

ON INITIALIZATION OF PRIMITIVE EQUATION MODELS

by

WILLIAM K-F. GRANT, JR.

B.S., Hampden-Sydney College
(1967)

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE
DEGREE OF MASTER OF SCIENCE

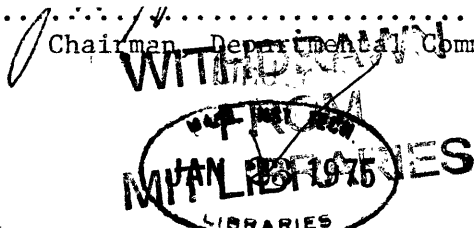
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

January, 1975

Signature of Author.....
Department of Meteorology, January, 1975

Certified by.....
Thesis Supervisor

Accepted by.....
Chairman, Departmental Committee on Graduate Students



ON INITIALIZATION OF PRIMITIVE EQUATION MODELS

William K-F. Grant, Jr.

Submitted to the Department of Meteorology on January 22, 1971 in partial fulfillment of the requirements for the degree of Master of Science.

ABSTRACT

Initialization procedures for primitive equation models based on the balance equation and the dissipative-iterative schemes developed by Nitta and Hovermale are examined. A simpler scheme, attributed to Okamura, and modified by Rivas, is also studied. A non-linear, shallow-water-equation model is used to numerically generate artificial data which simulates a balanced, initial state. This balanced state is then perturbed to simulate the effect of observational errors. Numerical experiments are performed to compare the recovery of balance by the various initialization procedures. Four iterative schemes and the consistent solution of the balance equation are compared. The common practice of restoring the mass field during the iterative process is shown to significantly decrease the rate of convergence of these schemes. However, the methods converge quite rapidly if the mass field is allowed to adjust freely. The high degree of balance recovered as a result of this adjustment far compensates for the possible increase in error of the adjusted fields. This error is considerably reduced by using an approximate form of the gradient wind equation to simply correct the geostrophically derived winds in data-sparse areas. The Okamura-Rivas scheme is shown to be significantly more efficient in decreasing the computation required for convergence and in achieving a more stable, balanced state. In the numerical experiments, the Okamura-Rivas scheme required an order of magnitude less computation than the balance-equation approach.

Thesis Supervisor: Eugenia E. Kálnay de Rivas

Title: Research Associate, Department of Meteorology

ACKNOWLEDGEMENTS

. 3 .

I am exceedingly grateful to my advisor, Dr. Eugenia E. Kálnay de Rivas for her patient and skillful guidance in this work. Professor Norman A. Phillips also contributed to the early planning of the experimental technique. Thanks are also extended to Professors Jule G. Charney and Edward N. Lorenz with whom I had some enlightening discussions on the physical significance of this problem. Mrs. Susan Wiseman cheerfully and carefully typed the thesis. My most personal debt is to my wife, Mary Rose, and my children, Sam and Tess, who all bore the burden of my erratic habits with patience and love during some trying periods.

The computations were performed, by remote terminal, at the Goddard Institute for Space Studies. My thanks are given to Doctors Robert Jastrow and Milt Halem for making these facilities available. My attendance at the Massachusetts Institute of Technology is sponsored by the United States Air Force under the auspices of the Air Force Institute of Technology.

TABLE OF CONTENTS

	<u>Page</u>
SECTION I INTRODUCTION	5
SECTION 2 THE BALANCE EQUATION APPROACH	9
SECTION 3 SEMI-PROGNOSTIC ITERATIVE APPROACH	13
SECTION 4 DESIGN OF NUMERICAL EXPERIMENTS	27
SECTION 5 RESULTS	36
5.1 Introduction	36
5.2 Initialization of the geostrophically perturbed state	40
5.3 Initialization of a randomly perturbed state	54
5.4 Gradient wind correction to the geostrophic velocity fields	59
SECTION 6 SUMMARY AND CONCLUSIONS	62

1. Introduction

It has long been known that the use of observed data directly in initial fields for numerical simulation, by primitive equation models¹, of atmospheric motion predicts gravity-inertia oscillations in excess of those observed in the atmosphere (Richardson, 1922). Such initial fields contain excessive imbalances between Coriolis and pressure forces, which arise primarily from measurement errors. Actual imbalance can and does occur in the atmosphere, but it is doubtful that the present sampling network can yield an accurate measure of the excess.

The fact that observed large-scale motions do seem to pass through a succession of quasi-balanced states suggests that these motions are subject to a process by which this equilibrium is maintained. It is generally accepted that the mutual adjustment between mass and velocity fields, under the influence of the earth's rotation and gravity, is the primary process which maintains this quasi-geostrophic balance. The adjustment process has been extensively studied (Rossby, 1937-1938a, 1937-1938b; Cahn, 1945; Bolin, 1953; Phillips, 1963). The dispersive character of gravity-inertial waves is fundamental to the maintenance of a balanced state by the adjustment process. A primitive equation model contains implicitly the adjustment mechanism as well as the capability to propagate gravity-inertial waves. However, the effect of erroneous imbalances in initial data is to excite gravity-inertial waves which strongly distort the initial tendencies of the dependent variables in a numerical forecast. It is this latter property which predominated in

¹The term "primitive equation model" is used here generically to mean any thermo-hydrodynamic model in which the hydrostatic balance relation replaces the prognostic equation of vertical motion.

Richardson's early experiment and discouraged further attempts to formulate primitive models for some time.

The theory of diagnostic balancing of initial data centers on the fact that gravity-inertia waves are characterized by divergence in the flow field. Thus they may be eliminated by purging the initial fields of their divergence, thereby reducing the chance of numerical propagation. Geostrophic balance satisfies the requirement for non-divergence but distorts the representation of curved flow. Charney (1955) first formulated "the" balance equation by explicitly removing divergence and its time derivations from a prognostic divergence equation. Phillips (1960) recognized that a small amount of divergence is needed in initial flow to properly represent the vertical motion. He suggested that a consistent estimate of the divergent wind field could be determined from the quasi-geostrophic omega equation and added to the non-divergent wind field determined by the balance equation. This suggestion has been extended, by several authors, to develop quite complicated initialization procedures. (See Hinklemann, 1961; Miyakoda, 1963; and Krishnamurti and Baumhefner, 1966 for example.)

The balance equation, by itself, is a mixed, hyperbolic-elliptic equation. It is easily solved only in the special case that it is purely elliptic. This restriction creates an ellipticity constraint on the mass (or pressure) field which may require its modification before the balanced winds may be determined from it. Flows which violate the ellipticity constraint are perfectly legitimate in the atmosphere. Miyakoda (1960) states that the region around an anticyclone, especially in front of a typhoon, may be hyperbolic. The treatment of this region

determines the prediction of the typhoon path. Artificial modification of this anticyclone, in order to make the region elliptic, may seriously affect the numerical forecast. The ellipticity constraint does not arise when the balance equation is part of a complete set of balanced equations in a filtered model (Charney, 1973). But for primitive equation models, it is solved independently to achieve balance in the initial data. The artificial modification of the data, which may be imposed by the ellipticity constraint, is a disadvantage of this approach. Nitta and Hovermale (1969) state that, in addition, the initialization procedures with filtered equations cannot produce a balanced flow which is in perfect agreement on the state of mutual adjustment for the primitive forecast equations. Nevertheless the balance equation approach has traditionally been adopted for initialization in such forecasts.

It is known that certain finite-difference, marching procedures selectively reduce the amplitude of high frequency oscillations. Miyakoda and Moyer (1968) first suggested an alternative approach to initialization using this property. The technique uses the primitive equations directly to march forward one step and then return to the initial time by reversing the time step. If this procedure is repeated in an iterative manner the high frequency gravity-inertial waves are eventually damped out. Miyakoda and Moyer originally suggested that the divergence should be explicitly set to zero after each time step to insure a balanced state. Nitta and Hovermale (1969) suggested that this restriction was not necessary. They proposed that the mutual adjustment properties of a particular model, which are uniquely implied in its mathematical representation, are sufficient to attain balance. Since the primitive equations are used directly

in the iterative process, the state of balance should be completely consistent with the model. A variation of this technique has been attributed to Okamura². A simpler dissipative scheme is used and fewer evaluations of the time variation of the variables are made.

In this paper, we will introduce a more flexible version of Okamura's scheme, suggested by Rivas³, which enhances its effectiveness without altering its simplicity. We will describe numerical procedures used in the balance equation approach and the iterative approach to initialization. The numerical properties of three iterative methods will be examined in some detail. The results of numerical experiments designed to compare the balance equation approach and the different iterative techniques will also be presented.

²See the appendix of Nitta (1969).

³Personal communication.

2. The Balance Equation Approach

A general divergence equation in pressure coordinates can be written as follows (Haltiner, 1971):

$$\underbrace{\left[\omega \frac{\partial \xi}{\partial p} + \delta^2 \right]}_D + \underbrace{\left[\frac{\partial \xi}{\partial t} + \vec{V} \cdot \nabla \delta + \nabla \omega \cdot \frac{\partial \vec{V}}{\partial p} \right]}_C + \underbrace{\left[\beta u - 2J(u, v) \right]}_B + \underbrace{\left[\nabla^2 \phi - f \xi \right]}_A = 0,$$

where δ is the two dimensional divergence on a constant pressure surface and all other symbols are in common usage. The bracketed terms are labeled A, B, C, D in decreasing order of magnitude for large scale motions. If term A alone is retained the equation becomes a statement of geostrophic balance for constant f . If term A is scaled as unity, term B is of order Rossby number, ($R_o \sim \frac{1}{10}$ for large scale motions), and terms C and D are respectively one and two orders of magnitude smaller. The balance equation, as proposed by Charney (1955), is obtained by neglecting terms C and D and assuming that the velocity is non-divergent, $\vec{V} = \hat{k} \times \nabla \psi$.

This yields:

$$\nabla^2 \phi - f \nabla^2 \psi - \nabla \psi \cdot \nabla f + 2 \left(\frac{\partial^2 \psi}{\partial x \partial y} \right)^2 - 2 \left(\frac{\partial^2 \psi}{\partial x^2} \frac{\partial^2 \psi}{\partial y^2} \right) = 0 \quad 2.1$$

where ∇ is a horizontal gradient operator on a constant pressure surface. Numerical and analytical techniques for solving this type of equation are most thoroughly developed for the case when it is purely elliptic. A general, second order, partial differential equation for the unknown ξ :

$$F(r, s, t, p, \xi, x, y; \xi) = 0, \quad 2.2$$

is elliptic if:

$$4F_r \cdot F_t - F_s^2 > 0, \quad 2.3$$

where the subscripts indicate partial differentiation and $r = \xi_{xx}$, $s = \xi_{xy}$, $t = \xi_{yy}$, $p = \xi_x$, $q = \xi_y$. If the stream function, ψ , is considered known from the observed velocity field, the balance equation is solved to obtain the geopotential field, ϕ , and is elliptic. In this case equation 2.1 is written in the notation of 2.2 as follows:

$$F(r, s, t, p, q, x, y; \phi) = r + t + g(\psi) = 0, \quad 2.4$$

with the ellipticity condition always satisfied:

$$4F_r F_t - F_s^2 = 4 > 0$$

On the other hand, if ϕ is taken directly from the data and equation 2.1 is solved for ψ equation 2.2 is written:

$$F(r, s, t, p, q, x, y; \psi) = f(r+t) + 2(rt-s^2) + f_x p + f_y q - \nabla^2 \phi = 0. \quad 2.5$$

The ellipticity condition is not trivially satisfied:

$$4F_r F_t - F_s^2 = 4(f+2r)(f+2t) - 16s^2 > 0, \text{ or}$$

$$(f + 2\psi_{xx})(f + 2\psi_{yy}) - 4\psi_{xy}^2 > 0. \quad 2.6$$

Substituting from equation 2.1 this becomes:

$$\nabla^2 \phi + \frac{f^2}{2} - \nabla f \cdot \nabla \psi > 0. \quad 2.7$$

If f is constant the ellipticity condition is:

$$\nabla^2 \phi + f^2/2 > 0. \quad 2.8$$

Houghton and Washington (1969) have demonstrated that the solution for ϕ yields more accurate results in the tropics for large scale motion while solution for ψ is preferred in extratropical regions. Thus the constraint of the ellipticity condition may apply to isolated regions of the geopotential field over a substantial portion of the globe. It should be emphasized that this restriction is primarily non-physical in nature in that it arises from inadequate numerical techniques for solving the general form of the balance equation.

In extratropical latitudes, the balance equation is usually solved by rewriting equation 2.1, after Petterson (1953) and Bolin (1955), as follows:

$$\nabla^2 \psi = -f \pm \sqrt{2\nabla^2 \phi + f^2 + A^2 + B^2 - 2(\nabla f \cdot \nabla \psi)}, \quad 2.9$$

where $A = \psi_{xx} - \psi_{yy}$ and $B = 2\psi_{xy}$ are deformation terms. The positive or negative sign of the radical is applicable in the northern or southern hemisphere respectively. A procedure for solving equation 2.9 for ψ , called the cycle-scan method by Miyakoda (1956, 1960) and Shuman (1957), consists of evaluating each side successively:

$$Q_{ij} = -f \pm \sqrt{2(\nabla^2 \phi)_{ij} + f^2 + A_{ij}^2 + B_{ij}^2}, \quad 2.10$$

$$(\nabla^2 \psi)_{ij} = Q_{ij}. \quad 2.11$$

Here (i,j) are indices of discrete grid points in (x, y) space and we have assumed that f is constant. The ellipticity condition for this case, Equation 2.8, also guarantees that the right hand side of 2.10 is real. A geostrophic approximation for ψ , or one derived from a previous forecast, is usually used to estimate A^2 and B^2 initially in 2.10, the cycle step. The scan step inverts equation 2.11 to determine a corrected ψ field. The process is then repeated using the corrected ψ to determine $A^2 + B^2$ in the cycle step. The procedure is repeated until the ψ input to a given cycle step agrees with the ψ output of the next scan step to within predefined limits.

Miyakoda (1960) discusses the necessity of maintaining integral properties in the consistent formulation of the finite difference algorithms for the ∇^2 operator and the deformation terms of the cycle step. Convergence of the cycle-scan method depends more critically on this formulation than on the precision of the method used to invert equation 2.11 in the scan step, since the cycle step restates the inversion problem with a better estimate of Q_{ij} at each scan. The basic requirement is that successive values of ψ converge.

More complicated simultaneous solutions of the balance equation and the omega equation couple the vertical variation of the flow field for discrete levels. As mentioned earlier, these techniques are more compatible with balance equation models than primitive equation models. In this paper we will apply our techniques only to one layer barotropic models. The simplified treatment of the balance equation approach as outlined above will be sufficient. In the next section we will examine the iterative approaches which use the primitive equations themselves to determine a more compatible initial state.

Three techniques for iterating a damped time scheme around the initial time will be studied. For simplicity of notation, let us adopt the following general form for a set of primitive equations:

$$\frac{d}{dt} U = \mathcal{F} U, \quad 3.1$$

where U is a column vector of the dependent variables and \mathcal{F} is a matrix differential operator which contains no explicit time derivatives. Equation 3.1 states that the operation of \mathcal{F} on the state U is equivalent to the local time derivative of that state. Using this notation we may readily convert Equation 3.1 into its finite difference equivalent. For example, a simple Euler marching scheme may be represented as follows:

$$U^{(\tau+1)} = U^{(\tau)} + F U^{(\tau)} = (I + F) U^{(\tau)}, \quad 3.2$$

where the elements of $U^{(\tau)}$ are now discrete values of the dependent variables at the time $\tau(\Delta t)$ and F is the finite difference equivalent of $\mathcal{F}dt$. The spatial derivatives in \mathcal{F} have been replaced by finite differences and the time increment is absorbed. We will use the symbol, \Rightarrow , to denote the conversion from continuous to discrete variations, $\mathcal{F}dt \Rightarrow F$.

The first method (NH1), uses an Euler-backward time scheme, as suggested by Nitta and Hovermale (1969), in the following steps:

$$U^* = U^{(v)} + F U^{(v)}, \quad 3.3a$$

$$U^{(\mu)} = U^{(v)} + F U^*, \quad 3.3b$$

$$U^{**} = U^{(\mu)} - F U^{(\mu)}, \quad 3.3c$$

$$U^{(v+1)} = U^{(\mu)} - F U^{**}. \quad 3.3d$$

Equations 3.3a-b represent a single forward time step using the Euler-backward scheme. In 3.3c-d the Euler-backward step is reversed to return to the original time, but the state $U^{(v+1)}$ may have different elements than $U^{(v)}$, as denoted by the superscript. A recursion relation between successive values of $U^{(v)}$ can be obtained by combining equations 3.3:

$$U^{(v+1)} = (I + F^2 + F^4) U^{(v)}. \quad 3.4$$

The second method (NH2), also suggested by Nitta and Hovermale (1969), uses a modified Euler-backward time scheme in a similar manner for each iteration:

$$U^* = U^{(v)} + \frac{1}{2} F U^{(v)}, \quad 3.5a$$

$$U^{**} = U^{(v)} + F U^*, \quad 3.5b$$

$$U^{(\mu)} = U^{(v)} + F U^{**}, \quad 3.5c$$

$$\bar{U} = U^{(\mu)} - \frac{1}{2} F U^{(\mu)}, \quad 3.5d$$

$$\bar{\bar{U}} = U^{(\mu)} - F \bar{U}, \quad 3.5e$$

$$U^{(v+1)} = U^{(\mu)} - F \bar{\bar{U}}. \quad 3.5f$$

Again we have explicitly returned to the initial time by reversing the direction of the modified Euler-backward time step in 3.5d-f. The recursion relation for successive values of $U^{(v)}$ at the initial time is:

$$U^{(v+1)} = (I + F^2 - \frac{1}{4} F^6) U^{(v)}. \quad 3.6$$

The third method uses a simple Euler time scheme:

$$U^* = U^{(v)} + F U^{(v)} \quad 3.7a$$

$$U^{**} = U^* - F U^* \quad 3.7b$$

The $(v+1)$ value of U is obtained by linear combination of $U^{(v)}$ and U^{**} , which again both represent different states at the same initial time:

$$U^{(v+1)} = (n+1)U^{(v)} - n U^{**} \quad 3.8$$

where n is a scalar number. This method is attributed to Okamura (Nitta, 1969), who suggested that $n = 2$ would maximize the damping properties of the method, as will be seen. Rivas⁴ has suggested that the method may be more adaptable to the requirements of a specific model if n is allowed to take on a finite sequence of values which are repeated during the iteration process. The added flexibility of this suggestion will become more apparent as we proceed. The recursion relation for this Okamura-Rivas (OR) method is:

$$U^{(v+1)} = (I + nF^2) U^{(v)} \quad 3.9$$

Let us now examine the stability requirements of these methods. Consider a single harmonic wave of frequency ω in time, $U = \hat{U} e^{i\omega t}$, where \hat{U} is a function of position only. Time differencing can be expressed explicitly for this wave:

$$\frac{\partial}{\partial t} U \Rightarrow F U^{(v)} = (i\omega \Delta t) U^{(v)} \quad 3.10$$

⁴Personal communication.

The respective recursion relations become:

$$U^{(v+1)} = [1 - (\omega \Delta t)^2 + (\omega \Delta t)^4] U^{(v)}, \quad (\text{NH1}) \quad 3.11a$$

$$U^{(v+1)} = [1 - (\omega \Delta t)^2 + \frac{1}{4} (\omega \Delta t)^6] U^{(v)}, \quad (\text{NH2}) \quad 3.11b$$

$$U^{(v+1)} = [1 - n(\omega \Delta t)^2] U^{(v)}. \quad (\text{OR}) \quad 3.11c$$

The damping factor, R , for each of these methods can be determined for the single wave case:

$$R = 1 - (\omega \Delta t)^2 + (\omega \Delta t)^4, \quad (\text{NH1}) \quad 3.12a$$

$$R = 1 - (\omega \Delta t)^2 + \frac{1}{4} (\omega \Delta t)^6, \quad (\text{NH2}) \quad 3.12b$$

$$R = 1 - n(\omega \Delta t)^2. \quad (\text{OR}) \quad 3.12c$$

Stability of the iterative methods requires that $|R| < 1$, which places a restriction on the size of the time increment, Δt :

$$\Delta t < \frac{1}{|\omega|}, \quad (\text{NH1}) \quad 3.13a$$

$$\Delta t < \left(\frac{2}{\omega^2}\right)^{\frac{1}{2}}, \quad (\text{NH2}) \quad 3.13b$$

$$\Delta t < \left(\frac{2}{|n|\omega^2}\right)^{\frac{1}{2}}. \quad (\text{OR}) \quad 3.13c$$

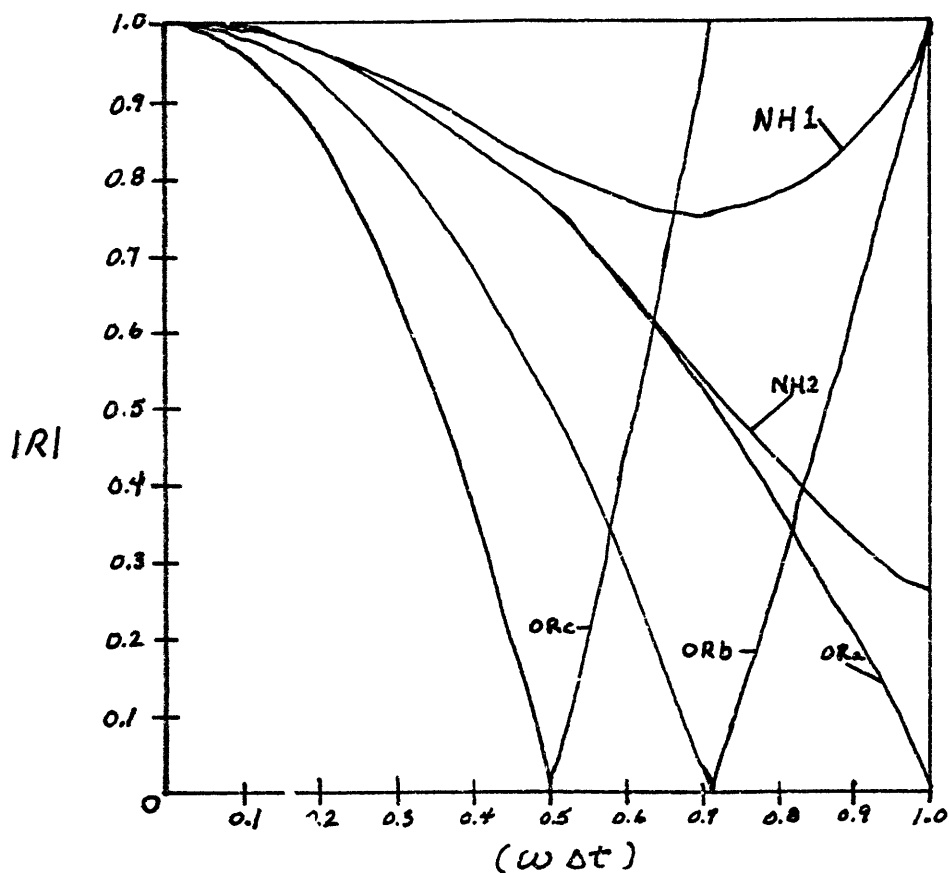
Here it is assumed that n is held constant in the OR method. To apply this type of a restriction to a method using actual data, which may contain a combination of many wave motions, we must obtain from physical considerations an upper limit to the significant gravity-wave frequency

range. This frequency, say ω_{\max} , which is dependent on the model characteristics as well as the physical nature of the data, can then be used to obtain an upper limit on the size of Δt by 3.13. From 3.13c it is now apparent that Okamura's choice of $n = 2$ minimizes the restriction on Δt . Figure 1A shows the variation of $|R|$ with the quantity $(\omega\Delta t)$ for a single oscillation as in Equations 3.12. Note that the NH1 and the OR ($n \geq 2$) methods exhibit the undesirable property that the damping factor increases at the very high ends of the frequency range allowed by 3.13. This can be corrected in the OR method by allowing n to repeatedly take on a sequence of values during the iterative process. The total damping factor for a sequence, say of 1, 1.6, 4, is the product of the damping factors at each n of the sequence. Therefore the damping is retained at high frequencies when $n < 2$ while also strongly damping mid-range frequencies when $n \geq 2$. Thus the OR scheme can be adapted to a greater selectivity in its damping properties to suit a specific model. Figure 1B compares the damping for an equal number of operations of F on U. The methods used and the relevant total damping factors are shown. Note that when n passes through the sequence 1, 1.6, 4 the damping is very similar to the case when n is constant at 2 in the OR method, except that damping at the high frequency end of the range is retained. Thus the condition 3.13c can be obtained for some equivalent n where a sequence is used. The numbers in this sequence shown were specifically chosen to correct the sharp rises in the single- n curves shown in figure 1A.

The operation of F on U in each time step represents a very complex one for sophisticated models although it is not necessary that F be

Figure 1A

Damping properties of the iterative schemes



NH1: $R = 1 - (\omega\Delta t)^2 + (\omega\Delta t)^4$

NH2: $R = 1 - (\omega\Delta t)^2 + \frac{1}{4}(\omega\Delta t)^6$

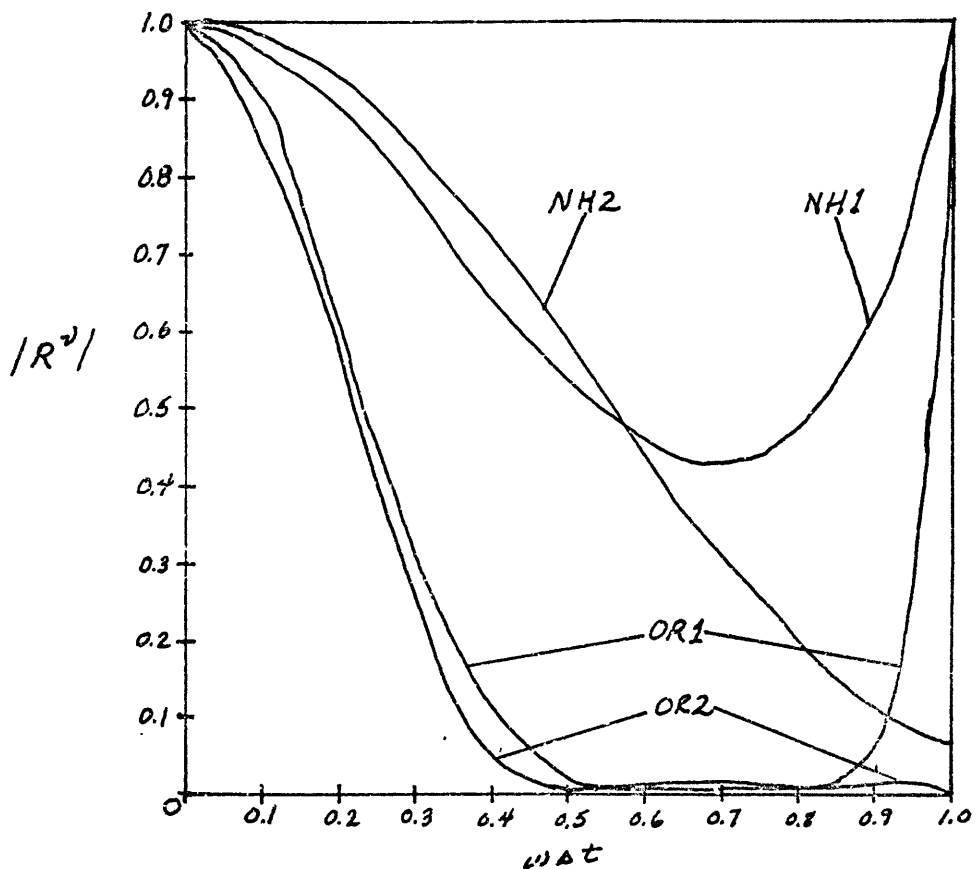
ORa; n=1: $R = 1 - (\omega\Delta t)^2$

ORb; n=2: $R = 1 - 2(\omega\Delta t)^2$

ORc; n=4: $R = 1 - 4(\omega\Delta t)^2$

Figure 1B

Relative efficiency of iterative schemes: F applied 12 times



$$\text{NH1: } \nu=3; R^3 = [1 - (\omega\Delta t)^2 + (\omega\Delta t)^4]^3$$

$$\text{NH2: } \nu=2; R^2 = [1 - (\omega\Delta t)^2 + \frac{1}{4}(\omega\Delta t)^6]^2$$

$$\text{OR1: } n=2; \nu=6; R = [1 - 2(\omega\Delta t)^2]^6$$

$$\text{OR2: } n=1, 1.6, 4; \nu=2; R = \{ [1 - (\omega\Delta t)]^2 [1 - 1.6(\omega\Delta t)^2] [1 - 4(\omega\Delta t)^2] \}^2$$

formulated explicitly. This operation primarily determines the cost of these iterative techniques in terms of computer utilization. Since the number of F operations is the same for all methods in Figure 1B, these curves provide an estimate of the relative efficiency of these schemes. Another important consideration in designing a specific model for the computer might be storage requirements. A rough estimate of the requirements of these methods is the number of sets of the dependent variables, U, which must be stored separately. Table 1 compares the three methods on the basis of these practical considerations.

Table 1

Method	#U matrices per iteration	#F operations per iteration
NH1	2	4
NH2	3	6
OR	3	2

In order to examine the convergence of these schemes, let us temporarily assume that a given model possesses a non-trivial, steady state solution. As the number of iterations increases all non-zero frequency motions are damped and the steady state solution is approached:

$$\lim_{v \rightarrow \infty} U_J^{(v)} = \lim_{v \rightarrow \infty} U_J^{(v+1)} = U_J, \quad 3.14$$

where J may be replaced by the symbols I, II, III to distinguish different numerical values for the dependent variables in the state U obtained by the different methods. These states must satisfy the recursion relations as follows:

$$U_I = (I + F^2 + F^4) U_I, \quad (\text{NH1}) \quad 3.15a$$

$$U_{II} = (I + F^2 - \frac{1}{4}F^4) U_{II} , \quad (\text{NH2}) \quad 3.15b$$

$$U_{III} = (I + nF^2) U_{III} . \quad (\text{OR}) \quad 3.15c$$

The error after the v^{th} iteration for the J^{th} method, $x_J^{(v)}$, is defined as:

$$X_J^{(v)} = U_J - U_J^{(v)} ,$$

and the iterations converge if and only if:

$$\lim_{v \rightarrow \infty} X_J^{(v)} = 0 \quad 3.16$$

In terms of a suitable matrix norm, denoted by the symbol, $|| () ||$, these criteria become:

$$|| I + F^2 + F^4 || < 1 , \quad (\text{NH1}) \quad 3.17a$$

$$|| I + F^2 - \frac{1}{4}F^4 || < 1 , \quad (\text{NH2}) \quad 3.17b$$

$$|| I + nF^2 || < 1 . \quad (\text{OR}) \quad 3.17c$$

The evaluation of Equations 3.17 may be very complex for sophisticated models, but necessary criteria may be derived by using the trace property of the matrices:

$$\text{TR} (F^2 + F^4) < 0 , \quad (\text{NH1}) \quad 3.18a$$

$$\text{TR} (F^2 - \frac{1}{4}F^4) < 0 , \quad (\text{NH2}) \quad 3.18b$$

$$\text{TR} (nF^2) < 0 . \quad (\text{OR}) \quad 3.18c$$

Further analysis is more easily performed with a specific model. For simplicity, consider a linear barotropic system on a flat, rotating plane (f-plane) given by the following equations:

$$\frac{\partial u}{\partial t} = f v - \frac{\partial \phi}{\partial x} \quad 3.19a$$

$$\frac{\partial v}{\partial t} = -f u - \frac{\partial \phi}{\partial y} \quad 3.19b$$

$$\frac{\partial \phi}{\partial t} = -H \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) \quad 3.19c$$

where (u, v) are the deviations of the horizontal velocity components from a basic state of rest, $\phi = gh'$ is the deviation of the geopotential about a mean value H , and f is a constant Coriolis parameter. In the notation of 3.2 a centered, second-order finite difference representation of F for the state $U = \begin{pmatrix} u \\ v \\ \phi \end{pmatrix}$ is:

$$F = \begin{pmatrix} 0 & \beta & -\gamma \nabla_y \\ -\beta & 0 & -\gamma \nabla_x \\ -\gamma H \nabla_x & -\gamma H \nabla_y & 0 \end{pmatrix} \quad 3.20$$

where (∇_x, ∇_y) are the x and y centered differences, $\beta = f\Delta t$, $\gamma = \frac{\Delta t}{2\Delta s}$, and $\Delta s = \Delta x = \Delta y$. Because we have chosen β , γ , and H constant the higher exponents of F are particularly simple:

$$F^2 = \begin{pmatrix} \gamma^2 H \nabla_x^2 - \beta^2 & \gamma^2 H \nabla_x \nabla_y & -\gamma \beta \nabla_y \\ \gamma^2 H \nabla_x \nabla_y & \gamma^2 H \nabla_y^2 - \beta^2 & \gamma \beta \nabla_x \\ \gamma \beta H \nabla_y & -\gamma \beta H \nabla_x & \gamma^2 H \nabla^2 \end{pmatrix},$$

$$F^4 = \alpha F^2, \quad F^6 = \alpha^2 F^2,$$

where $\alpha = \gamma^2 H \nabla^2 - \beta^2$, and $\nabla^2 = \nabla_x^2 + \nabla_y^2$. 3.21

Consider now a simple harmonic wave of mode (p, q) in (x, y) space:

$$U^{(v)} = \tilde{U}^{(v)} e^{i(pj\Delta s + qm\Delta s)}, \tag{3.22}$$

where $x = j\Delta s$ and $y = m\Delta s$. The finite difference operators (∇_x, ∇_y) become (l, k) , where $l = 2i \sin p\Delta s$ and $k = 2i \sin q\Delta s$. For this simple case F becomes an algebraic matrix:

$$F = \begin{pmatrix} 0 & \beta & -\gamma l \\ -\beta & 0 & -\gamma k \\ -\gamma H l & -\gamma H k & 0 \end{pmatrix}, \quad F^2 = \begin{pmatrix} \gamma^2 H l^2 - \beta^2 & \gamma^2 H l k & -\gamma \beta k \\ \gamma^2 H l k & \gamma^2 H k^2 - \beta^2 & \gamma \beta l \\ \gamma \beta H k & -\gamma \beta H l & \gamma^2 H (l^2 + k^2) \end{pmatrix},$$

$$F^4 = \alpha F^2, \quad F^6 = \alpha^2 F^2, \quad \alpha = \gamma^2 H (l^2 + k^2) - \beta^2.$$

Note also that the trace of F^2 is simply expressed and is always less than zero for this mode:

$$\begin{aligned} \text{Tr}(F^2) &= 2[\gamma^2 H (l^2 + k^2) - \beta^2] = 2\alpha \\ &= -2(\Delta t)^2 \left\{ f^2 + \frac{H}{(\Delta s)^2} [\sin^2(p\Delta s) + \sin^2(q\Delta s)] \right\} \end{aligned} \tag{3.23}$$

The necessary conditions for convergence, 3.18, now become:

- $2\alpha(1+\alpha) < 0, \text{ or } -\alpha < 1$ (NH1)
- $2\alpha(1-\frac{1}{4}\alpha^2) < 0, \text{ or } |\alpha| < 2$ (NH2)
- $2\alpha n < 0, \text{ or } n > 0$ (OR)

We obtain explicit restrictions on Δt for the first two methods by expanding α :

$$\Delta t < \left\{ f^2 + \frac{H}{(\Delta s)^2} [\sin^2(p \Delta s) + \sin^2(g \Delta s)] \right\}^{-1/2}, \tag{NH1} 3.24a$$

$$\Delta t < \left\{ f^2 + \frac{H}{(\Delta s)^2} [\sin^2(p \Delta s) + \sin^2(g \Delta s)] \right\}^{-1/2} \sqrt{2}. \tag{NH2} 3.24b$$

For the OR method the necessary condition that $n > 0$ is rather obvious, but it seems to be insufficient for our purpose. We may easily correct this by noting the analogy between the right hand sides of 3.24 and those of 3.13a-b. If we maximize the quantity in braces we will obtain the following conditions necessary for convergence:

$$\Delta t < \left[f^2 + \frac{2H}{(\Delta s)^2} \right]^{-1/2}, \tag{NH1} 3.25a$$

$$\Delta t < \sqrt{2} \left[f^2 + \frac{2H}{(\Delta s)^2} \right]^{-1/2}. \tag{NH2} 3.25b$$

Since the conditions 3.13a-b are both necessary and sufficient it would seem that 3.25a-b are also and that $\omega_{\max} = \left[f^2 + \frac{2H}{(\Delta s)^2} \right]^{1/2}$. Since ω_{\max} is not dependent on the iterative method, we may assume that the necessary and sufficient condition for convergence of the CR method is analogous to 3.13c:

$$\Delta t < \left[\frac{2}{n} \left(f^2 + \frac{2H}{(\Delta s)^2} \right) \right]^{1/2}. \tag{OR} 3.25c$$

Note the similarity between the ω_{\max} derived by numerical analysis and that which is determined from perturbation analysis for surface gravity-inertial waves in a linear barotropic model.

The physical significance of the convergence of the iterative methods can be carried one step further in the case of this model. By Equations 3.15 the steady-state U_j must be a solution of a complete set

of homogeneous equations:

$$(1+\alpha) F^2 U_I = 0 \quad (\text{NH1}) \quad 3.26a$$

$$(1-\frac{1}{4}\alpha^2) F^2 U_{II} = 0 \quad (\text{NH2}) \quad 3.26b$$

$$n F^2 U_{III} = 0 \quad (\text{OR}) \quad 3.26c$$

In our model the determinant of F^2 is always zero for the case of a single mode (p, q) . The resulting non-trivial solution is in geostrophic balance, in finite difference form, provided the conditions 3.25 are satisfied for each method:

$$U_J = \begin{pmatrix} -\frac{1}{\beta} \frac{\partial \phi_J}{\partial x} \\ + \frac{1}{\rho} \phi_J \\ \phi_J \end{pmatrix} = \begin{pmatrix} -\frac{1}{f} \frac{\partial \phi_J}{\partial s} \\ + \frac{1}{f} \frac{\partial \phi_J}{\partial \Delta s} \\ \phi_J \end{pmatrix}, \quad J=I, II, III$$

Geostrophy is also the analytical solution for steady-state motion by perturbation methods for this model.

Now let us consider the case where a model does not have a non-trivial steady state solution. The inclusion of variable Coriolis parameter or map factors are significant cases where this may occur. These iterative methods will presumably approach a trivial state if continued long enough. For such models it will be necessary to terminate the iterative process before the meteorologically significant motions are also damped out. These techniques would then be most effective for motions in which there is a large gap in frequency between gravity-inertial waves and meteorologically significant waves.

Although the preceding analysis is much simplified by our model choice, it is intended that the conditions 3.25 may be useful as a first guess for more complex models in the same manner that similar arguments are used in restricting the size of Δt for marching schemes. It should also be emphasized that these procedures do not contain any restriction equivalent to the ellipticity constraint of the balance equation approach.

One further point is of significance. Nowhere in the above analysis is it required that the geopotential field or velocity field be fixed while the other is allowed to reach balance. This further restriction may be imposed by restoring the initial geopotential field, for example, after each iteration. The conditions 3.25 remain the same, and the procedure converges to a steady-state geostrophic balance, but the numerical values of the fields would naturally be different. As will be seen from the numerical experiments, this procedure is somewhat questionable. It is important to remember that these iterative methods implicitly rely on the adjustment process, particular to a given model, to seek a state of balance.

4. Design of Numerical Experiments

To test the various initialization techniques a non-linear, shallow-water model is used, defined by the following equations:

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} = f v - g \frac{\partial h}{\partial x} \quad 4.1a$$

$$\frac{\partial v}{\partial t} + u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} = -f u - g \frac{\partial h}{\partial y} \quad 4.1b$$

$$\frac{\partial h}{\partial t} + u \frac{\partial h}{\partial x} + v \frac{\partial h}{\partial y} = -h \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) \quad 4.1c$$

where h is the height of the free surface, and (u, v) are the horizontal velocity components. The Coriolis parameter, $f=10^{-4} \text{sec}^{-1}$, is constant. The actual numerical model is derived from the flux form of these equations:

$$\frac{\partial \phi u}{\partial t} = - \frac{\partial \phi u u}{\partial x} - \frac{\partial \phi u v}{\partial y} + f \phi v - \phi \frac{\partial \phi}{\partial x} \quad 4.2a$$

$$\frac{\partial \phi v}{\partial t} = - \frac{\partial \phi u v}{\partial x} - \frac{\partial \phi v v}{\partial y} - f \phi u - \phi \frac{\partial \phi}{\partial y} \quad 4.2b$$

$$\frac{\partial \phi}{\partial t} = - \frac{\partial \phi u}{\partial x} - \frac{\partial \phi v}{\partial y} \quad 4.2c$$

where $\phi = gh$. In order to avoid non-linear instability, an energy-conserving scheme is chosen to evaluate the right hand sides of Equations 4.2, following the method developed by Lilly (1965), Bryan (1966), and Rivas (1971). The finite-difference equations at a grid point (i, j)

are:

$$\begin{aligned} \left(\frac{\partial \phi u}{\partial t}\right)_{ij} &= -D_x(\bar{u}^x \phi \bar{u}^x)_{ij} - D_y(\bar{u}^y \phi \bar{v}^y)_{ij} + f(\phi v)_{ij} - \phi_{ij} D_x(\bar{\phi}^x)_{ij}, \\ \left(\frac{\partial \phi v}{\partial t}\right)_{ij} &= -D_x(\bar{v}^x \phi \bar{u}^x)_{ij} - D_y(\bar{v}^y \phi \bar{v}^y)_{ij} - f(\phi u)_{ij} - \phi_{ij} D_y(\bar{\phi}^y)_{ij}, \\ \left(\frac{\partial \phi}{\partial t}\right)_{ij} &= -D_x(\phi \bar{u}^x)_{ij} - D_x(\phi \bar{v}^y)_{ij}, \end{aligned} \quad 4.3$$

where the average and difference operators are defined as follows:

$$(\bar{A}^x)_{ij} = \frac{1}{2} (A_{i+\frac{1}{2},j} + A_{i-\frac{1}{2},j}), \quad 4.4a$$

$$D_x A_{ij} = \frac{1}{\Delta s} (A_{i+\frac{1}{2},j} - A_{i-\frac{1}{2},j}). \quad 4.4b$$

Analogous definitions are made for the y-direction, and $\Delta x = \Delta y = \Delta s = 250$ km.

To formulate the energy equation we use the following relationships:

$$\left(\frac{\partial \phi u^2/2}{\partial t}\right)_{ij} = u_{ij} \left(\frac{\partial \phi u}{\partial t}\right)_{ij} - u_{ij}^2/2 \left(\frac{\partial \phi}{\partial t}\right)_{ij}$$

$$\left(\frac{\partial \phi v^2/2}{\partial t}\right)_{ij} = v_{ij} \left(\frac{\partial \phi v}{\partial t}\right)_{ij} - v_{ij}^2/2 \left(\frac{\partial \phi}{\partial t}\right)_{ij}$$

$$\left(\frac{\partial \phi^2/2}{\partial t}\right)_{ij} = \phi_{ij} \left(\frac{\partial \phi}{\partial t}\right)_{ij}$$

The definitions 4.4 and their use in Equations 4.3 are chosen such that the total energy, summed over a finite grid, is changed only by a flux through the boundaries or by sources and sinks:

$$\frac{d}{dt} TE \propto \sum_{ij} \left[\left(\frac{\partial \phi u^2/2}{\partial t} \right)_{ij} + \left(\frac{\partial \phi v^2/2}{\partial t} \right)_{ij} + \left(\frac{\partial \phi^2/2}{\partial t} \right)_{ij} \right] \quad . 29 .$$

= Boundary terms + sources - sinks.

This formulation conserves the total energy aside from truncation errors in a marching scheme.

The purpose of our numerical experiments is to compare the iterative and balance-equation solutions to an artificially generated initial state which is as closely in balance as possible. The velocity and height fields of the balanced data will be altered or perturbed by various methods and the initialization techniques used to restore balance. We will be interested in whether the unperturbed balance is recovered by the initialization. We will also compare the height variation at an arbitrarily chosen, fixed point, P, as a forecast is made from the unperturbed and initialized data.

The initial, balanced state is itself generated by integrating the model for a finite time, T, with an artificial source term, S(x, y, t), added to Equation 4.1c. The integration is started at rest with a level surface at 3 km. A leapfrog marching procedure is used with a forward Euler step every 24 leapfrog steps to avoid the separation of fields at odd and even steps. The spatial variation of the source function is a double sine wave in x and y:

$$S(x, y, t) = \hat{S}(t) \sin\left(\frac{2\pi}{L} x\right) \sin\left(\frac{2\pi}{L} y\right),$$

where L = 4000 km. It is convenient to visualize the state produced as a checkerboard pattern of highs and lows extending periodically over the

infinite f-plane. Numerical computations are made on a rectangular grid which extends one wavelength of the source function in the x-direction (east) and half a wavelength in the y-direction (north). Therefore we are concerned only with the unit cell of the checkerboard pattern consisting of one low and one high. The boundary conditions are periodic, but the north-south boundaries are matched diagonally to preserve the checkerboard periodicity. The grid specifications and boundary conditions are shown in Figure 2. Equations 4.3 are evaluated in the interior region enclosed by solid lines. The unit cell also includes the boundary points connected by dashes. Note that with periodic boundary conditions, the total energy is changed by the source term only. Also the source adds no net mass (geopotential) to the system due to its sinusoidal variation.

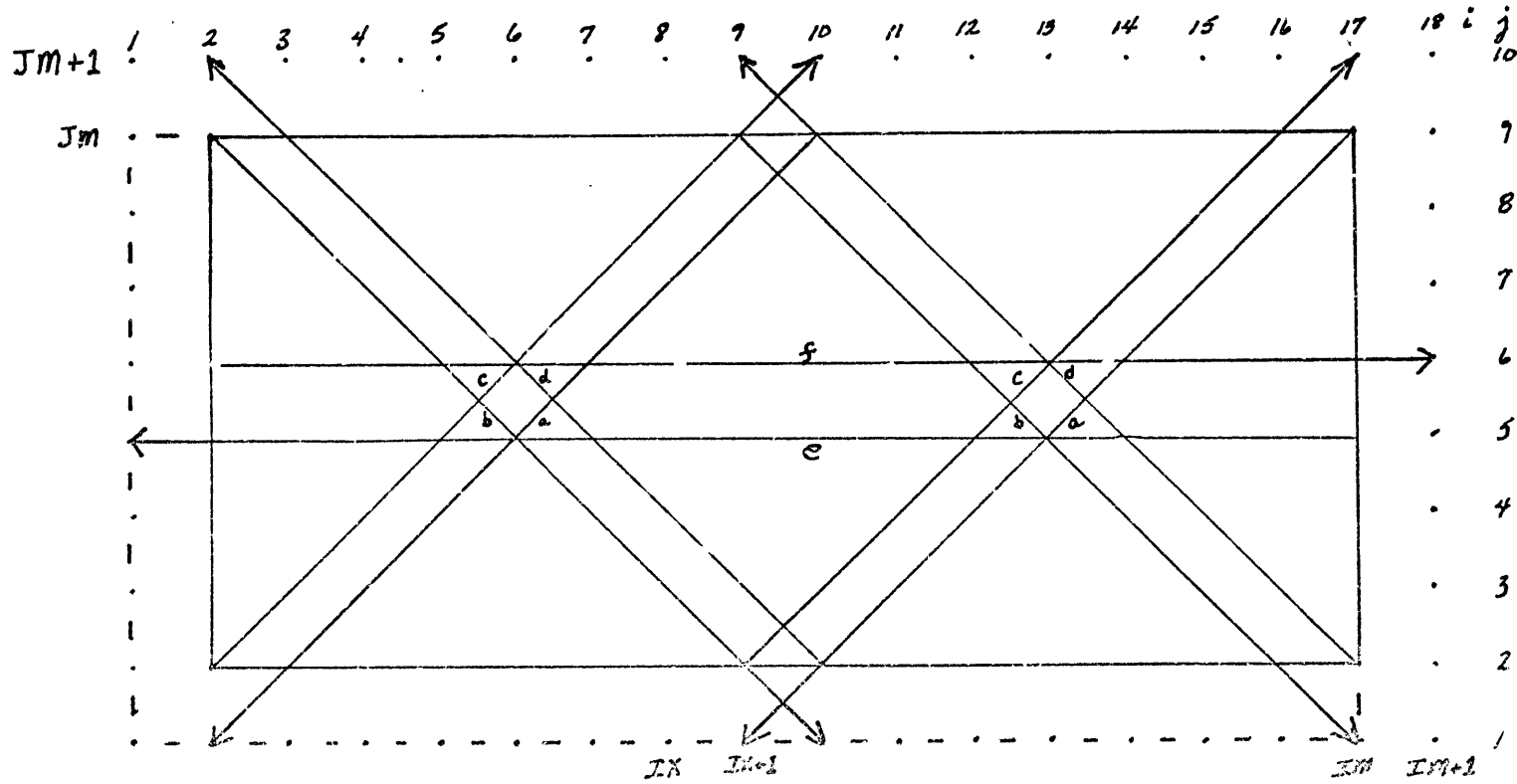
The strength, $\hat{S}(t)$, of the source function is chosen such that the resulting state of balance will assume the role of a meteorologically significant wave. The deviation of the free surface from its mean height is shown in Figure 3A. The contours are drawn at 6 decameter intervals, the low is 34 dekameters below the mean height, and the high is 15 dekameters above it. The associated velocity fields (not shown) are anti-cyclonic around the low and cyclonic around the high with a maximum speed of about 30 m/sec.

The balance of these initial fields is critically dependent on the time variation of the source term. The field shown in Figure 3A is generated with the following form:

$$\hat{S}(t) = \frac{A\pi}{2T} \sin\left(\frac{\pi}{T}t\right) \quad 4.5$$

Figure 2

Grid System and Boundary Conditions:



Active Region: 16×8 grid points; $i = 2, 3, \dots, 17$; $j = 2, 3, \dots, 9$.

Unit Cell: 17×9 grid points; $i = 1, 2, \dots, 18$; $j = 1, 2, \dots, 10$.

Boundary Conditions:

a. $A(i, 1) = A(i + 16, 10)$

b. $A(i + 16, 1) = A(i, 10)$

c. $A(i, 11) = A(i + 16, 2)$

d. $A(i + 16, 11) = A(i, 2)$

$i = 2, 3, \dots, 17$

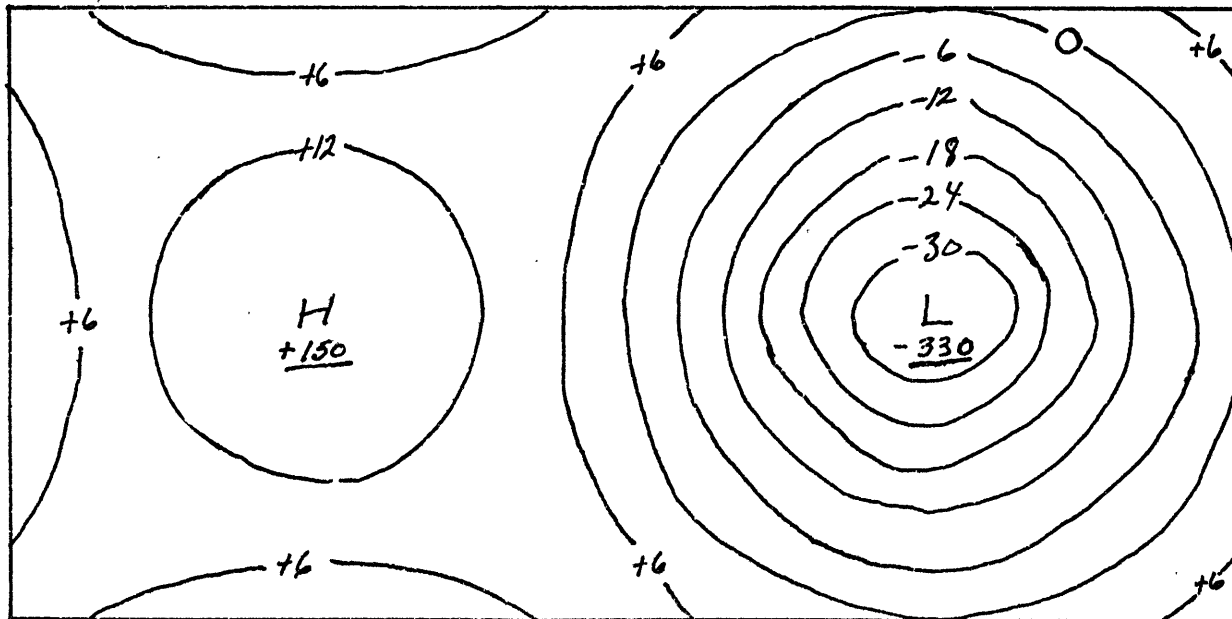
e. $A(1, j) = A(17, j)$

f. $A(18, j) = A(2, j)$

$j = 1, 2, \dots, 11$

Figure 3A

Height deviation field of reference state



Contours labelled in dekameters;
Low and high centers labeled in meters.

where the integration is terminated at time, $T = 8$ days, and $A = 1.01 \times 10^4 \text{ m}^2/\text{sec}^2$ is the integrated strength. This sinusoidal form was found to yield the most balanced state. However, it is also necessary to add the source in very small increments ($\Delta t = 5$ min) in order to minimize the generation of imbalances in the state produced. Figure 3B indicates the degree to which balance is achieved. The variation at the point P, where $i = 4$, $j = 4$, is shown, on a very expanded scale, as a forecast is made from the initial state. The leapfrog scheme is again used with $\Delta t = 12$ min. Note that the amplitude of the gravity waves is only about 0.2 meters. This indicates a very high degree of balance. Similar attempts to generate initial data were made using an exponential form and a linear form for the source strength. The integrated strength was the same. The resultant forecasts showed gravity waves with amplitudes about 25 and 100 times greater, respectively.

The initial fields generated by the sinusoidal variation of the source strength serve as a reference balance which we seek in our initialization procedures. The initial fields and the 48-hour-forecast fields are used as standards in computing rms departures of the velocity and height over the interior region.

The balance-equation initialization follows "Scheme C" of Miyakoda (1960). Successive over-relaxation is used to invert equation 2.11. A geostrophic first guess for the stream function, ψ , is obtained from the geopotential field. Following Miyakoda's suggestion, the ψ -field obtained from each scan is compared to the average of the two previous ψ -fields to test for convergence. The procedure is stopped when $\psi_{i,j}$ agrees to within four significant digits at all points.

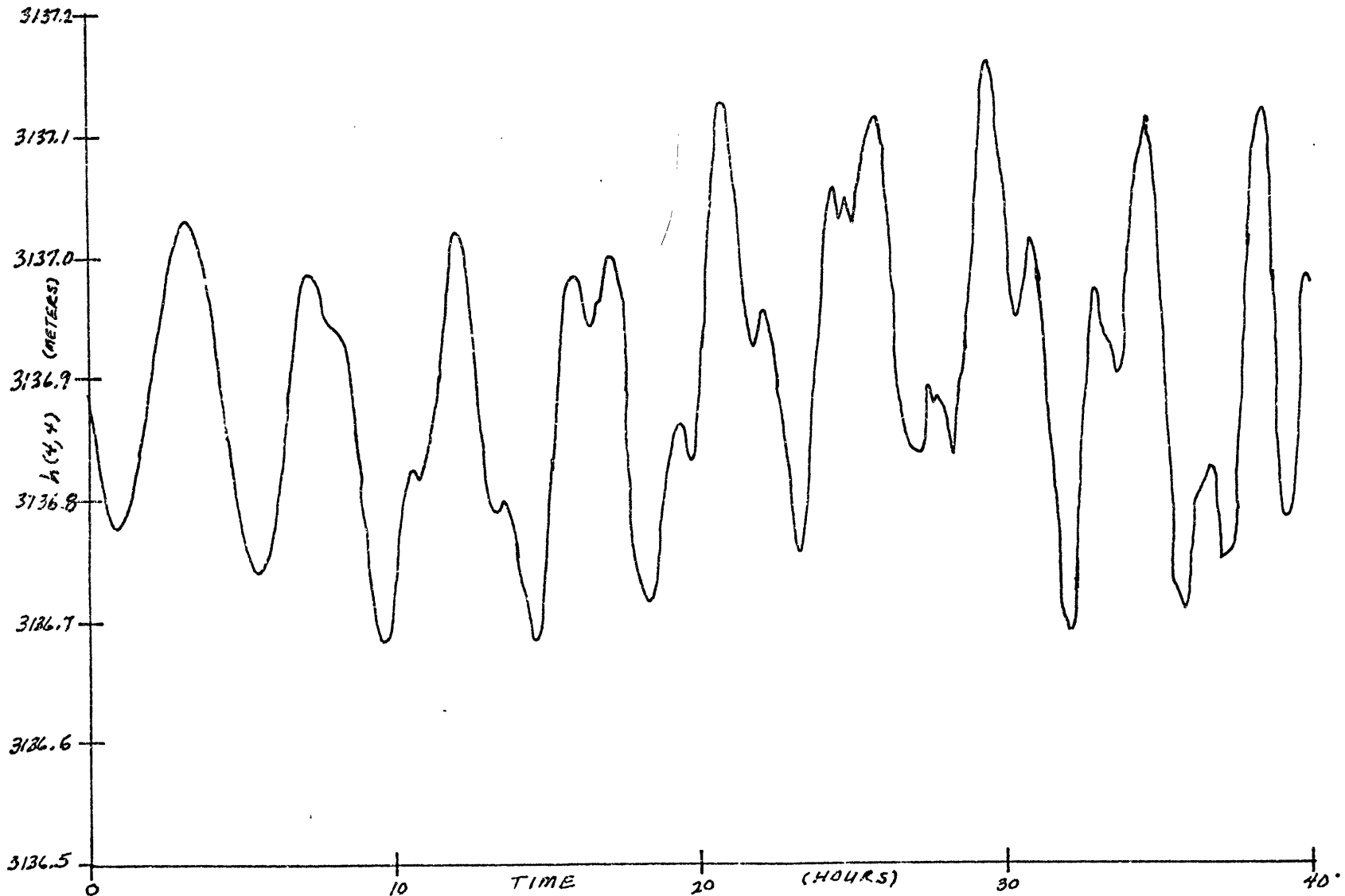



Figure 3B
 Height variation at the point (4, 4) during forecast from reference state

The iterative techniques are carried out as in Section 3. The NH1 technique uses an Euler-backward time scheme as in Equations 3.3, while NH2 uses a modified Euler-backward scheme as in Equations 3.5. Two versions of the Okamura-Rivas scheme will be examined, using the Euler time step as in Equations 3.7. For the first version (OR1), the scalar n in Equation 3.8 has the value 2 for each iteration. In the second (OR2), n scales through the sequence 1, 1.6, 4 successively as the iterations proceed. The option of restoring the geopotential field after each iteration is incorporated into the iterative methods. The exercise of this option will be seen to have a significant effect on the convergence rate of the iterative approaches.



5. Results

5.1 Introduction

By carefully applying artificial forcing to a shallow-water-equation model, as described in Section 4, initial geopotential and velocity fields have been generated which are in balance for all practical purposes. This state assumes the role of a synoptic wave in our idealized system. If we integrate forward in time, again using the leapfrog schemes described in Section 4, but with the source term removed, this synoptic wave is approximately stationary. This steady-state flow, regardless of its simplicity, represents the meteorologically significant motion in our system and will serve as the reference state. Note that the amplitude of the synoptic wave is about 250 meters in the mean, and the maximum wind speed is about 30 m/sec.

In order to introduce initial imbalances, which represent the effect of observational errors, the reference fields are perturbed in some way. For our first experiment, we assume the "observed" geopotential field contains no error, but that the error in the velocity field is very large. Therefore we will use directly the geopotential field of the reference data and perturb the balance by replacing the velocity field by one which is geostrophically determined from the geopotential field. The perturbation is quantitatively evaluated by determining the rms departure of the perturbed velocity field from the reference velocity field over the

interior region of the grid:

$$\delta V_{rms} = \left[\frac{1}{(IM-1)(JM-1)} \sum_{i=2}^{IM} \sum_{j=2}^{JM} (u_{ij}' - u_{ij})^2 + (v_{ij}' - v_{ij})^2 \right]^{\frac{1}{2}} \quad 5.1a$$

where the primes denote the perturbed fields. The various initialization techniques can then be applied to the perturbed state in order to evaluate the reduction in this rms velocity error. Since we propose to allow the geopotential field adjust during some of our experiments we will also measure the departure of the height field from the reference height field in a similar way:

$$\delta h_{rms} = \left[\frac{1}{(IM-1)(JM-1)} \sum_{i=2}^{IM} \sum_{j=2}^{JM} (h_{ij}' - h_{ij})^2 \right]^{\frac{1}{2}} \quad 5.1b$$

The rms errors in the height and velocity fields will measure the extent to which the reference fields are recovered by a given initialization procedure but not the extent to which the original balance is restored. In order to evaluate the balance achieved after initialization, we will measure the amplitude of gravity waves at a point P, arbitrarily chosen to be the grid point $i = j = 4$, as a forecast is made from the initialized state. In Figure 3B the reference forecast at this point showed a height variation with less than 0.2 meters amplitude. This variation is of negligible consequence to our 250 meter synoptic wave. Any oscillation of the height at P which has a significantly larger amplitude will be interpreted as residual gravity-inertial waves arising from imbalance which has not been eliminated from the perturbed state by the initialization procedure. Although the amplitude of gravity waves at a single

point may not represent those of the entire grid we are interested chiefly in relative comparisons. We also measure the rms errors in the height and velocity fields after a 48-hour forecast. The 48-hour fields forecasted from the reference state are used as standard (unprimed) fields in Equations 5.1.

To directly simulate observational errors, we will also perturb the initial fields by adding a field of normally distributed random numbers to the balanced fields. The mean and standard deviation of these random numbers can be specified to represent typical errors in atmospheric measurements. The remainder of this experiment proceeds as in the case of the geostrophically perturbed state. A further attempt was also made to correct to geostrophically determined field by means of an approximation based on the gradient wind equation. The justification of this correction is perhaps best left for later discussion.

In the results to be given here, four iterative methods are compared both to the balance-equation initialization and among themselves. They are the NH1, NH2, OR1, and OR2 methods described in detail earlier. Note that the NH1 and OR1 schemes do not damp the highest frequency gravity waves while the NH2 and OR2 methods do provide sufficient damping at this frequency. The damping properties of these methods were shown in Figure 1.

In order to maximize the damping of each of the iterative methods the time increments used were separately determined to be the maximum, in whole minutes, for which that method remained stable using the non-linear, shallow-water model. Table 2 compares these experimentally determined values to the upper limits of the linear stability criteria as expressed

Table 2

Method	Δt (nonlinear)	Δt (linear)
NH1	16 min	17 min
NH2	22 min	24 min
OR1	16 min	17 min
OR2	17 min	--
Leapfrog (forecast)	12 min	17 min

in 3.25. A similar comparison is also made for the leapfrog forecast scheme. Note that the iterative schemes in general seem to adhere more closely to the linear criteria than does the leapfrog scheme. An exact linear criterion for the OR2 method is not readily available. In this case the sequence, 1, 1.6, 4, of the values for n in Equation 3.8, was chosen by graphical inspection of Figure 1 to specifically correct the lack of damping at very high frequencies in OR1, while providing more efficient damping at intermediate frequencies. This sequence is not necessarily an optimum one but is used to illustrate the improvement which can be achieved in the Okamura-Rivas scheme by allowing n to vary.

In applying the four iterative methods a significant difference will be seen between the case where the geopotential field is restored after each iteration and the case where the geopotential field is freely allowed to adjust with the velocity field. This difference is examined in the results of initializing the geostrophically perturbed state.

5.2 Initialization of the Geostrophically Perturbed State.

The geostrophically determined velocity field departs from the reference velocity field by an rms error of 7.7 m/sec. Figure 4A shows the height variation at the point P as the forecast proceeds without further initialization. The gravity waves generated by the geostrophic perturbation have a maximum amplitude of about 125 meters and are sufficient to strongly distort the synoptic wave, which has a 250 meter amplitude. After 48 hours the rms velocity error is about 8 m/sec while an rms error of 29 meters has developed in the height field. Since initial imbalances cause distortions of the initial tendencies it is the amplitude of the height variation that better represents the amount of imbalance in the perturbed state.

If the ellipticity condition, 2.8, is everywhere satisfied, the balance-equation can be solved without altering the geopotential field. Our reference state is of sufficient amplitude to cause the ellipticity constraint to be violated at a few points around the perimeter of the high. A correction is made to the geopotential field to satisfy the ellipticity condition. The left hand side of expression 2.8 is evaluated, at all grid points, as follows:

$$\chi_{ij} \equiv \frac{4(\bar{\phi}_{ij} - \phi_{ij})}{(\Delta S)^2} + \frac{f^2}{2}, \quad 5.2a$$

where $\bar{\phi}_{ij}$ is the average geopotential of the four grid points adjacent to (i, j). At those points where $\chi_{ij} < 0$, the value of ϕ_{ij} is decreased

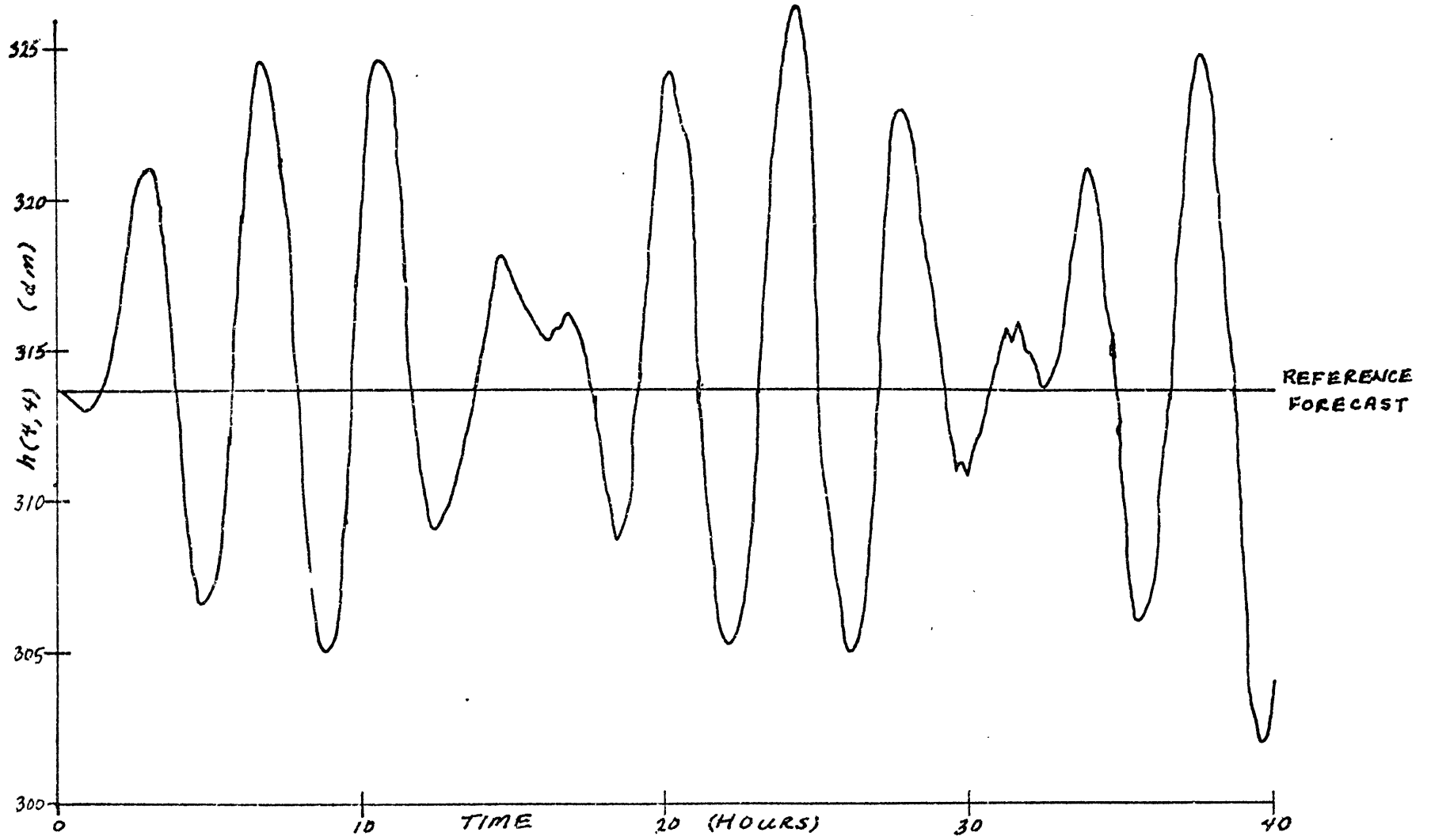


Figure 4A

Height forecast at (4, 4) after geostrophic perturbation

by the following formula:

$$\phi'_{ij} = A \phi_{ij} + B \bar{\phi}_{ij}, \quad 5.2b$$

subject to the restrictions that:

$$A + B = 1 \quad 5.2c$$

$$\frac{4(\bar{\phi}_{ij} - \phi'_{ij})}{(\Delta S)^2} + \frac{f^2}{2} = -\alpha \chi_{ij}, \quad 5.2d$$

where α is a small non-negative number. After replacing ϕ_{ij} with ϕ'_{ij} at the relevant points, new x_{ij} are calculated at all points. The process is then repeated until all x_{ij} are non-negative, hence the ellipticity condition is everywhere satisfied. In the present case, with $\alpha = 0$, this process is repeated 5 times. The height field is corrected at a total of 8 interior grid points and the maximum change is 0.5 meters. The resulting rms error in the height field is about 0.09 meters.

After 28 cycle-scans of the balance equation the stream function converged to our specifications (Section 4). The rms velocity error was reduced to 0.7 m/sec. Figure 4B shows the residual gravity waves in a forecast made from the initialized state. The vertical scale here is 8 times that of Figure 4A. The maximum amplitude is about 3 meters. After 48 hours the rms velocity and height errors are 1 m/sec and 1.5 m respectively. The balance equation does provide a significant restoration of both the initial fields and their state of balance, in this case.

In additional experiments, we varied the strength of the source used to generate the reference state. Stronger flow patterns do not

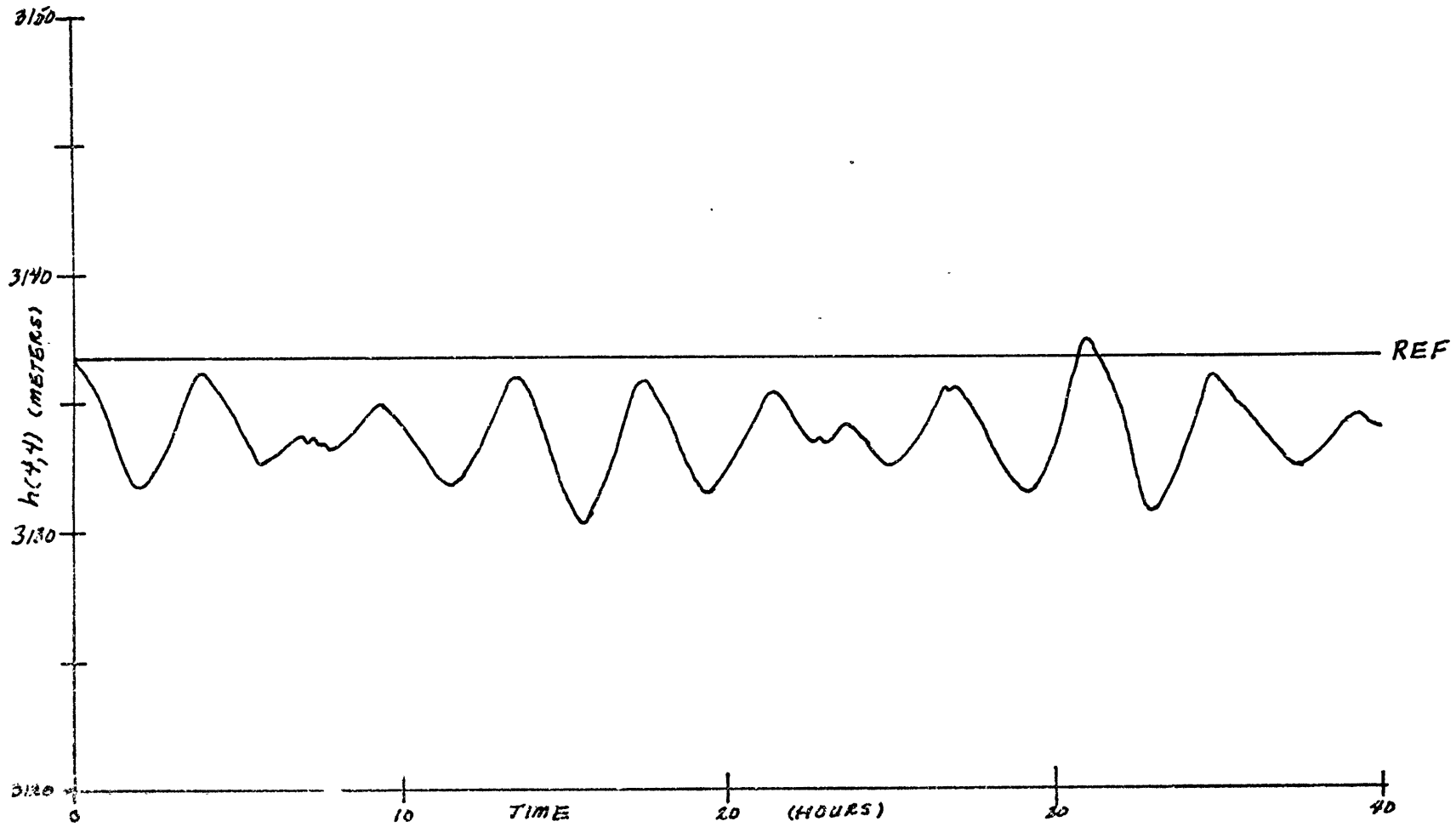


Figure 4B

Height forecast at (4, 4) after balance equation initialization

affect the iterative initialization methods, but the ellipticity constraint on the balance-equation procedure is violated at an increasingly large number of points. The procedure for "correcting" the geopotential field described in Equations 5.2 becomes insufficient to satisfy the ellipticity condition as this number increases. Therefore we were unable to obtain solutions to the balance equation for sources that were about 20% stronger. We will return to this point in later experiments.

For each of the iterative methods, 150 iterations were performed. If we assume that each evaluation of equations 4.3 is equivalent to one time step in a forward forecast, and use the leapfrog time increment, the equivalent time traversed in 150 iterations is 5 days for the NH1 method, 7.5 days for NH2, and 2.5 days for the OR methods. This represents a considerable equivalent forecast, especially for the NH methods. Figure 5A shows the decrease in the rms velocity errors for each method as the number of iterations increases. Here the reference geopotential field is restored after each iteration. The horizontal lines represent the rms errors of the geostrophic perturbation and the balance-equation initialization. A considerable decrease in the rms velocity errors is shown. After 150 iterations this error is reduced to 2.7 m/sec (NH1), 1.8 m/sec (NH2), 1.3 m/sec (OR1), and 1.1 m/sec (OR2). However, 150 iterations do not seem sufficient to cause any method to converge to a steady value of the rms error.

Figure 5B shows the same kind of plot where the geopotential fields are allowed to adjust. Only the small area shown in the upper left corner of Figure 5A has been represented in Figure 5B to better resolve the

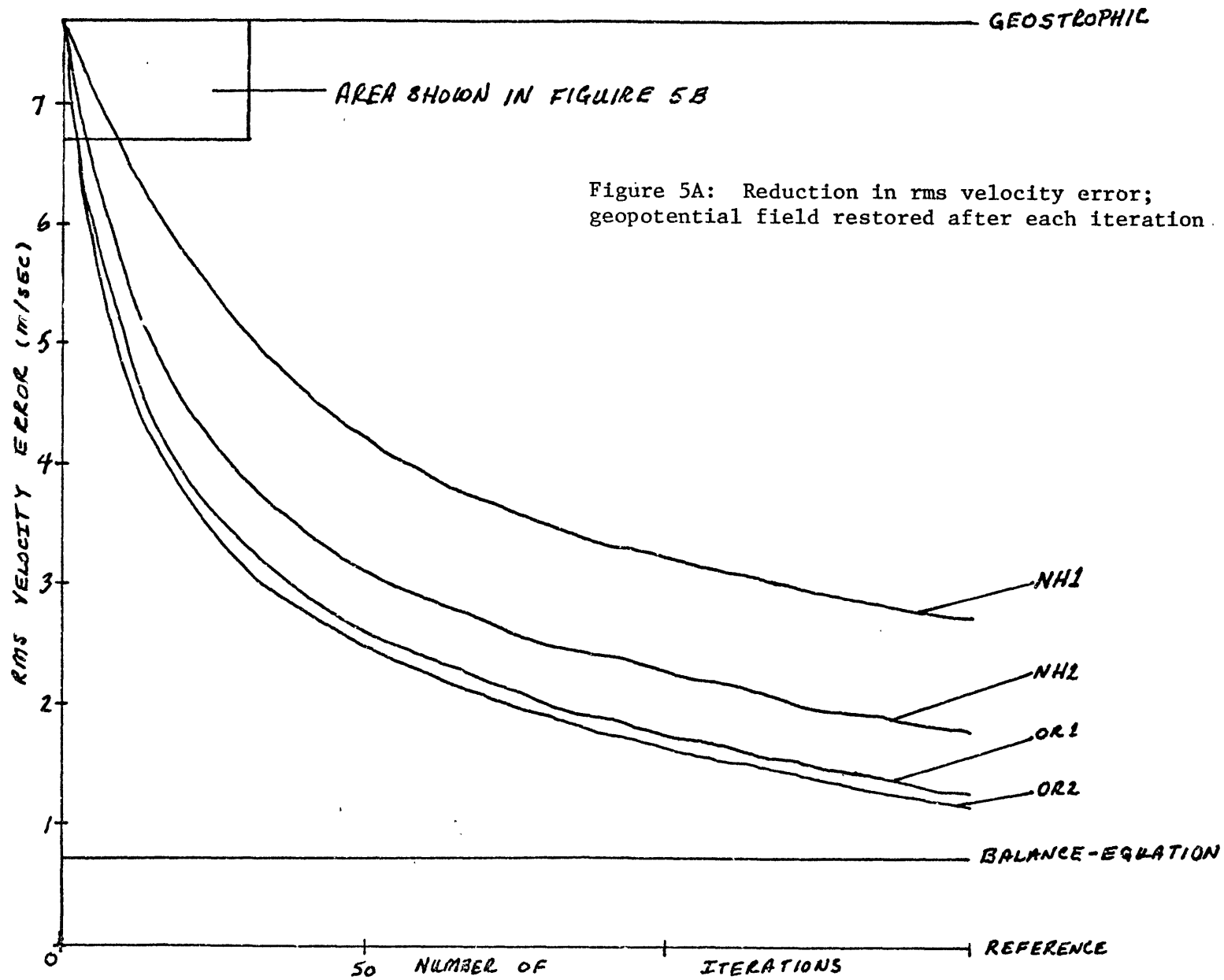


Figure 5B: Reduction in rms velocity error, Geopotential field adjusts

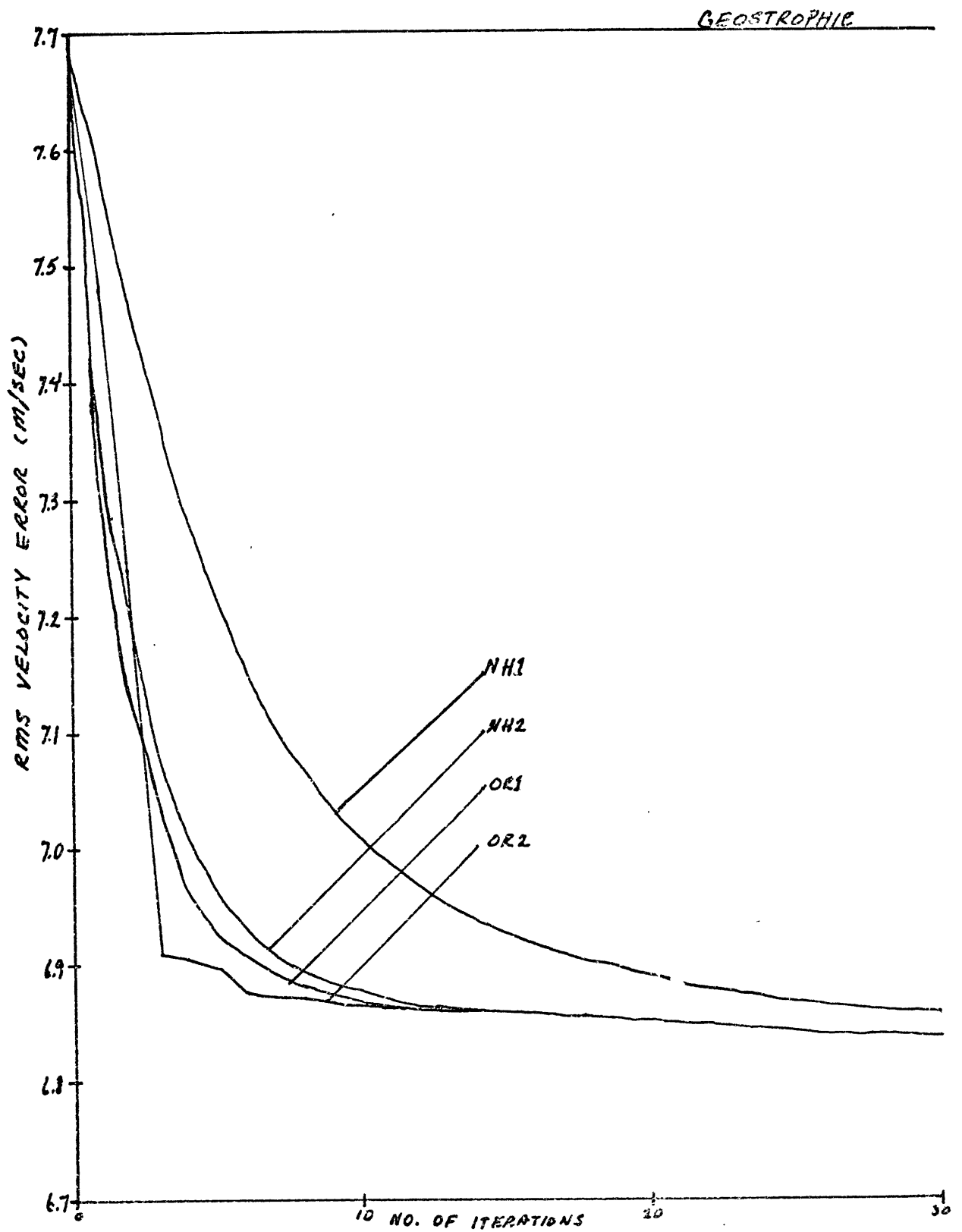
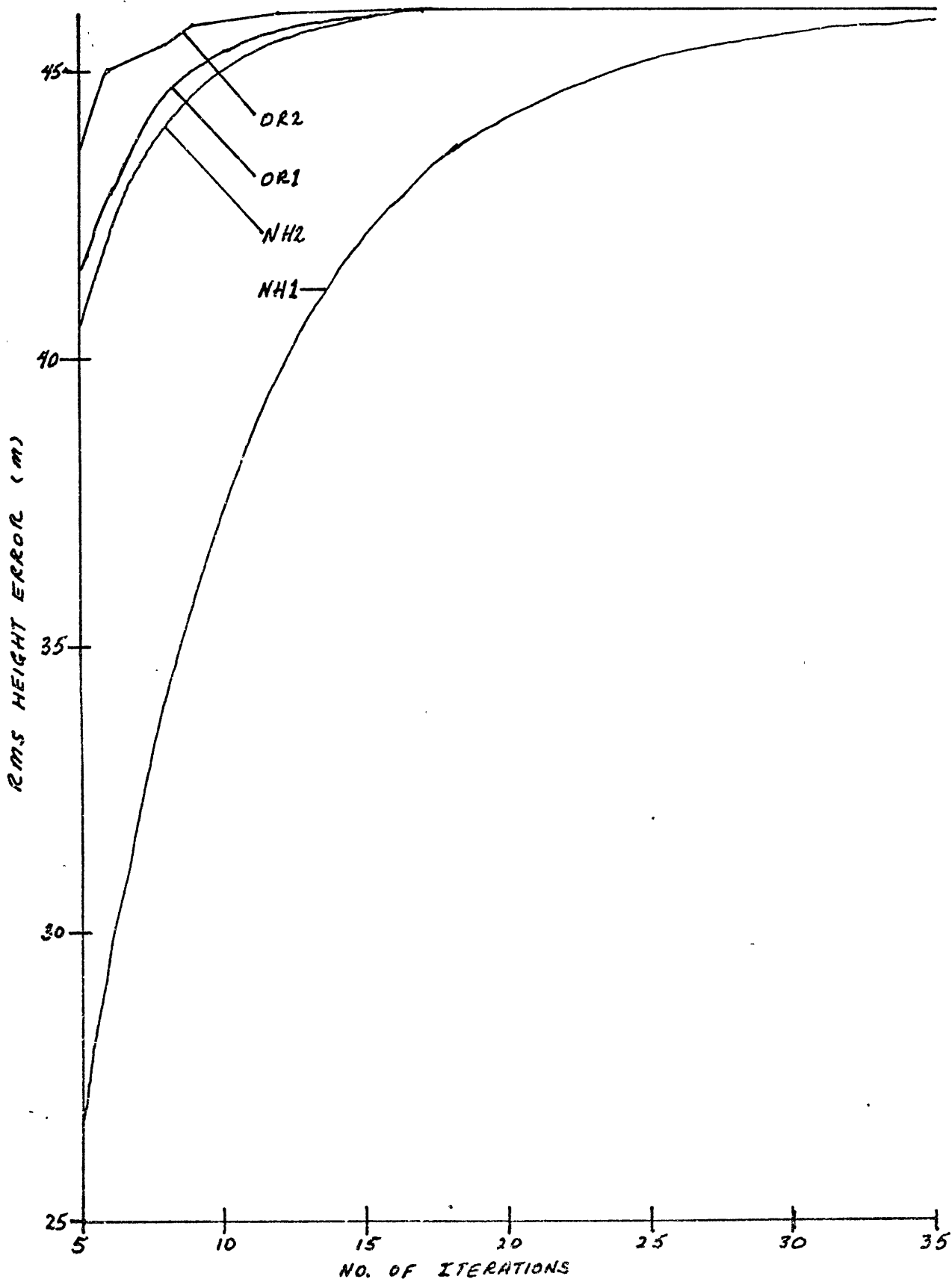


Figure 5C: Increase of rms height error, as geopotential field adjusts 47.



different methods. Only 30 iterations are shown. In this case the geopotential field is altered during the initialization process. Figure 5C shows the resulting increase in the rms height error. Again only the interesting portion of the variation is shown. The immediately obvious result of allowing the geopotential field to adjust is that the rms errors converge very rapidly to steady values, but these values represent a significant departure from the reference state. All four methods reach a steady rms velocity error of about 6.9 m/sec and create a steady rms height error of 46 meters. These values remain steady throughout the full 150 iterations. To compare the convergence rates we estimate the number of iterations required to reach the steady values: 40 (NH1), 15 (NH2), 15 (OR1), and 12 (OR2). The equivalent time traversed during these iterations is 32 hours (NH1), 18 hours (NH2), 6 hours (OR1), and 5 hours (OR2). These equivalent times are much shorter than those where the geopotential field is restored. Note that the variation of n in the OR2 method causes a variation in the smoothness of the error curves in Figures 5B and 5C.

Figures 6A, 6B, 6C and 6D compare the forecasts of the height at P. The upper curves show the forecast after initialization when the geopotential field is allowed to adjust while the lower curves show the same forecast for the case in which the geopotential is restored after each iteration. The scale is the same as that of Figure 4B. Note that in all cases, 150 iterations are not sufficient to completely damp the gravity waves generated by the geostrophic perturbation if the geopotential field is restored. The maximum residual amplitudes are 10m (NH1), 7m (NH2), 5m (OR1), and 4m (OR2). The OR methods, in particular, are quite

Figure 6A
Height forecast at (4, 4) after NH1 initialization

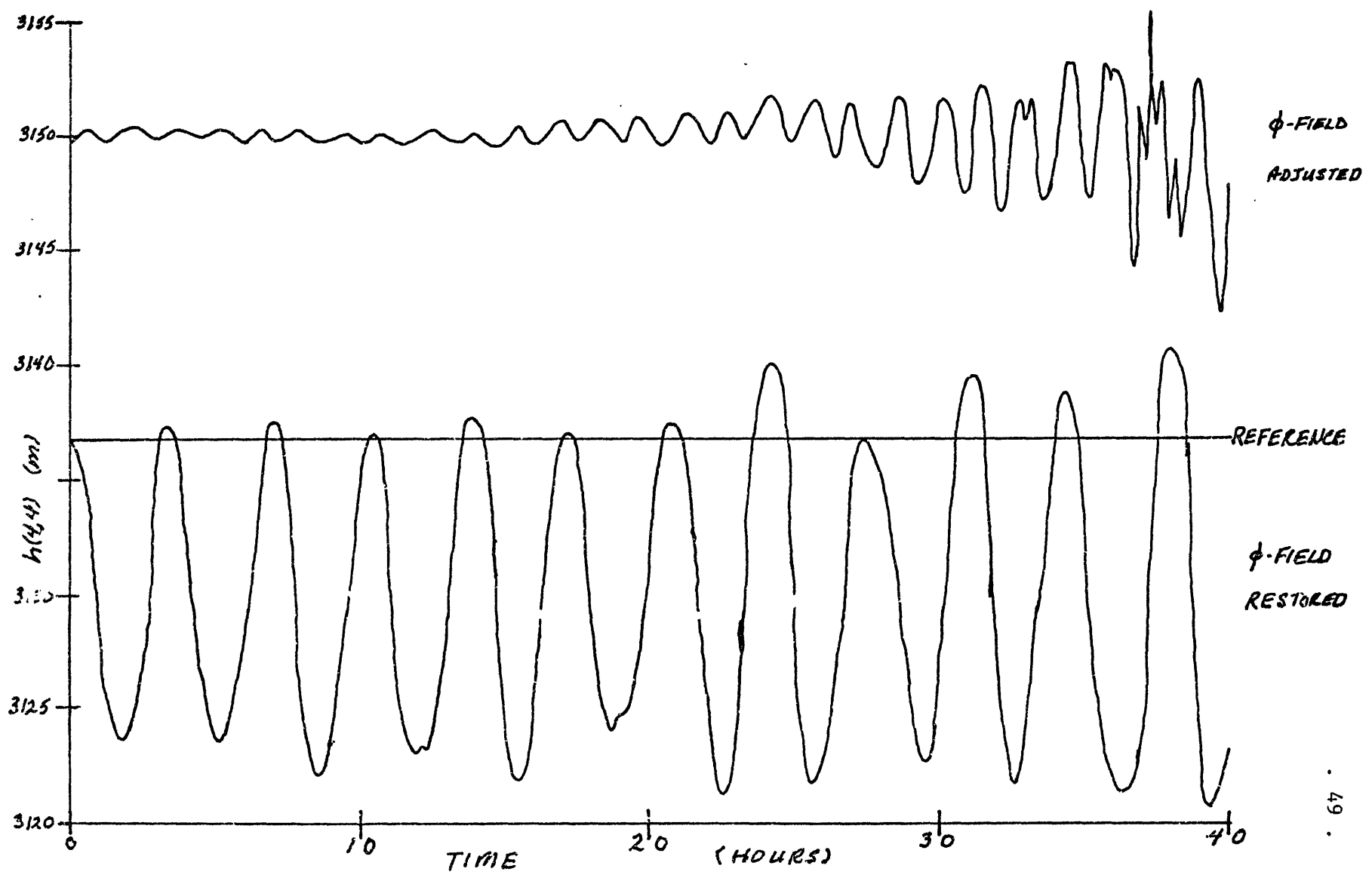


Figure 6B

Height forecast at (4, 4) after NH2 initialization

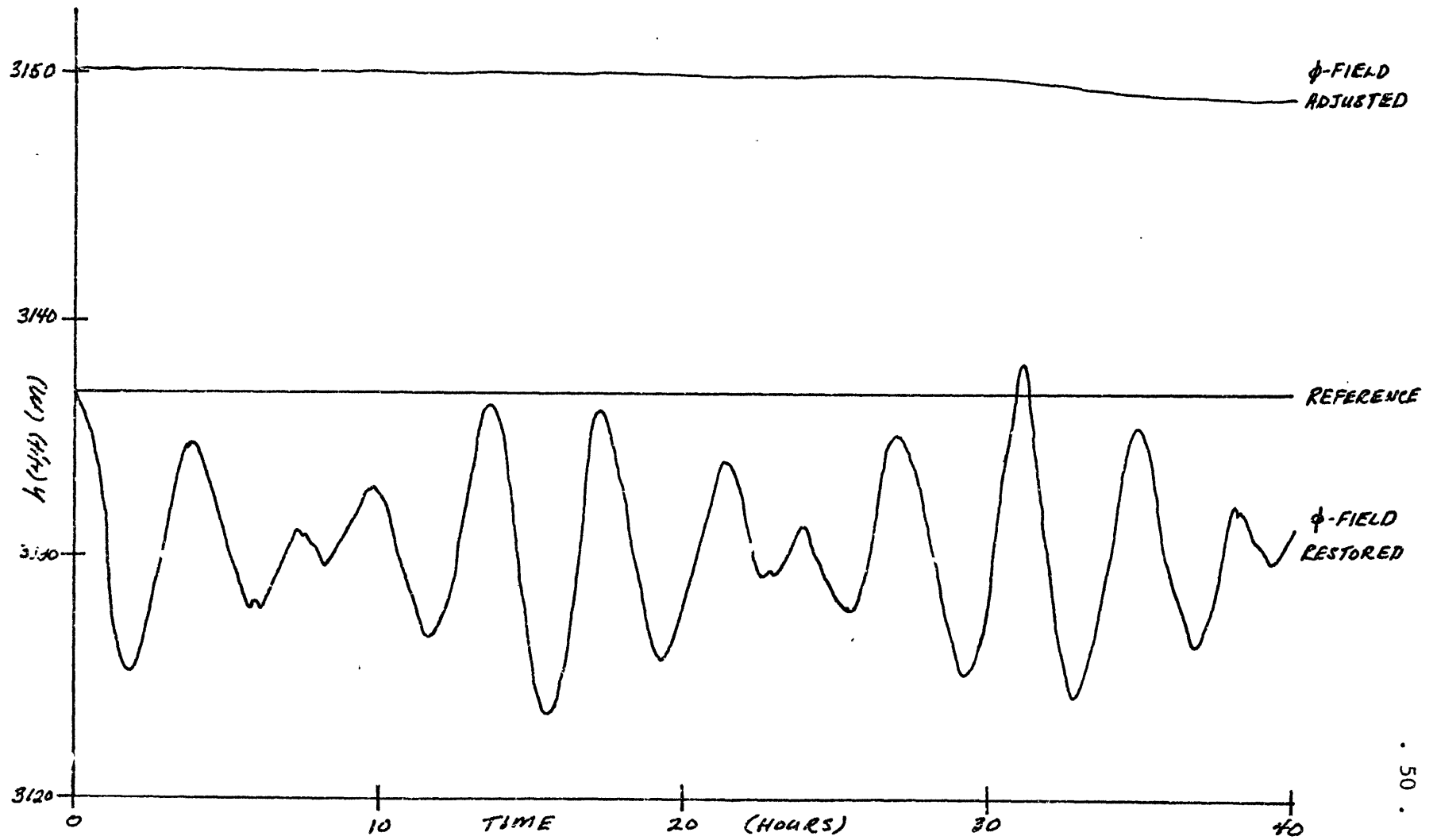


Figure 6C:

Height forecast after ORI initialization

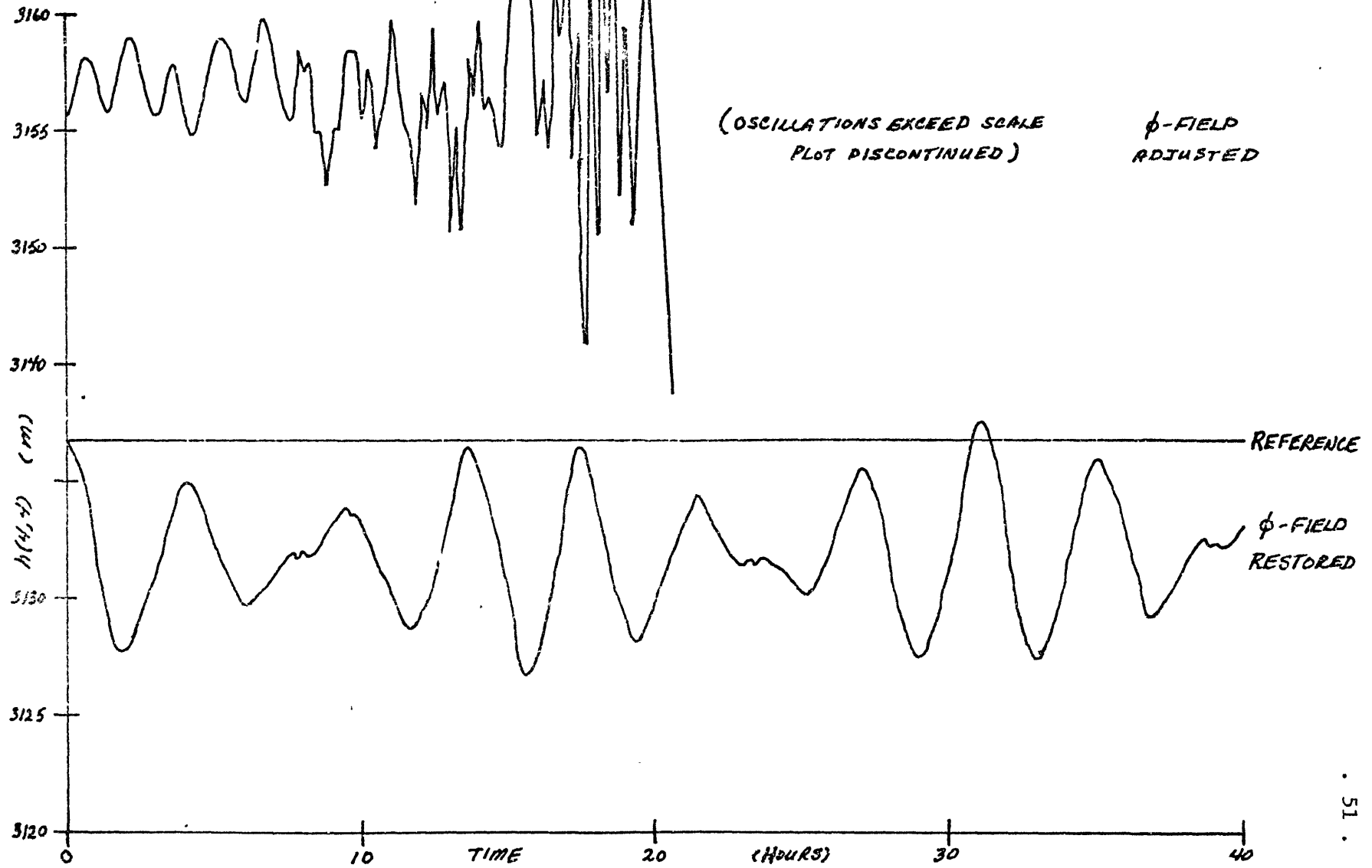
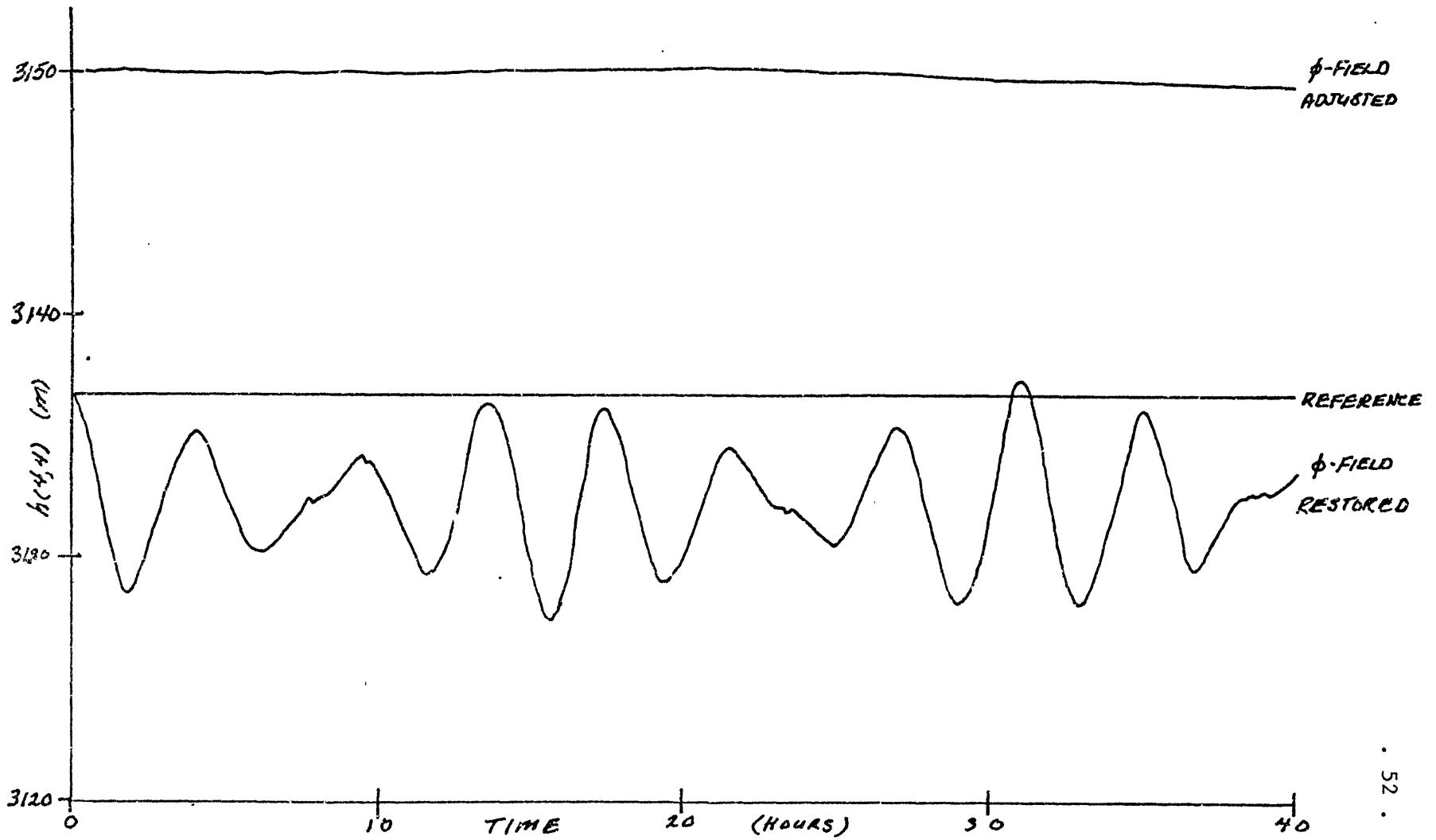


Figure 6D

Height forecast after OR2 initialization



effective in reducing the amplitudes but not as effective as the balance equation. It is also disturbing that we have not better recovered the balance by using the iterative techniques.

When the geopotential field is allowed to adjust a quite different forecast is also obtained from all the iterative initialization methods. The amplitude of the oscillation increases from about 0.2m to 10m after NH1 initialization and from 1.4 to 200m after OR1 initialization. The height variation after OR1 initialization exceeds our scale beyond 20 hours and the plot is discontinued. These excessive oscillations are probably due to aliasing of very high frequency oscillations which are not eliminated by the NH1 and OR1 methods. On the other hand, the NH2 and OR2 methods completely eliminate these oscillations from the forecast if the geopotential field is allowed to adjust. This attainment of balance more than compensates for the initial increase in the rms errors. The high frequency oscillations can also be eliminated with the NH1 and OR1 methods by decreasing Δt but this would also decrease their efficiency. According to Equations 3.12 the damping decreases as the square of the time increment. Therefore, if we halve the time increment, for example, to maximize the damping of the highest frequency waves, the number of iterations required to reach steady values of the rms errors is increased by a factor of about 4.

The rms height and velocity errors remain virtually constant during the 48-hour forecast from the initial balance recovered by the NH2 and OR2 schemes. The final height error is still 46m and the velocity error is about 6.7 m/sec for both methods. Since the balance is restored these rms errors now accurately reflect the net effect of these two methods.

It is now apparent that only by completely allowing the geopotential field to adjust are we able to reach a state of balance in a reasonable number of iterations. It should be emphasized that the OR2 method is about 3 times more efficient than NH2 due to the fewer number of operations needed in each iteration. The rms height error increases by 7.5m (NH1), 18m (NH2), 21m (OR1), and 12m (OR2, $n = 1$) in the first iteration of each method. If we attempt to correct these errors by restoring the geopotential field we create a significant reversal of this adjustment of the height field. Therefore, by restoring the geopotential field, we also restore a significant portion of the original imbalance. This effect accounts for the slow convergence of the iterative methods when the geopotential field is restored.

The amount of error generated as both fields adjust is determined by the amount and nature of the imbalance which exists in the perturbed state. Due to the curvature and strength of the reference flow, the geostrophic approximation seems to have rather strongly perturbed the original balance. In making this approximation we have not utilized the fact that we do have measurements of the velocity field. We will now perturb both reference fields with random errors which simulate the effect of actual observational errors.

5.3 Initialization of a Randomly Perturbed State

The reference state is now perturbed by adding a normally distributed random error field of zero mean to the geopotential and velocity

fields. Three separate experiments are performed with different rms values of the height error. For Case I, no error is introduced in the height field. For Cases II and III this rms error has the values 5 and 10 meters, respectively. In all three cases, an rms error of 3 m/sec is added to each component of the velocity field. The forecasts from the perturbed state showed that the effect of this perturbation is similar to that shown in the upper curve of Figure 6C. The amplitude of the oscillations are about 250 meters (Case I), 350 meters (Case II), and 425 meters (Case III), after 48 hours. In all three cases the synoptic wave in the reference state is completely obliterated by these oscillations. The perturbing influence of the random errors is predominately concentrated in small scales, whereas the geostrophic perturbation predominantly affects larger scales of our system.

Only the NH2 and OR2 iterative methods are applied in these experiments and their behavior is very similar to that shown in Subsection 5.2. The geopotential field is allowed to adjust in all cases and 150 iterations are again performed. Both methods converge rapidly to nearly steady rms errors in the height and velocity fields although these errors slightly decrease as the iterations proceed. The small scale gravity waves are again eliminated by both methods, although a very large scale oscillation remained with an amplitude of a few meters and a period which exceeds the 48 hour forecast interval. These slight departures from the previous experiment are believed to occur because the samples of random numbers selected slightly alter the character of our synoptic wave, causing it to oscillate very slowly. These effects do not represent a lack of

balance or a significant alteration of the convergence properties of these iterative methods.

Since a good balance is again achieved by both techniques, we are most interested in the size of the rms departures from the reference state after initialization. These errors are listed in Table 3 for all three cases. Two effects are noticeable here. First the adjusted errors are

Table 3

RMS error in height perturbation	RMS error in velocity perturbation	Method	Adjusted rms height error	Adjusted rms velocity error
Case I				
0 m	4.2m/sec	NH2	6.4m	2.0 m/sec
Case II		OR2	6.2	2.0
5	4.2	NH2	6.5	1.9
Case III		OR2	6.3	1.8
10m	4.2	NH2	6.6	1.9
		OR2	6.5	1.8

smaller than those of the initialization of the geostrophic perturbation, namely, 46m and 6.8 m/sec. The second is that the adjusted height errors seem relatively insensitive to the variation of the height perturbation. Both effects are again a result of allowing the geopotential field to reach balance with the velocity field as determined by the adjustment process. The adjustment process distributes the energy of impulsive imbalances between gravity wave energy and the total energy of the synoptic flow. The relative distribution of this energy depends on the scale of the imbalance. The actual dependence is determined by the ratio of this

scale and the Rossby radius of deformation, which, in turn, depends on f (Rossby, 1937-38b; Bolin, 1953; Charney, 1973). But f is constant in our system. The energy of small scale imbalances, as produced by random height errors, is converted more into gravity wave energy which is then dissipated by the iterative methods. Thus the residual error of the synoptic wave is less.

The current tendency in primitive-equation modeling, in order to utilize advances in computer technology as well as growing resolution of atmospheric observations, is to decrease the size of the space increments. This results in an increased limitation on the size of the time increment necessary for stability of both marching procedures and our iterative techniques, however, the damping characteristics of the iterative methods change relatively to correctly adapt to the decreased time increment. Even if the error in measurements does not significantly decrease, the scale of initial random imbalances in observed data will decrease. The iterative procedures should respond favorably to this trend to decrease the scale of imbalances.

In contrast, the ellipticity constraint on the balance equation becomes excessively stringent as the spatial resolution is decreased. For example, we rewrite the ellipticity condition, 2.8, in the following form:

$$\Delta h \equiv h - \bar{h} < \frac{f^2 \Delta s^2}{8g}, \quad 5.3$$

where \bar{h} is again the average height of the four grid points adjacent to a central point whose height is h . In this form the ellipticity condition is a restriction on the difference, Δh , for a given Δs . Note that the

restriction depends on the square of Δs . If we halve the space resolution of a model, the ellipticity constraint is 4 times as restrictive. In our system $\Delta s = 250$ km and $f = 10^{-4}$ sec⁻¹. The difference, $\Delta h = 8$ meters, at a grid point is sufficient to violate the ellipticity constraint. For a space increment of 125 km, the ellipticity condition is violated at a point where $\Delta h = 2$ meters! As was noted in the geostrophic perturbation case, this violation of the ellipticity constraint occurs at a large number of points for stronger amplitude reference states. Experiments with $\Delta s = 125$ km also confirmed that the geopotential field could not be corrected by the procedure in 5.2, even where random height errors of small magnitude were used to perturb the standard reference state.

Also with $\Delta s = 250$ km, the ellipticity condition is violated at a large number of points where the height field is perturbed by random errors of Case III. The procedure in 5.2 again failed to correct the geopotential field. For Case II the ellipticity condition was satisfied everywhere, by the procedure in 5.2, but, even so, the balance-equation-solution procedure did not converge. It is pertinent to remark here that the simultaneous adjustment of the geopotential and velocity fields is not possible in the framework of the balance equation approach.

In discussing random errors we have tacitly assumed that the decrease in spatial resolution is justified by the density of observations. In data sparse regions it is often easier to obtain accurate height fields from a few data points than to obtain accurate velocity fields. The geostrophic approximation is often used. In the next experiment we will

briefly describe a simple method for correcting the geostrophic winds by utilizing an approximation to the gradient wind equation.

5.4 Gradient Wind Correction of the Geostrophic Velocity Fields

The gradient wind equation may be expressed as follows:

$$V - V_g = -\frac{V^2}{fr} \quad 5.4a$$

where V is the gradient wind, V_g is the geostrophic wind, and r is the radius of curvature of the streamlines. This equation is strictly valid only for circular steady flow. The solution for V better approximates the curvature effects of the flow than does the geostrophic wind. However, the equation is quadratic in V and is subject to a restriction, similar to the ellipticity condition on the balance equation, so that the solution is real:

$$V_g + \frac{fr}{4} > 0 \quad 5.4b$$

If we assume that the right hand side of Equation 5.4a is small compared to V_g , we may write $V = V_g(1+\epsilon)$ and neglect terms of the order of ϵ^2 . We obtain then an approximation for ϵ :

$$\epsilon = -\frac{V_g}{fr + 2V_g} \quad 5.4c$$

The evaluation of ϵ at each point of the field involves the estimation of the radius of curvature r :

$$r = - \frac{(1 + y'^2)^{3/2}}{y''} \quad 5.4d$$

where $y' = \frac{dy}{dx}$ is the slope of the streamline, $\psi(x, y) = \text{constant}$:

$$y' = - \frac{\psi_x}{\psi_y} \quad 5.4e$$

and

$$y'' = - \frac{\psi_x^2 \psi_{yy} - 2 \psi_x \psi_y \psi_{xy} + \psi_y^2 \psi_{xx}}{\psi_y^3} \quad 5.4f$$

Combining Equations 5.4d-f, we obtain an expression for r in terms of the stream function:

$$r = - \frac{(\psi_x^2 + \psi_y^2)^{3/2}}{\psi_{xx} \psi_y^2 - 2 \psi_x \psi_y \psi_{xy} + \psi_x^2 \psi_{yy}} \quad 5.4g$$

Since r is only used to evaluate ϵ , we may approximate ψ by the geostrophic stream function. With this approximation and the resulting computation of ϵ over the entire field we may determine γ everywhere. In the regions where the constraint 5.4b is not satisfied, i.e., where r is large and negative, ϵ is no longer a small number. In these regions no correction of the geostrophic wind is made since the curvature is small.

This correction procedure was applied to the geostrophically perturbed state before initialization. The resultant rms error in the

corrected velocity field is 3.8 m/sec as compared to 7.7 m/sec for the purely geostrophic error. A forecast directly from the corrected state showed gravity waves with a maximum amplitude of 12 meters as compared to 125 meters in the forecast from the uncorrected state.

The OR2 method was applied for 150 iterations with the geopotential field allowed to adjust. Rapid convergence occurred as in previous experiments. The forecast after initialization also showed that gravity waves had been eliminated. The departure of the adjusted rms errors are compared to those after initialization of the uncorrected state in Table 4.

Table 4

Perturbation	Adjusted rms height error	Adjusted rms velocity error
Geostrophic	46 m	6.7 m/sec
Gradient	5.5 m	2.9 m/sec

This simple correction does seem to significantly decrease the departure of the adjusted fields from the reference state. However, these very encouraging results may be enhanced by the circular symmetry of our flow system (refer to Figure 3A) and its steady-state character.

6. Summary and Conclusions

The balance equation approach, as represented in this study, provides adequate, but not complete, balance in the initial state for a primitive equation forecast. However this approach depends very critically on the ellipticity condition, which, in its simplest form, is a restriction on the maximum amount by which a given height can differ from the average height of the neighboring field. We have reported that this restriction is severely violated around strong anticyclones and when the resolution of a model is increased. More importantly the ellipticity condition is violated by measurement errors typical of those occurring in atmospheric observations, even for moderate circulations modeled at moderate resolution. The failure of the "correction" procedure in equations 5.2, to satisfy the ellipticity constraint in these cases merely emphasises the limited means available for solving the general form of the balance equation.

The iterative technique, on the other hand, is not restricted. The primary criticism of this technique has been the excessive amount of computation required to reach a state of balance. Two results of this study have indicated that this amount of computation can be substantially decreased. Firstly, the simpler Okamura-Rivas schemes reduce the amount of computation per iteration by a factor of 2 over the NH1 scheme and 3 over the NH2 scheme. The variation of n in the OR2 scheme further reduces the number of iterations required for convergence by virtue of its increased efficiency in damping at intermediate frequencies, but the more important effect of this flexibility is the increased damping of the high

frequency waves. It is this latter effect which increases the stability of the initial balance with respect to the primitive forecast for which the iterative technique is designed. In our experiments, the OR2 scheme is more than an order of magnitude faster than the balance equation approach.

The second result of this study is the reduction of the number of iterations required for convergence when the initial fields are allowed to adjust. The number of computations can be decreased by a factor of more than 10 over the method in which the mass is restored. The slow convergence of the restorative-iterative method results from the partial restoration of the original imbalance at each iteration. The more important effect of allowing the mass field to adjust is that the state of balance attained fully realizes the compatibility of using the primitive equations directly in the iterative process. These features far outweigh the error in the mass field incurred as a result of its adjustment. Nevertheless this error is a real problem.

Winninghoff (1973) has suggested a scheme for partial restoration of the mass field to reduce this error, but this again compromises the practical efficiency of the technique as well as its compatibility with the primitive forecast. Our experiments, in which random errors of 10m and 4.2 m/sec (rms) were introduced in the height and velocity fields, respectively show that the errors, after using the iterative technique for adjustment, are only about 6.5m and 2.9 m/sec. These errors are typical of real atmospheric measurements. We conclude that, where the density of observations is sufficiently high to justify the assumption

that measurement errors are random, the observed fields of both velocity and mass should be used to reduce the error incurred in allowing both fields to adjust during the iterative procedure. The smoothing incorporated in conventional analysis techniques may also help to reduce the error in observed fields. In data-sparse regions the geostrophic approximation may be necessary. In this case the gradient wind correction described in Equations 5.4, may considerably reduce the initial wind error. In our experiment this correction of the initial wind field had the effect of reducing the adjusted rms height error from 46 to 5.5m. Another method to improve the initial estimation of the wind field could be to solve the balance equation on a coarse grid in data-sparse regions and then interpolated to finer resolution. In any case, we conclude that the best procedure is to first obtain a good estimate of the initial fields and then apply the iterative techniques, allowing the free adjustment of the mass field.

This study has been limited in some important areas. The effect of the iterative schemes on the model which includes the variation of the Coriolis parameter and terms accounting for the earth's curvature should be determined in further study. These features do not permit the existence of a non-trivial, steady-state flow. The rapid convergence shown here when the initial fields are allowed to adjust may help to decrease the damping of slow frequency synoptic waves, since fewer iterations are required. In particular, the flexibility of the variable parameterization in the Okamura-Rivas scheme may aid in this respect. We have not fully exploited this capability in this study. The ultimate

test will naturally be the use of the iterative technique in operational forecasts using sophisticated models and real data. But it is hoped that sufficient evidence is given here to indicate that, when the adjustment process is allowed to freely act, the iterative technique, and particularly the Okamura-Rivas scheme, can provide a viable alternative to the balance equation approach to initialization.

REFERENCES

- Bolin, B., 1953: The adjustment of a non-balanced velocity field towards geostrophic equilibrium in a stratified fluid. Tellus 5, 373-385.
- Bolin, B., 1955: Numerical forecasting with the barotropic model. Tellus 7, 27-49.
- Bryan, K., 1966: A scheme for numerical integration of the equations of motion on a irregular grid free from nonlinear instability. Monthly Weather Review 95, 39-40.
- Cahn, A., 1945: An investigation of the free oscillations of a simple current system. Journal of Meteorology 2, 113-119.
- Charney, J.G., 1955: The use of the primitive equations of motion in numerical prediction. Tellus 7, 22-26.
- Charney, J.G., 1973: Planetary fluid dynamics. Dynamic Meteorology, P. Morel editor, D. Reidel Publishing Company, Dordrecht, Holland, 97-351.
- Haltiner, G.J., 1971: Numerical Weather Prediction. John Wiley and Sons, Inc., New York. 317 pages.
- Hinkelmann, K., 1961: Non-characteristic filtering of meteorological noise waves. Deutscher Wetterdienst, West Germany. 19 pages. (unpublished manuscript).
- Houghton, D. and W. Washington, 1969: On global initialization of the primitive equations: Part I. Journal of Applied Meteorology 8, 726-737.
- Krishnamurti, T.N. and D. Baumhefner, 1966: Structure of a tropical disturbance based on solutions of a multilevel baroclinic model. Journal of Applied Meteorology 5, 396-406.
- Lilly, D.K., 1965: On the computational stability of numerical solutions of time-dependent nonlinear geophysical fluid dynamics problems. Monthly Weather Review 93, 11-26.
- Miyakoda, K., 1956: On a method of solving the balance equation. Journal of the Meteorological Society of Japan 24, 364-367.
- Miyakoda, K., 1960: Numerical solution of the balance equation. Technical Report of the Japan Meteorological Agency, No. 3, 15-34.

- Miyakoda, K. and R.W. Moyer, 1968: A method of initialization for dynamical weather forecasting. Tellus 20, 115-128.
- Nitta, T., 1969: Initialization and analysis for the primitive equation model. Proceedings of the International WMO/IUGG Symposium on Numerical Weather Prediction in Tokyo, 1968, the Meteorological Society of Japan, pp. VI-11 to VI-20.
- Nitta, T. and J.B. Hovermale, 1969: A technique of objective analysis and initialization for the primitive forecast equations. Monthly Weather Review 97, 652-658.
- Petterssen, S., 1953: On the relation between vorticity, deformation and divergence and the configuration of the pressure field. Tellus 5, 231-238.
- Phillips, N.A., 1960: On the problem of initial data for the primitive equations. Tellus 12, 121-126.
- Phillips, N.A., 1963: Geostrophic motion. Reviews of Geophysics 1, 123-176.
- Richardson, L.F., 1922: Weather Prediction by Numerical Process. Cambridge University Press, London, 236 pages.
- Rivas, E.K. de, 1971: The Circulation of the Atmosphere of Venus. Ph.D. Thesis, Department of Meteorology, Massachusetts Institute of Technology, Cambridge, 271 pages.
- Rossby, C.-G., 1937-1938a: On the mutual adjustment of pressure and velocity distributions in certain simple current systems, I. Journal of Marine Research 1, 15-28.
- Rossby, C.-G., 1937-1938b: On the mutual adjustment of pressure and velocity distributions in certain simple current systems, II. Journal of Marine Research 1, 239-263.
- Shuman, F.G., 1957: Numerical methods in weather prediction: I The balance equation. Monthly Weather Review 85, 329-322.
- Winninghoff, F.J., 1973: Note on a simple restorative-iterative procedure for initialization of a global forecast model. Monthly Weather Review 101, 79-84.