted as a stremfunction for the flow in the ϕ-p plane. Adiabatic initialisation (top) supressed the upper tropospheric return flow in the Hadley cell as well as the mid-troposheric vertical motions. With diabatic tendencies included (bottom), the tropical circulation is more intense. A further indication of the positive impact is the time-averaged (6.6. - 11.6) initialised velocity potential at 200 mb (Fig. 13). With diabatic initialisation (bottom), divergent motions in the tropics are stronger than for the adiabatic scheme (top). The effect on the spin-up of the model is demonstrated in Figs. 14 and 15, which show the diabatic heating (averaged over 2 weeks) for the slab 700-300 mb in the northern hemisphere for the initialised analysis (left) and the 1 day forecast (right). For the method of calculation see White, 1983. Fig.14 gives the results for the first two weeks of September, when adiabatic initialisation was still in use. Initially, there is little diabatic heating in the tropics, whereas the 24 hour forecast tends to overshoot when compared with the level achieved in later stages of the forecast. After the introduction of diabatic initialisation into operations, the initial diabatic heating rates (Fig. 15, left) are closer to the 24 hour forecast values (Fig. 15, right). Even the diabatic heating fields for the storm tracks in the North-Atlantic are similar. The excessive convective activity over East-Africa is currently being investigated.

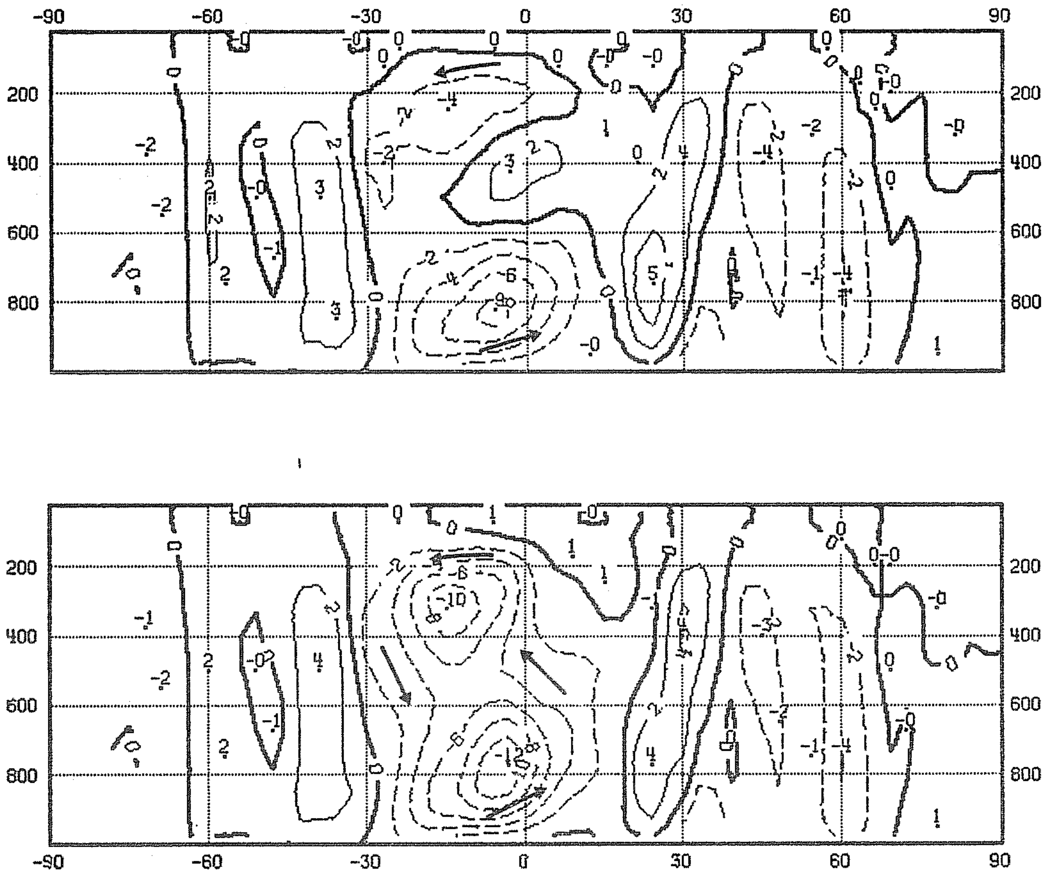

Fig. 12 Zonally averaged, vertically intergated mass flux in 10^{10} kg/s for 12 GMT 11 June 1979 for adiabatic (top) and diabatic (bottom) initisation.

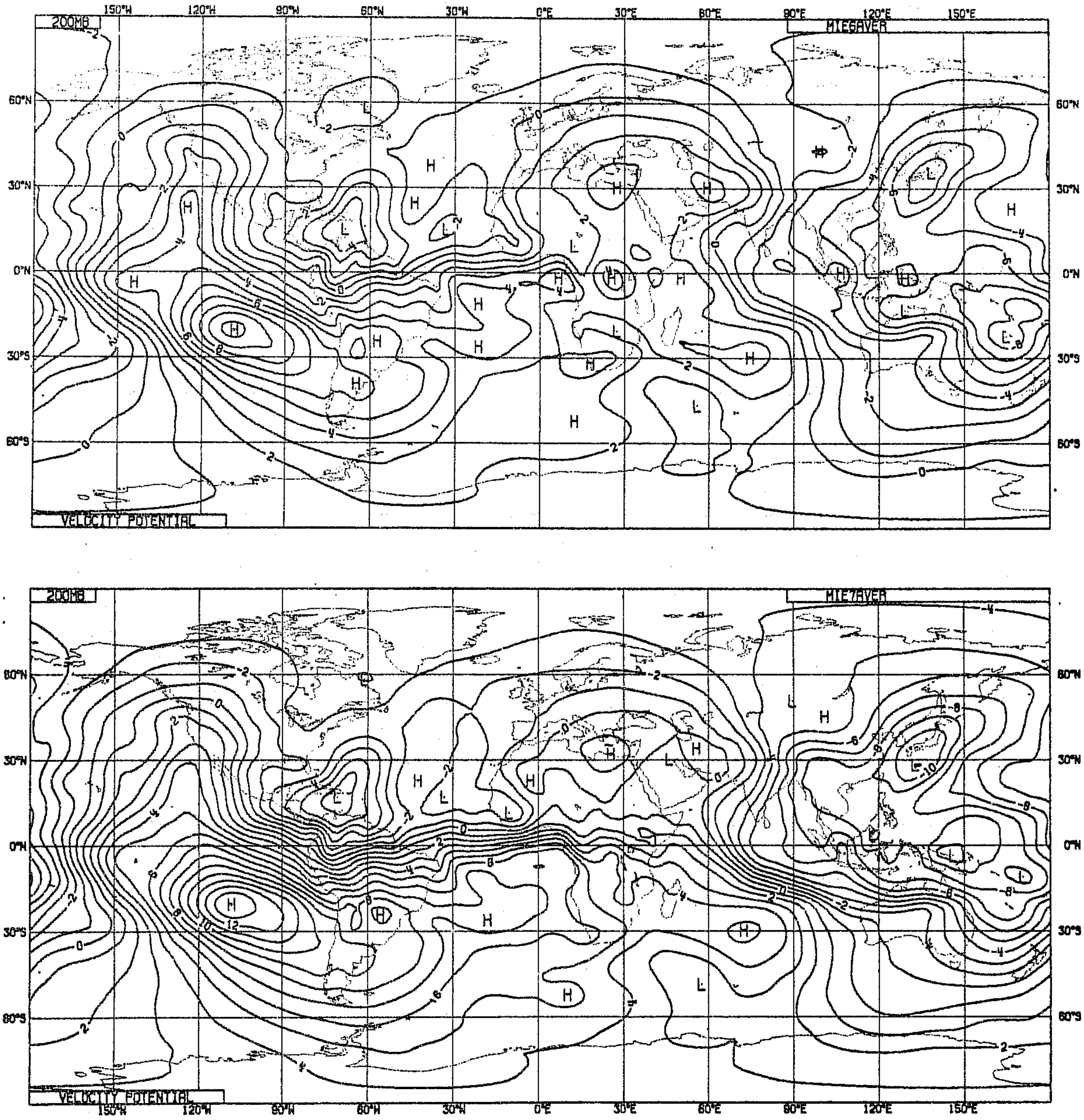

Fig. 13 Time-averaged (6-11 June 1979) initialised velocity potential at 200 mb in 10^6 m^2/s for adiabatic (top) and diabatic (bottom) initialisation.

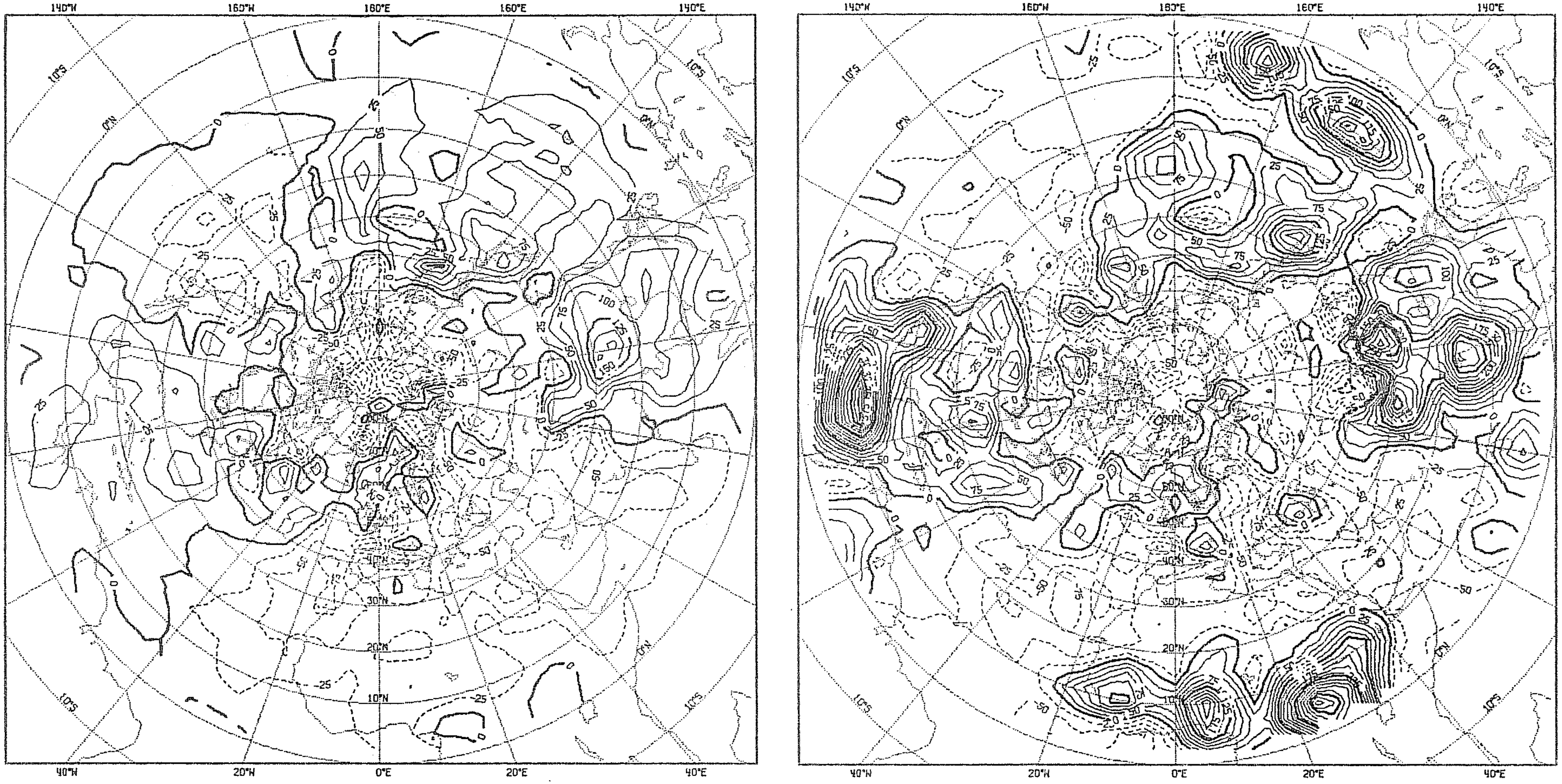

Fig. 14 Time averaged (1-15 September 1982) diabatic heating in W/m^2 for the slab 700 to 300 mb before the introduction of diabatic initialisation. Left for initialised analysis, right for 24 hour forecast.

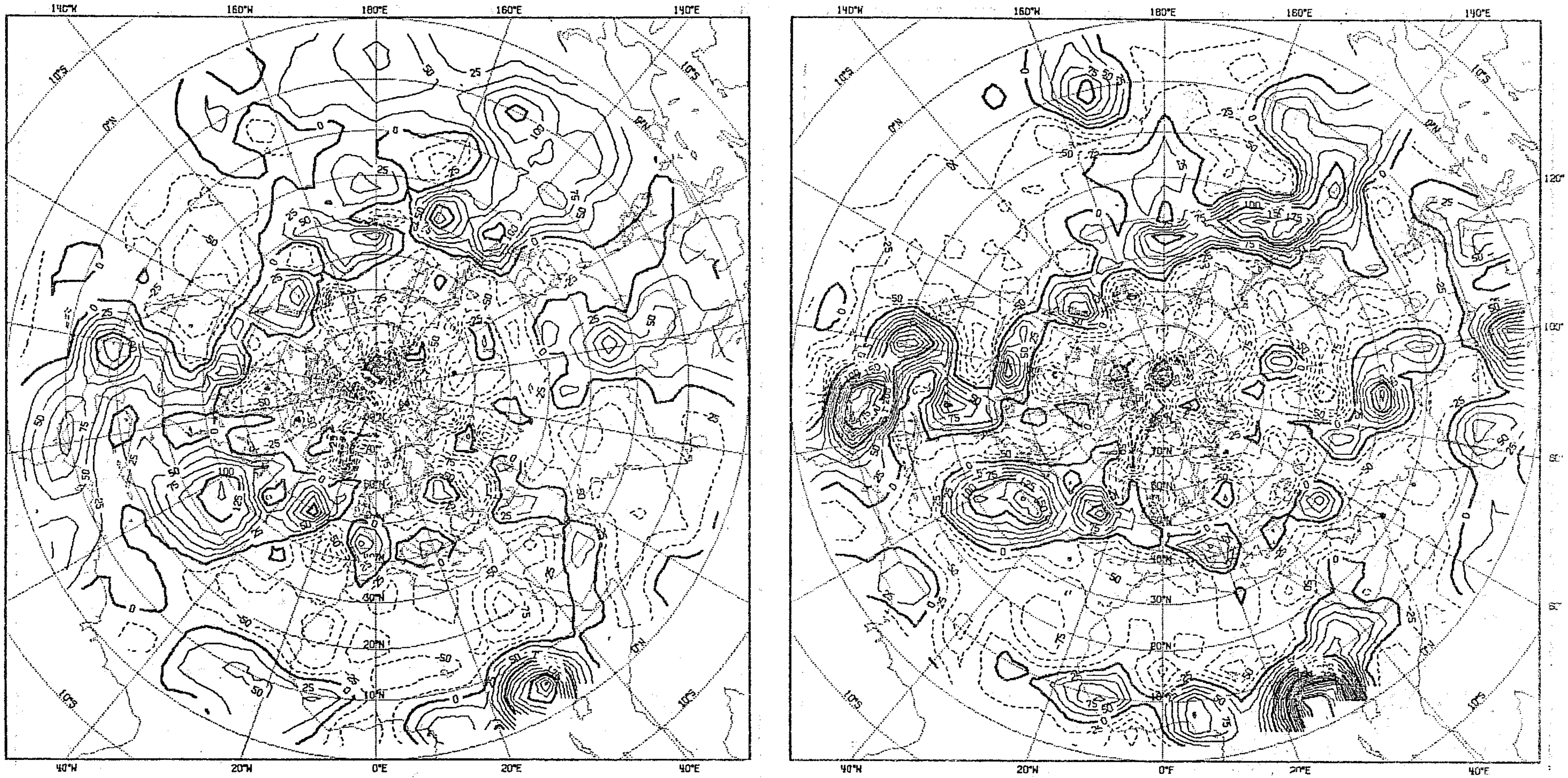

Fig. 15 Time averaged (1-15 October 1982) diabatic heating in W/m^2 for the slab 700 to 300 mb after the introduction of diabatic initialisation. Left for initialised analysis, right for 24 hour forecast.

Tropical scores for 200 mb winds (Fig. 16) show an improvement for the October 82 ensemble forecasts compared to the October 81 cases. The improvement is mainly due to improved skill in the large- and medium scale waves. As the forecasts are verified against the initialised analysis, part of the improvement is due to a more realistic verifying analysis.

8. SOME REMARKS FOR USERS

When using uninitialised or initialised analyses, some points should be kept in mind. The uninitialised analysis usually is closer to the data. It should therefore be used where a close fit to observation matters. However, it should not be used for calculations of divergence or derived quantities, such as vertical velocity or surface pressure tendency. The initialised analysis is much better suited for these purposes, although its balance - especially in the tropics - is model dependent. Also, a forecast should in general only be started from an initialised analysis. A close fit to observations does not necessarily mean that the observations are actually accepted by the model. Finally, when judging initialisation changes, the three-dimensional structure of both mass- and windfield should be taken into account. Invariably, big initialisation changes indicate problem areas in the analysis.

Acknowledgement

I should like to thank Glenn White, who allowed me to publish some of his figures. R.Daley and A.Hollingsworth carefully reviewed the manuscript.

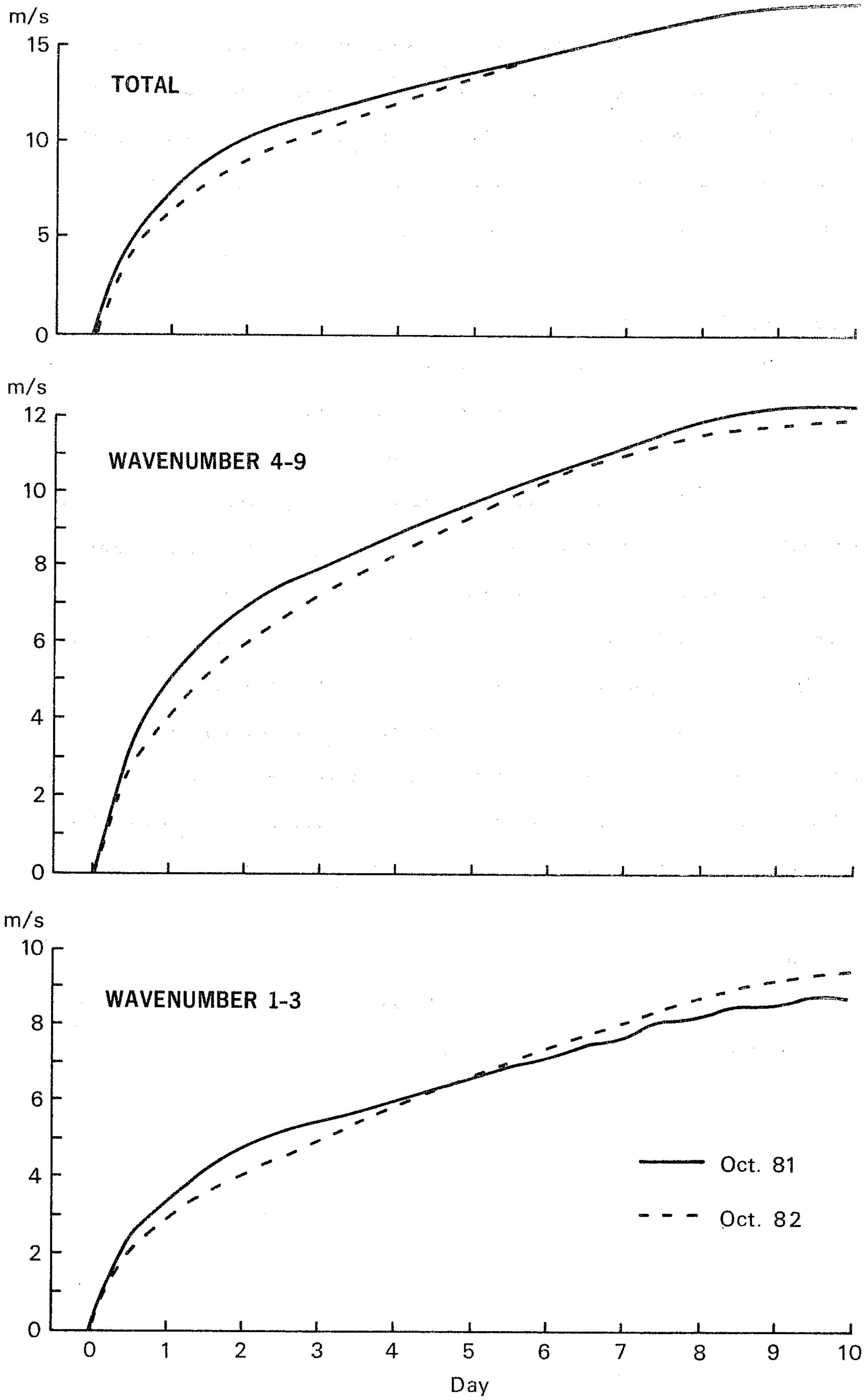

Fig. 16 RMS error of vector wind in the tropics (35^oS to 32.5^oN) at 200 mb averaged over all October 1981 forecasts (full) and October 1982 (dashed) forecasts.

REFERENCES

Bengtsson L., 1980: Current problems in four-dimensional data assimilation. ECMWF seminar proceedings on 'Data assimilation methods', 195-218. (Available from ECMWF).

Cats, G. and Wergen,W., 1982: Analysis of large scale normal modes by the ECMWF analysis scheme. To be published in ECMWF Workshop Proceedings on 'Current problems in data assimilation'.

Daley,R., 1980: Normal mode initialization. ECMWF Seminar Proceedings on 'Data assimilation methods', 107-158. (Available from ECMWF).

Lönnberg,P., 1982: The ECMWF analysis system. ECMWF Seminar Proceedings on 'Interpretation of numerical weather prediction products'.

Machenhauer,B., 1977: On the dynamics of gravity oscillations in a shallow water model, with application to normal mode initialization. Beitr.Phys.Atmos. 50, 259-271.

Temperton,C. and Williamson,D.L., 1981: Normal mode initialization for a multi-level grid-point model, Part 1: Linear aspects. Mon.Wea.Rev., 109,729-743.

White,G.H., 1983: The global circulation of the atmosphere December 1981 - November 1982 based upon ECMWF analyses. Technical Note, Dept.of Meteorology, Reading University, UK, in preparation.