VARIATIONAL IMPLICIT NORMAL MODE INITIALIZATION FOR NUMERICAL WEATHER PREDICTION MODELS

by

Luc Fillion

A Thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Department of Meteorology

McGill University, Montréal

May 1991

ABSTRACT

Using the technique of *Implicit nonlinear Normal Mode Initialization* (INMI), we consider the extension of this method in a variational framework for initializing NWP models. The fact that INMI works in physical space instead of the normal mode space, offers the possibility to efficiently reconsider Daley's (1978) original proposal.

The first part of this study gives the formulation of a *variational* INMI scheme (VINMI) for shallow-water models using Daley's approximation. Applications of VINMI to a regional finite elements shallow-water model with a fairly large number of degrees of freedom (~ 30,000) is first considered. It is shown that the VINMI scheme has *the same balancing benefits* as INMI but also efficiently control the changes that are made to the analyzed fields in a way consistent with their estimated error variance. Results of a two day forecast also indicate that VINMI has the possibility to *significantly* affect the slowly evolving component of the flow.

The second part of this research extends the VINMI scheme to a global spectral shallow-water model with triangular truncation at zonal wavenumber 63. Similar conclusions as the regional model are obtained in the present context (including a reasonable computer cost), thus extending the range of applicability of the method to those models which are more commonly used in NWP. Simple connections between INMI and quasi-geostrophic theory are also used to asses the convergence problem of the method in the context of height-constrained initialization.

The third aspect of the study shows that the VINMI procedure can be generalized in a consistent way to multilevel NWP models. Several tests indicate the efficiency and robustness of the scheme in an operational context.

Finally, a general outlook of possible future developments of the variational scheme is presented at the end of the thesis.

RÉSUMÉ

On considère l'extension dans un cadre *variationnel* de la méthode d'initialisation *implicite* (ou intrinsèque) non linéaire par modes normaux (INMI). La formulation dans l'espace physique de l'INMI rend l'idée originale de Daley (1978) plus intéressante à considérer dans un cadre opérationnel.

La première partie de ce travail offre une formulation *variationnelle* de l'INMI (VINMI) basée sur l'approximation de Daley. La méthode est ensuite appliquée à un modèle aux éléments finis pour les équations de Saint-Venant ayant un assez grand nombre de degrés de liberté (~ 30,000). Il est démontré que la VINMI peut fournir un équillibre dynamique comparable à celui obtenu par l'INMI, mais permet en plus de controler les changements faits aux champs d'analyse. Un résultat important dans ce contexte est que la VINMI a un impact non négligeable sur les champs prévus après 48 heures d'intégration.

En seconde partie, on a étendu la VINMI à un modèle spectral global des équations de Saint-Venant avec une troncature spatiale triangulaire au nombre d'onde zonal 63. On y a observé le même type de résultats que pour la version à domaine limité (y inclus un coût informatique abordable) donnant ainsi un champ d'application plus vaste de la méthode, étant donné l'utilisation assez répandue de la méthode spectrale dans les centres de recherche en prévisions numériques. Le problème de convergence de la VINMI dans un contexte de contrôle du géopotentiel uniquement fut ensuite examiné a partir de liens avec la théorie quasi-géostrophique.

La VINMI est finalement considérée dans un contexte barocline. On a établi une formulation généralisée de la VINMI qui opère complètement dans l'espace physique mais qui conserve l'aspect important de ¨ restriction verticale ¨ du schéma de Machenhauer (1977). La méthode fut testée dans le contexte d'un modèle de prévision opérationnel Canadien.

ACKNOWLEDGEMENTS

The author would like to thank Dr. Clive Temperton for initiating and supporting the present work while he was at the " Division de recherche en prévision numérique " (DRPN) at Dorval. His continuous and kind guidance during the course of this research are greatly acknowledged.

I am also grateful to Drs. Tom Warn (McGill University) and André Robert (Université du Québec à Montréal) who acted as thesis director and advisor respectively. The author also greatly benefited from unavoidable daily interactions with Dr. Andrew Staniforth at DRPN.

Thanks are also due to Drs. Harold Ritchie and Jean Côté at DRPN for providing the computer code of their spectral shallow-water models used in chapter IV of this thesis.

I am also grateful to the research staff at DRPN for their interest during the regular seminars, and to the "division informatique du RPN" for their expert programming support.

I wish to thank also Drs. Jacques Derome (McGill university) and Michel Béland (chef du RPN) who made this type of research feasible.

Finally, I acknowledge the remarkable assistance of my colleagues Pierre Gauthier and Sylvie Gravel at RPN during the redaction of this thesis.

TABLE OF CONTENTS

List of Figures

List of Tables

STATEMENT OF ORIGINALITY

The original findings of this thesis are :

1) The formulation and application of an efficient variational implicit nonlinear normal mode initialization method for

   a. a regional finite-element shallow-water model

   b. a global spectral shallow-water model.

2)

   a. The formulation of a variational implicit nonlinear normal mode initialization method for a multilevel model.

   b. Application of the VINMI-3D to an operational regional baroclinic model for numerical weather prediction.

The VINMI method as formulated in the present study allows for the first time in the field of nonlinear normal mode initialization for numerical weather prediction models, to balance the analyzed fields in a consistent manner with their known statistical error distributions in three dimensional space.

MANUSCRIPTS AND AUTHORSHIP

The candidate has the option, subject to the approval of the Department, of including as part of the thesis the text, or duplicated published text (see below), of an original paper or papers. In this case the thesis must still conform to all other requirements explained in Guidelines Concerning Thesis Preparation. Additional material (procedural and design data as well as descriptions of equipment) must be provided in sufficient detail (e.g. in appendices) to allow a clear and precise judgement to be made of the importance and originality of the research reported. The thesis should be more than a mere collection of manuscripts published or to be published. It must include a general abstract, a full introduction and literature review and a final overall conclusion. Connecting texts which provide logical bridges between different manuscripts are usually desirable in the interests of cohesion.

It is acceptable for theses to include as chapters authentic copies of papers already published, provided these are duplicated clearly on regulation thesis stationery and bound as an integral part of the thesis. Photographs or other materials which do not duplicate well must be included in their original form. In such instances, connecting texts are mandatory and supplementary explanatory material is almost always necessary.

The inclusion of manuscripts co-authored by the candidate and others is acceptable but the candidate is required to make an explicit statement on who contributed to such work and to what extent, and supervisors must attest to the accuracy of the claims, e.g. before the Oral Committee. Since the task of the Examiners is made more difficult in these cases, it is in candidate's interest to make the responsabilities of authors perfectly clear. Candidates following this option must inform the department before it submits the thesis for review.

The text of the above shall be cited in full in the introductory sections of any theses to which it applies.

Chapter I

Introduction

I.1 The initialization problem

One of the earliest problems encountered in the development of numerical weather prediction (NWP) with the primitive equation models, concerned the generation of spurious fast oscillations during the forecast integration. Since that time, it has been recognized that the problem of properly setting the initial data for such models requires special treatment. In fact when directly inserting initial conditions for mass and wind fields obtained from an analysis scheme, it quickly becomes apparent that high frequency oscillations with unacceptably large amplitudes are generated during the integration of the model. Figure 1.1 for example, illustrates this behaviour for a 48 hour forecast using the Canadian regional finite-elements model (RFE) operational at CMC (Centre Météorologique Canadien) starting with analyzed fields. Fast oscillations in the surface pressure field are present with an amplitude of 2 mb, a feature which is rare in the routinely observed surface pressure data at weather reporting stations throughout the world. Fast oscillations of the gravity wave type are known to exist in the real atmosphere but their associated energy represents a small part of the total motion which evolves on much slower timescales.

Depending on the diffusion properties of the model (i.e. numerical and explicit diffusion), these oscillations will persist typically for the first 12-24 hours of a forecast unless the initial mass and wind fields are properly adjusted. This initial adjustment is referred to as *initialization*. Since most current operational analysis schemes do not produce "balanced" (to be defined later) initial data, these unrealistic large amplitude oscillations can seriously damage the quality of the forecasts during the early stages of integration. A classical disastrous aspect of these fast oscillations, is that they have associated with them a considerable amount of divergence and vertical motion which can completely obscure the

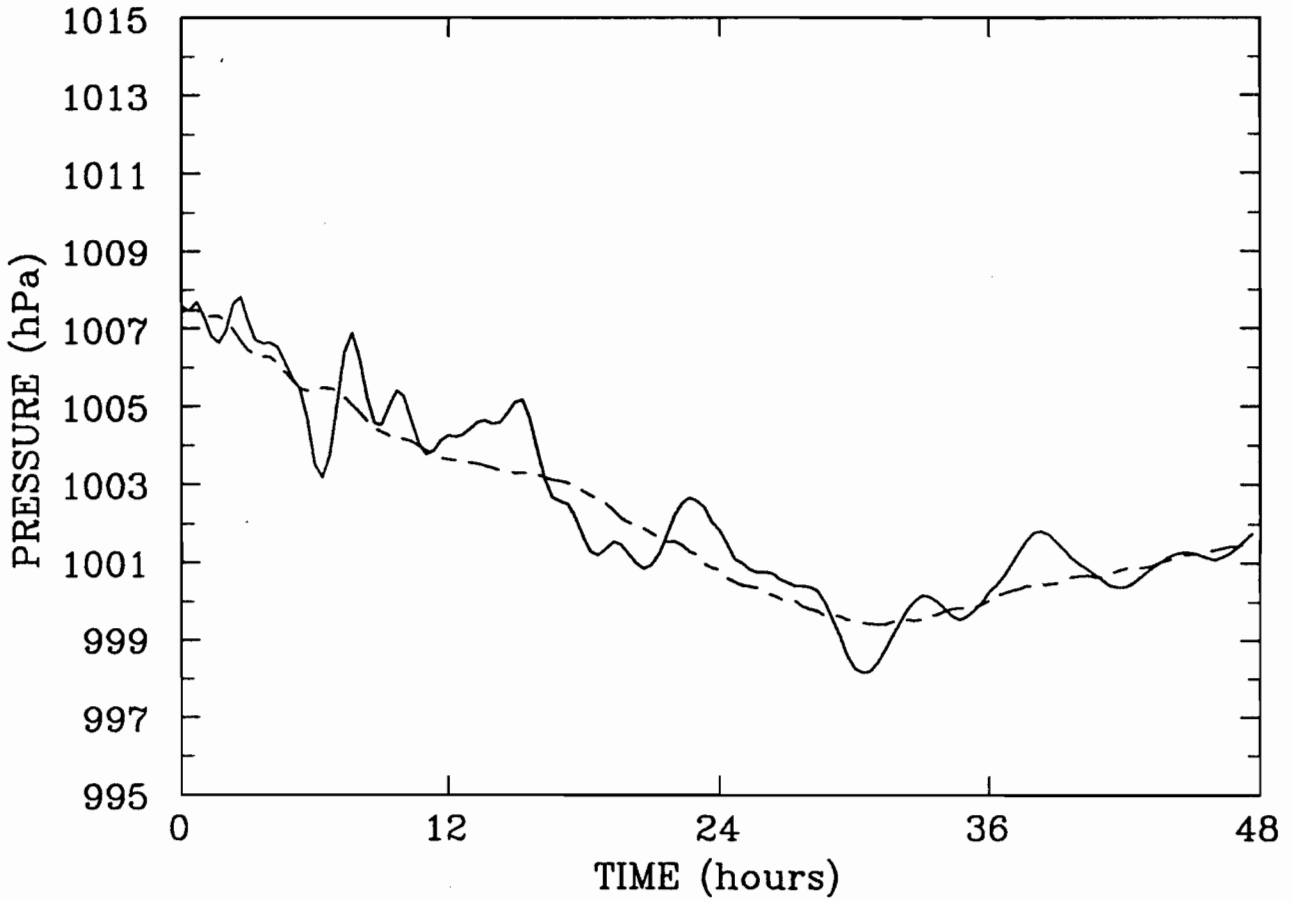


Fig.1.1 Time-trace of surface pressure for a point in the Northern-Atlantic as forecast by the operational Canadian RFE model. Solid line: no initialization. Dashed line: after three iterations of Implicit Nonlinear NMI (*after Temperton and Roch 1991*).

slowly evolving meteorologically significant cloud and precipitation patterns.

I.2 The initialization methods

The problem of " initial imbalance " may be attacked by imposing diagnostic balancing relationships on the initial data in order to mimic the approximate type of "balance" observed in long term integrations of the model. One of the most widely used operational analysis scheme at the present time, " statistical optimal interpolation (OI) " (Gandin 1963), uses geostrophic balance in a multivariate analysis mode (Rutherford 1972). Ignoring some other well known deficiencies of the OI schemes, geostrophy is easily shown to be a too crude approximation. Williamson et al. (1981) examined the accuracy of OI analyses when geostrophy is used to determine covariances between analyzed variables. Their study also clearly demonstrated an improved behaviour of the slow and fast (transient inertia-gravity) components (based on Tribbia's 1981 axisymmetric vortex) when a more accurate nonlinear relationship is used. The earliest attempt to force a nonlinear balance between dynamical variables of the model was based on the **nonlinear balance equation** (Charney 1955) obtained by setting $D = \delta_t D = 0$ in the divergence equation (where D is the horizontal divergence of the wind field) thereby giving a relation between the rotational part of the wind field and the geopotential (see also section 2.4). The nonlinear balance equation gives no information on the initial divergent wind field and Phillips (1960) pointed out that initial fields with vanishing divergence lead to an imbalance with subsequent high frequency oscillations. He proposed that the nonlinear balance equation be supplemented by a form of the classical **omega equation** to obtain a divergent initial wind field. Further discussions concerning the use of the balance equation are given in chapter II and IV.

During the period 1968 to 1977, the **dynamic initialization** method was introduced. This consists of integrating the model forward and backward around the initial

time to allow the model to adjust before starting a forecast. The procedure used a special time integration scheme that selectively damps the amplitude of certain high frequency modes. The earliest contributions are those of Miyakoda and Moyer (1968) (using a limited area two layer baroclinic model with Euler-backward time integration scheme), Temperton (1973) (with the shallow-water equations), and Temperton (1976) (similar to Nitta and Hovermale 1969, on a five level hemispheric model). One disadvantage of this method relates to the lack of control on the magnitude of the changes to the slow component of the analyzed fields during this cycling process. Also, dynamic initialization would seem to eliminate fast modes by introducing an artificial numerical diffusion. One consequence could be that it leads to a *false* balance, i.e., an effective dissipative term is added to the right-hand side of the fast mode equations which changes the nature of the balance. Consequently, it is not clear that this method can be justified in the context of initialization. The method does have the advantage that it is easy to apply (even for the non-hydrostatic primitive equations, Tanguay et al. 1990).

Meanwhile, the need to improve our understanding of the type of "balance" developped by primitive equation models for weather forecasting forced the development of more sophisticated initialization methods. This lead naturally to the **normal mode initialization (NMI)** method. By considering the discretized form of the model equations, linearized about a basic state atmosphere, the resulting linearized dynamics describes the evolution of a broad spectrum of free modes called *normal modes*. An essential characteristic of this spectrum is the appearence of two widely different groups of frequencies referred to as *slow* (Rossby) and *fast* (gravity-inertia) modes. Once the structure and frequencies of the normal modes has been obtained, the model equations may be rewritten in *normal mode space* (i.e. the linear space spanned by the normal modes). Balancing schemes devised to force the fast gravity modes to evolve slowly in time by properly specifying their amplitudes in normal mode space at the initial time are termed *normal mode initialization schemes*. A number of NMI techniques for the primitive

equations exist at this time and some of them will be considered more precisely in the next chapter.

As mentioned previously, it inevitably turns out that analyzed fields have some projection on the fast modes with amplitudes not reflecting the appropriate "balance" required by the model itself. Altough this is generally negligible for mean or long term forecasts (> 2 days), the proper initial setting of these modes is a fundamental problem when viewed from the point of view of data assimilation.

Dickinson and Williamson (1972) were the first to use the model's normal modes in the initialization problem. Although it required the construction of the normal modes of the numerical model, the approach was attractive in that it dealt *directly* with the spectrum of frequencies and structure of the fast modes. This aspect was attractive compared to quasi-geostrophic theory since it improved the compatibility of the balancing schemes and the underlying model. The limited benefit of Dickinson and Williamson's experiments using linear NMI (i.e. eliminate the initial amplitude of the fast modes in the analyzed fields) was rapidly improved by considering the nonlinear nature of the dynamical system written in normal mode space. Machenhauer (1977) devised his nonlinear balancing scheme by observing (i.e by running a spectral shallow-water model) that the nonlinear forcing term for the fast modes varied "slowly" (i.e. on a much slower timescale than the amplitude of the fast modes). His scheme (as will be discussed in subsequent chapters) thus required the vanishing of the *time tendency* of the amplitudes of the fast modes.

Baer (1977) and "later" Baer and Tribbia (1977) used a Rossby number expansion of the scaled primitive equations in normal mode space to devise a "filtering procedure" of the fast oscillations. In operational practice, Machenhauer's scheme was preferred and introduced by Andersen (1977), Daley (1979), and Temperton and Williamson (1979) for global models. Ballish (1980) analyzed the application of Baer and Tribbia's initialization scheme for the NMC (National Meteorological Centre) global spectral model and found similar performance of the scheme as compared to Machenhauer's scheme. Until now,

numerical modellers have concentrated their efforts on the application of the nonlinear NMI to their models, discovering various limitations of Machenhauer's scheme regarding the number of vertical modes to be initialized, the proper treatment of diabatic effects, etc. These specific problems are well documented in Temperton and Williamson (1979) and Daley (1981) for example. There is also the possibility of examining the behaviour of the fast modes that we usually initialize when the model is " properly balanced " and try to assess the degree of accuracy of the preceding initialization methods. This type of pioneering work has been done by Errico (1984, 1988 a,b,c, 1989 a,b) for climate models and a mesoscale model both using the primitive equations. Further research of this type is of crucial importance if the nature of dynamical balance is to be *accurately* characterized.

Ballish (1980) was the first to show that the main benefits of nonlinear NMI (i.e. balancing the highest frequency gravity-inertia oscillations of the model) could be obtained by neglecting the " beta terms " in his linearized system used to determine the normal modes, a technique which he called *non-normal mode initialization.* This reasoning was based on the ideas of the **bounded derivative initialization (BDI)** method of Kreiss (1979, 1980). This latter initialization method was first used in the meteorological context by Browning et al. (1980). The BDI method has a firm mathematical sounding, and deals with the specification of initial data (including boundary conditions if necessary) for ordinary or partial differential (hyperbolic) systems having different time scales.

The practical use of nonlinear NMI in regional models received more attention at the beginning of the 1980's . The nonseparability of the linearized equations due to the appearence of the horizontally variable map scale factor and Coriolis parameter (for projected grids) render the determination of the normal modes of the model intractable. Brière (1982) and Bourke and McGregor (1983) reexamined Ballish's results and succeeded in applying the ideas of nonlinear NMI for their regional models using specific assumptions to eliminate the beta terms in their linearizations. Their results together with Ballish's results pointed out an interesting direction of research in the domain of

initialization, that is: "what is the least drastic approximation in the linearization of the primitive equations that permits the physical space formulation of nonlinear NMI ? ".

Temperton (1985) was the first to establish a general formulation for what he called **implicit nonlinear NMI** using a simple algebraic reformulation of the standard "explicit " nonlinear NMI procedure. It was later shown by Temperton (1988, 1989, 1991) and Juvanon du Vachat (1986, 1988) that implicit normal mode initialization is also feasible by including most of the beta-terms of the usual linearization considered in explicit NMI. The technique has been extended successfully to the operational Canadian RFE model (Temperton and Roch 1991) an to the operational global spectral model at the European Centre for Medium Range Weather Forecasts (ECMWF) (Temperton 1990). We finally note that the BDI method may give some answers to the previously stated question. The connections between the BDI and nonlinear NMI methods were examined in part by Kasahara (1982a). We note however that in order to characterize the complicated type of adjustment required for the various normal modes of the model, the explicit knowledge of the structure of certain of these modes is important. An example of this is given by the treatment of equatorially trapped modes characterizing the tropical circulations. On the other hand, the very high frequency gravity modes may simply be considered as noise and treated with a physical space initialization. Considering the future demands of the "next generation" models (i.e. global models with very high and variable resolution), the treatment of the latter group of modes by such a scheme will become highly desirable. As it stands presently, the application of the BDI method on the sphere results in two different balancing schemes due to the separate scaling of the dynamical equations in the tropics and extratropics. These schemes were formulated by Browning et al. (1980) whereas Kasahara (1982) extended these schemes for the multilevel case. On the sphere, some sort of combination of these two schemes would be required.

I.3 Constrained initialization

The preceding sections have discussed the problem of dynamic imbalance and various initialization methods to remedy this problem. We now indicate two issues relating to initialization schemes of the form M(R,G) = 0 (where R is a given vector of the amplitude of the slow modes at the initial time, and G is the vector of gravity modes to be specified by the nonlinear constraint M), i.e. nonuniqueness and accuracy. The question of nonuniqueness may be stated as "Even if we know the location of the manifold characterizing balanced states accurately, what do we change to ensure that M = 0 , G, R or both ? " . This nonuniqueness is linked with the accuracy problem since current balancing schemes for NWP models are only approximations of the type of balance developped by the model after a certain period of time. Moreover, the errors introduced by the inherent limitations of the analysis scheme when determining the Rossby mode amplitudes, also introduces inaccuracies which may result in large changes on the reliable data (e.g. surface pressure). Variational methods allow one to make unique corrections based on error statistics, thus *complementing* the balancing scheme. The pioneering work of Daley (1978) dealt precisely with this aspect of the analysis-initialization problem. His " variational nonlinear NMI " scheme forced the initialized state to be as "close" as possible to the analyzed fields by minimizing a functional of the weighted changes of the analyzed fields. The essential features of Daley's scheme may be illustrated by using Leith's (1980) diagram (see Fig. 1.2). In phase space, each model state may be represented by a unique point having both slow and fast mode components (R,G). This is also valid for the original analyzed fields which is represented in Fig. 1.2 by point O. Balanced model states are located on a manifold described by M. Starting from the analyzed state O, successive application of the initialization scheme (i.e. adjustment of the fast mode amplitudes G) and the minimization algorithm (i.e. alteration of the slow mode amplitudes R) leads to the "optimal" balanced state I. This process was originally formulated in normal mode space for a spectral shallow-water model and its operational use could not be envisaged due to

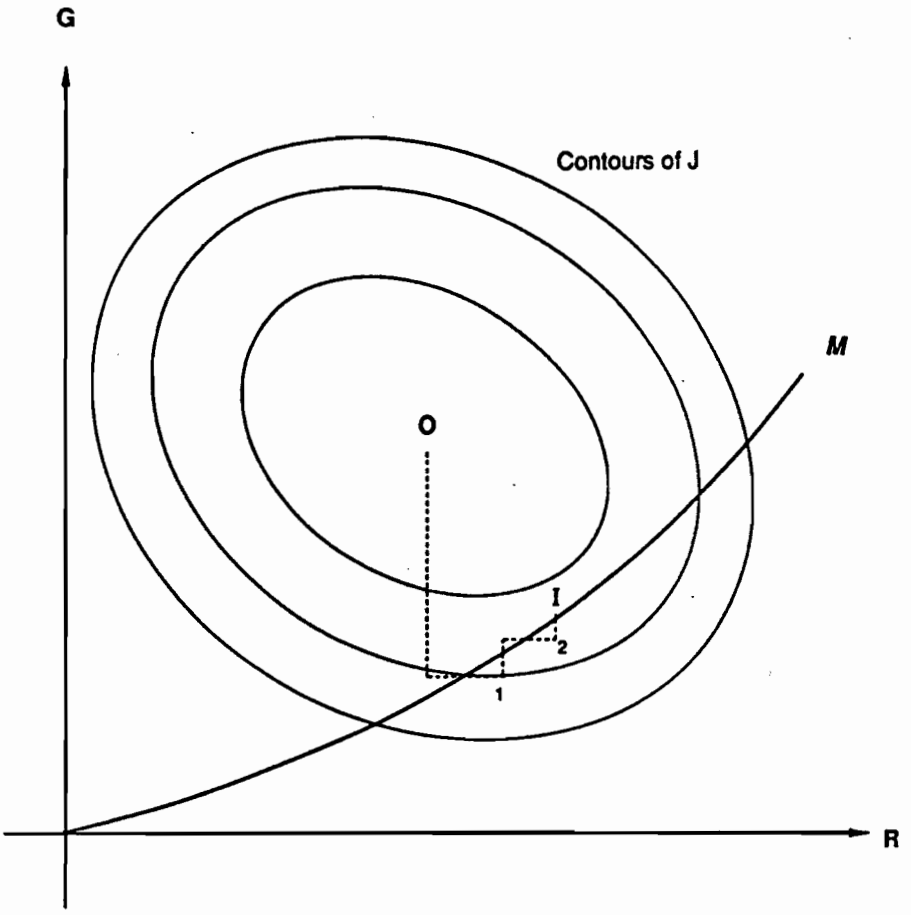


Figure 1.2. Variational initialization in phase space.

prohibitive computational cost. Some attempts to render the variational extension of nonlinear NMI schemes tractable were considered by Tribbia (1982) and Temperton (1984). Their schemes suffered from restrictions on the choice of weight functions used to limit the changes.

It is precisely the purpose of this thesis to reconsider Daley's variational proposal of nonlinear NMI using the implicit formulation. Chapter II includes a review of the principal initialization schemes applicable for NWP models and discusses briefly their relationships. In chapter III and IV the feasibility of variational INMI (VINMI) schemes for shallow-water models (regional and global) with present typical horizontal resolution of NWP models is demonstrated. The extension of the VINMI scheme to multilevel models is considered in chapter V. The method is then tested in the context of the presently operational regional finite-elements Canadian (RFE) model. Finally, chapter VI summarizes the essential findings of this research together with possible future developments of the method.

Chapter II

Initialization methods for primitive equation models and their relationship

The purpose of this chapter is to briefly review and clarify the theoretical and practical aspects of various initialization methods. The reader will find a more complete and detailed discussion of each of these methods in references cited in the text. For simplicity we restrict considerations to an f-plane shallow-water model without external forcing (i.e. orographic and thermal forcing). Further discussion of the relation between implicit normal mode initialization and quasi-geostrophic theory will be given in chapter IV. We begin by describing the basic shallow-water system.

II.1 The shallow-water model

The discussion to follow will be based on the shallow-water model which describes the hydrostatic motion of an homogeneous fluid with a free upper surface, although the concepts can be extended to more complicated baroclinic models. The shallow-water model may be termed the " kernel problem " since as will be shown in chapter V, initialization of baroclinic models can in a certain sense be reduced to the initialization of a set of shallow-water models differing only in their associated *equivalent depths*. The momentum and continuity equations are

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v}\cdot\nabla\mathbf{v} + \nabla\phi + f\mathbf{k}\times\mathbf{v} = 0 \qquad (2.1.1)$$

$$\frac{\partial \phi}{\partial t} + \mathbf{v}\cdot\nabla\phi + \phi\,\nabla\cdot\mathbf{v} = 0 \qquad (2.1.2)$$

where

$\phi = \Phi + \phi'$ where ϕ' is the geopotential deviation about Φ

Φ = mean geopotential depth

g = acceleration due to gravity

ρ = constant density of the fluid

μ = viscosity of the fluid : taken to be zero here (i.e. inviscid fluid)

Ω = angular speed of the earth

L = typical horizontal length scale characterizing shallow-water perturbations

x,y,z = east-west (north-south, vertical) spatial coordinates

$\mathbf{v} = (u,v)$ = zonal and meridional wind components.

In this chapter, we will concentrate on the dynamical equations themselves while ignoring the effects of lateral boundaries. Figure 2.1 shows the essential parameters characterizing the model.

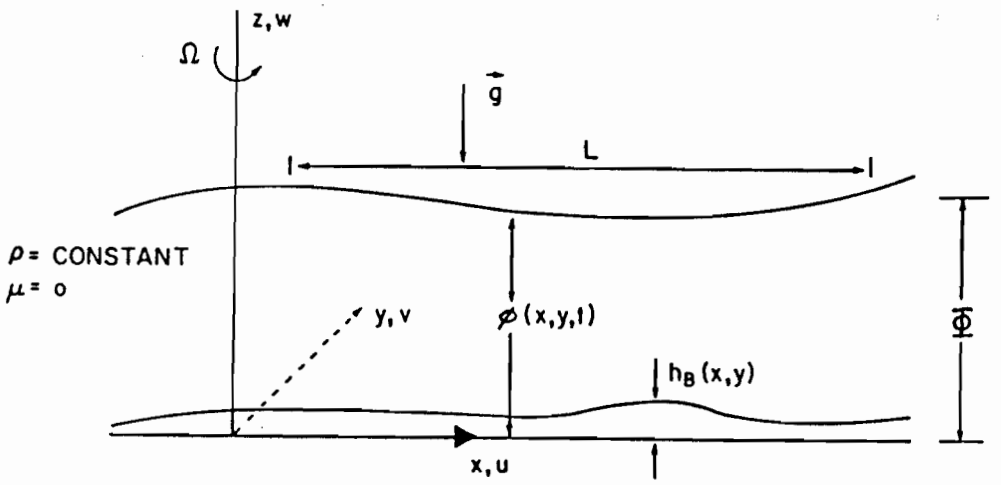


Figure 2-1. The shallow-water model.

Equations (2.1.1) and (2.1.2) may be written in the compact form

$$\mathbf{x}_t + \boldsymbol{L}\,\mathbf{x} + \boldsymbol{N}(\mathbf{x}) = 0 \qquad (2.1.3)$$

where

$$\boldsymbol{L} \equiv \begin{bmatrix} 0 & -f & \partial_x \\ f & 0 & \partial_y \\ \Phi\,\partial_x & \Phi\,\partial_y & 0 \end{bmatrix} \quad ; \quad \mathbf{x} \equiv \begin{bmatrix} u \\ v \\ \phi \end{bmatrix} \quad ; \qquad (2.1.4)$$

$$\boldsymbol{N} \equiv \begin{bmatrix} u\,u_x + v\,u_y \\ u\,v_x + v\,v_y \\ u\,\phi_x + v\,\phi_y + \phi\,(u_x + v_y) \end{bmatrix} \equiv \begin{bmatrix} N_u \\ N_v \\ N_\phi \end{bmatrix} \quad . \qquad (2.1.5)$$

II.2 The normal modes of the shallow-water equations

We now briefly discuss the nature of the linear normal modes of the shallow-water system. There are two classes; low-frequency quasi-geostrophic modes generally associated with the meteorologically significant part of the motion and high-frequency inertial-gravity modes which carry a small portion of the overall energy of the flow. This is most easily seen by considering small disturbance to a motionless background state $(u,v,\phi) = (0,0,\Phi)$. The motion is then described by the linear operator $\boldsymbol{L}$ to leading order. For periodic solutions characterized by a wavevector $\mathbf{k} = (k_x, k_y)$ and frequency σ of the form

$$\mathbf{x} = \mathbf{x}^o \exp\,(\,ik_x x + ik_y y - i\sigma t\,) \qquad (2.2.1)$$

where $\mathbf{x}^o$ represents the amplitude of the perturbation. Substituting (2.2.1) into the linearized form of (2.1.3) (i.e. where f and ϕ are replaced by f_o and Φ resp.), the eigenfrequencies are obtained by requiring that

$$\begin{vmatrix} -i\sigma & -f_o & ik_x \\ -f_o & -i\sigma & ik_y \\ i\Phi k_x & i\Phi k_y & -i\sigma \end{vmatrix} = 0 \quad . \qquad (2.2.2)$$

The eigenvalues σ are real for a given set of (k_x, k_y) and are given by:

$$\sigma = \sigma_R = 0 \quad ,$$

$$\sigma = \sigma_G = \pm \left[f_o^2 + \Phi(k_x^2 + k_y^2)\right]^{1/2} \quad .$$

Thus, on an f-plane, the slow (Rossby) modes are stationary while the fast (gravity-inertial) modes appear as a pair of modes traveling in the **k** and -**k** directions Further, the vorticity-divergence (ζ, D) form of the momentum equations linearized about a resting basic state are

$$\frac{\partial \zeta}{\partial t} + fD = 0 \quad ; \quad \frac{\partial D}{\partial t} - f\zeta = -\nabla^2 \phi' \quad ,$$

where

$$\zeta \equiv \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \quad ; \quad D \equiv \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \quad .$$

Using the previous results, it is clear that gravity-inertial waves possess both divergence and vorticity due to coupling by rotation. Rossby modes however, are divergence free due to the assumption of constant f. Further, the principle of conservation of potential vorticity

$$Q = \frac{\zeta + f}{\phi}$$

for the full nonlinear shallow-water equations reduces, in the present linear context, to conservation of the linearized potential vorticity q (normalized by Φ^2), i. e.

$$\frac{\partial q}{\partial t} = 0 \quad \text{where} \quad q \equiv \Phi\zeta - f\phi' \quad .$$

For plane waves, the previous result implies that gravity-inertial waves have zero linearized potential vorticity. The vanishing of the frequency of the slow modes implies that the evolution is determined by the nonlinear terms.

The dispersion diagram in this case is represented in figure 2.2a. When strong variations of the Coriolis parameter are considered, as in the case of global or equatorial models for example), the *frequency* separation is less clear-cut as some modes have intermediate frequencies. One then has to use arbitrary selection criteria (principally on the frequency of these modes) to define the eigenfrequencies as fast or slow. In multilevel

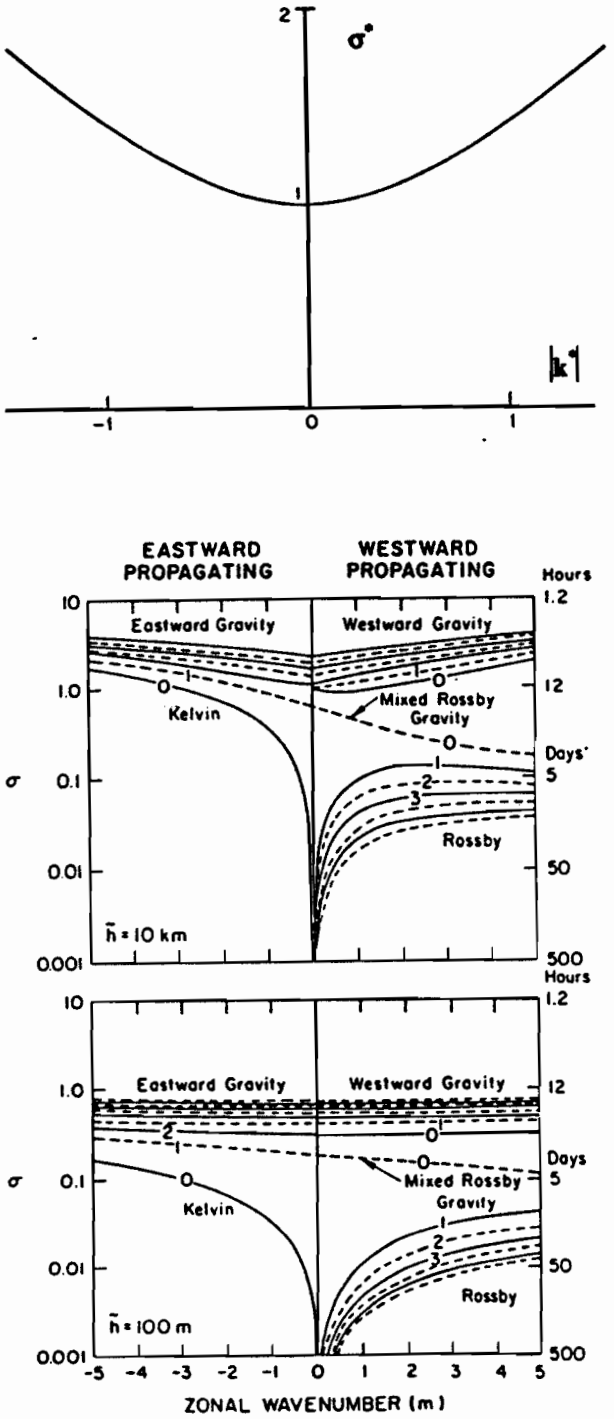


Fig. 2.2 Dispersion diagram for (a) f-plane shallow-water models in normalized frequency and wavenumber coordinates. The normalization factors are respectively f and $\Phi^{1/2}$, and the Rossby modes are stationary. (b) a global spectral shallow-water model with equivalent depths 10 km and 100 m. (*Daley 1986*)

models, the frequency separation also depends on the vertical structure or "equivalent depth" (see chapter V). As an example of this, Fig. 2.2b shows the dispersion diagram for two vertical modes of equivalent depth 10 km and 100 m. The lowest frequency "band" characterizes slow modes while the highest (eastward-westward) inertia-gravity frequency band characterizes modes that are initialized in nonlinear NMI schemes. Intermediate modes (Kelvin and mixed Rossby-Gravity) can be regarded as fast or slow depending on the wave-number domain considered. These last modes generally are much more difficult to initialize. The detailed structure of the normal modes of the linearized shallow-water equations on the sphere may be found in Longuet-Higgins (1968), Kasahara (1976), and more recently in Müller (1989).

II.3 The nondimensional form of the shallow-water equations

It will be usefull for later comparison of various initialization methods to deal with the nondimensionalized form of the model. Let us scale the variables as

$$x^* = \frac{x}{L} \quad ; \quad y^* = \frac{y}{L} \quad ; \quad t^* = \frac{t}{T} \quad ; \quad v^* = \frac{v}{V} \quad ; \tag{2.3.1}$$

where L, T, and V are typical values of the length, time, and velocity scales of the problem. The Coriolis parameter f will be set to a constant value of f_o. If we introduce an *advective* time scale

$$T = \frac{L}{V} \quad , \tag{2.3.2}$$

and a geostrophic scaling of the geopotential perturbation ϕ', i.e.

$$\phi' = \phi_s \phi^* \quad \text{where} \quad \phi_s \equiv f_o V L \quad ,$$

then

$$\phi = \Phi + \phi' = \Phi (1 + \varepsilon F_r \phi^*) \quad , \tag{2.3.3}$$

where

$$F_r \equiv \frac{f_o^2 L^2}{\Phi} = \text{Rotational Froude number} \quad ,$$

$$\varepsilon \equiv \frac{V}{f_o L} = \text{Rossby number} \quad .$$

If the horizontal length scale is fixed by requiring $F_r = 1$, i.e. if

$$L = L_R \equiv \frac{\Phi^{1/2}}{f_o} = \text{Rossby radius of deformation} \quad ,$$

$$\phi = \Phi (1 + \varepsilon \phi^*) \quad , \tag{2.3.4}$$

then, (2.1.1) and (2.1.2) become on dropping asterisks

$$\varepsilon \frac{dv}{dt} + \mathbf{k} \mathbf{x} \mathbf{v} = - \nabla \phi \quad , \tag{2.3.5}$$

$$\varepsilon \frac{d\phi}{dt} = - (1 + \varepsilon \phi) \nabla \cdot \mathbf{v} \quad , \tag{2.3.6}$$

which may also be written as

$$\mathbf{x}_t^* + \varepsilon^{-1} \boldsymbol{L}^* \mathbf{x}^* + \boldsymbol{N} (\mathbf{x}^*) = 0 \quad , \tag{2.3.7}$$

where the linear operator $\boldsymbol{L}^*$ is the same as $\boldsymbol{L}$ in (2.1.4) with f and Φ replaced by unity, and $\mathbf{x}^*$ is the nondimensional state vector.

II.4 Quasi-geostrophic balancing

Historically, the quasi-geostrophic balancing method was the first to be applied operationally in numerical weather prediction. It has since been replaced by other more sophisticated schemes. The common basis of initialization methods is to limit the amplitude of the high-frequency oscillations (gravity-inertia) largely characterized by their associated divergence field. Obviously, a flow evolving in such a way that the divergence field is identically zero for all times is free from high-frequency gravity-inertia waves since the principal mechanism for producing them is absent. This would be the case for example with the non-divergent barotropic vorticity equation. In the special context envisaged in this chapter, the shallow-water equations do posses these high-frequency linear free modes (

see section II.2). One might intuitively expect that setting the initial amplitude of the flow to be non-divergent and further that the time tendency of the divergent field be zero initially i.e. $D = \delta_t D = 0$ when $t = 0$, should be sufficient to remove (at least for a certain period of time) these high-frequency oscillations. It is nevertheless quite clear that this is at best approximate if one is to allow for a slow, weakly divergent meteorologically significant Rossby wave type of motion.

Since the divergence of synoptic-scale disturbances is small in comparison with the vorticity field, it is convenient to separate the wind into rotational and divergent parts using Helmholtz's decomposition theorem. Since the divergence is small according to (2.3.6) (i.e. is zero to $O(\varepsilon)$), we may write

$$\mathbf{v} = \mathbf{v}_\psi + \varepsilon\, \mathbf{v}_\chi\,, \quad \mathbf{v}_\psi = \mathbf{k}\times\nabla\psi, \quad \mathbf{v}_\chi = \nabla\chi \quad , \tag{2.4.1}$$

where ψ and χ are the dimensionless streamfunction and velocity potential. Applying the horizontal divergence operator to (2.3.5) and using (2.4.1), we get:

$$\varepsilon^2 \left\{\frac{\partial D}{\partial t} + \nabla\cdot(\mathbf{v}_\psi\cdot\nabla\mathbf{v}_\chi) + \nabla\cdot(\mathbf{v}_\chi\cdot\nabla\mathbf{v}_\psi)\right\} + \varepsilon^3\,\nabla\cdot\left(\mathbf{v}_\chi\cdot\nabla\mathbf{v}_\chi\right) +$$

$$+\ \varepsilon\,\nabla\cdot\left(\mathbf{v}_\psi\cdot\nabla\mathbf{v}_\psi\right) + \left(\nabla^2\phi - \zeta\right) = 0 \tag{2.4.2}$$

$$\text{where} \quad D = \nabla\cdot\mathbf{V}_\chi \ \text{ and } \zeta = \mathbf{k}\cdot\nabla\times\mathbf{V}_\chi \quad .$$

Thus to order ε^2 we get

$$\nabla^2\phi - \zeta + \varepsilon\,\nabla\cdot\left(\mathbf{v}_\psi\cdot\nabla\mathbf{v}_\psi\right) = 0 \quad ,$$

or in a more familiar form (Charney 1955) as

$$\nabla^2\psi - 2\,\varepsilon J(u_\psi, v_\psi) = \nabla^2\phi \tag{2.4.3}$$

which is the **classical nonlinear balance** equation relating the rotational part of the wind field with the geopotential field.

A diagnostic equation for the divergent part of the wind field may be obtained to the same order of approximation as (2.4.3), i.e. by neglecting terms of order ε^n where $n > 1$.

At this order of approximation, we write the vorticity and continuity equations (using (2.4.1), (2.3.5), and (2.3.6)) as

$$\varepsilon \frac{\partial \phi}{\partial t} + \varepsilon \, V_\psi \cdot \nabla \phi + \varepsilon \, D = 0 \quad , \tag{2.4.4a}$$

$$\varepsilon \frac{\partial \zeta}{\partial t} + \varepsilon \, V_\psi \cdot \nabla \zeta + \varepsilon \, D = 0 \quad . \tag{2.4.4b}$$

By taking the Laplacian of (2.4.4a) and combining with (2.4.3) and (2.4.4b), the desired diagnostic equation is

$$\nabla^2 D - D = V_\psi \cdot \nabla \zeta - \nabla^2 (V_\psi \cdot \nabla \phi) \quad .$$

This is the usual **quasi-geostrophic divergence equation** on an f-plane (see Haltiner and Williams 1980, chapter 3). Using a simple linearized shallow-water model on an f-plane, Phillips (1960) demonstrated that *both* the quasi-geostrophic divergence and balance equations are required to suppress gravity-inertia oscillations in the system. He further suggested a more general quasi-geostrophic set of divergence and balance equations to balance primitive equation models.

The balance equation (2.4.3) can be used in two different ways to initialize a primitive equation numerical model. One can determine the initial rotational part of the wind from the "observed" (or analysed) geopotential field. This is known as the direct balance equation. One can also attempt to determine geopotential from the observed wind. This method is referred to as solving the reverse balance equation. Apart from the obligation to specify consistent boundary conditions to solve (2.4.3), one faces a special problem of solvability with the direct balance equation.

For the reverse balance equation, (2.4.3) is an elliptic partial differential equation for ϕ and the resulting Poisson problem is well-posed (i.e. it has a unique solution which depends continuously on the data) provided one specifies the values of the streamfunction

or its derivative (i.e. Dirichlet or Neumann) at the boundaries of the domain of integration of the model. For the global problem no such boundary conditions are required.

The direct equation is of the " Monge-Ampère " type. As a boundary value problem, the nonlinear equation (2.4.3) admits at most two solutions if its coefficients *and the right-hand-side* of (2.4.3) (owing to the nonlinearity of the equation) are sufficiently well behaved. Equation (2.4.3) may be written as

$$E(\psi_{xx}\psi_{yy} - \psi_{xy}^2) + A\psi_{xx} + 2B\psi_{xy} + C\psi_{yy} - D = 0 \quad , \tag{2.4.5}$$

and it can be shown that (as an elliptic boundary value problem) it admits at most two solutions (Courant and Hilbert 1962, Vol. 2) provided the criterion

$$AC - B^2 - DE > 0 \quad , \tag{2.4.6}$$

is satisfied everywhere in the interior domain (Rellich 1934). Here we have

$$A = C = -\frac{1}{2}, \quad B = 0, \quad D = -\frac{1}{2}\nabla^2\phi, \quad E = \varepsilon \quad , \tag{2.4.7}$$

while (2.4.6) becomes

$$\varepsilon\nabla^2\phi + \frac{1}{2} > 0 \tag{2.4.8}$$

If ψ is a solution of (2.4.5), then using (2.4.6) we have the inequality

$$(E\,\psi_{xx} + C)\,(E\,\psi_{yy} + A) - (E\,\psi_{yy} - B)^2 > 0 \quad .$$

It follows that both $(E\psi_{xx} + C)$ and $(E\psi_{yy} + A)$ are either always greater than or always less than zero in the domain of interest. This means that there are two types of solutions of (2.4.5), one yielding $(1 + \varepsilon\nabla^2\psi < 0)$ and the other yielding $(1 + \varepsilon\nabla^2\psi > 0)$ in the domain. In dimensional form, this means that we have negative (resp. positive) absolute vorticity everywhere in the domain. The meteorologically significant solution retained for (2.4.5) is characterized by the requirement that in the Northern Hemisphere we have

$$\frac{1}{2} + \varepsilon\,\psi_{xx} > 0 \quad \text{and} \quad \frac{1}{2} + \varepsilon\,\psi_{yy} > 0$$

and in the Southern Hemisphere

$$\frac{1}{2} + \varepsilon\,\psi_{xx} < 0 \quad \text{and} \quad \frac{1}{2} + \varepsilon\,\psi_{yy} < 0 \quad .$$

By continuity we must let both ψ_{xx} and ψ_{yy} vanish at the equator since f = 0 there. It is however quite clear that the ellipticity condition is violated only when the scaling fails. Kasahara (1982b) examined the appearence of non-elliptic geopotential regions based on level IIIb FGGE data and found that the breakdown of (2.4.8) was frequent in the tropics. In the past, the ellipticity problem was avoided by altering the mass field. It is important to stress here that for reasons of consistency, the violation of the ellipticity criteria implies that the basic assumptions used to derive the balance equation do not apply. This point will be reexamined in the context of height-constrained implicit normal mode initialization in chapter IV. It should also be noted that important divergent circulations in the tropics were not taken into account in the previous derivation of the nonlinear balance equation.

II.5 The Bounded Derivative method

Browning et al. (1980) were the first to introduce the application of the " Bounded Derivative Initialization method (BDI) " for initializing the primitive equations model. The basic principle of this method was formulated by Kreiss (1979,1980) for systems governed by ordinary or partial differential equations. An interesting aspect of Kreiss's principle is that it may be used either in normal mode or physical space. The bounded derivative method can also be applied to systems with open boundaries (Browning and Kreiss 1982) where NMI is hardly applicable. Kreiss examined the requirement on the initial data to ensure slow time scale evolution of a given dynamical system for a certain period of time (to be defined). In order to simply express Kreiss's principle, let us call $\mathbf{x}(t)$ the state vector of the *nondimensionalized* dynamical system having a number of degree of freedom of the model (not necessarily the normal mode components). Kreiss then used the following observation for systems having slow and fast time scales: if $\mathbf{x}(t)$ varies slowly, then its first few time derivatives must satisfy

$$\frac{d^n x(t)}{dt^n} = O(1) \quad ; \quad n = 1, \ldots, p$$

where $p > 0$ and t is a slow timescale.

Applying this principle at the initial time $t = 0$ constrains the initial state to allow a slow time variation of $\mathbf{x}(t)$ for a given period of time where $0 < t < T = O(1)$. Note that it is not necessary to bound all derivative up to order p, this is automatic once the p^{th} order derivative has been bounded. Applications of this principle to ordinary differential equations are discussed in Kreiss (1979). The reader will also find in Kasahara (1982a) a simple application of the bounded derivative principle in the context of the normal mode form of the shallow-water equations and a comparison with Baer and Tribbia's (1977) filtering procedure. Also, Kasahara (1982a) examined (using the beta-plane approximation) the relationship between the BDI, nonlinear NMI and quasi-geostrophic theory for baroclinic primitive equations. To add to Kasahara's work, we establish the BDI method in physical space in a form appropriate to compare with the recent formulation of implicit nonlinear NMI (Temperton 1988, 1989 and Juvanon du Vachat 1986, 1988). We restrict the present analysis to the f-plane approximation.

We proceed in an analogous manner as in Browning et al. (1980) to derive the BDI scheme for the shallow-water equations. From (2.3.7), the first-order derivative is unity if and only if (dropping the "star" notation) $\boldsymbol{L}\,\mathbf{x}/\varepsilon$ is order unity, i.e., if and only if

$$\boldsymbol{L}\mathbf{x} = \varepsilon\,\mathbf{r} \tag{2.5.1}$$

and $\mathbf{r}$ is O(1). Second, to compute the second order time derivatives, we rewrite system (2.3.7) by using (2.5.1), i.e.

$$\mathbf{x}_t + \mathbf{r} + \boldsymbol{N}(\mathbf{x}) = 0 \tag{2.5.2}$$

therefore, the second order time derivative of $\mathbf{x}$ is of the order of unity if and only if $\mathbf{r}_t$ is of the order of unity. Using (2.5.1) and (2.5.2), it follows that

$$-\varepsilon\,\mathbf{r}_t = -\boldsymbol{L}\,\mathbf{x}_t = \boldsymbol{L}\,\mathbf{r} + \boldsymbol{L}\boldsymbol{N}(\mathbf{x}) \quad . \tag{2.5.3}$$

Combining (2.5.1) and (2.5.3) yields

$$L^2 \mathbf{x} = -\varepsilon L N(\mathbf{x}) - \varepsilon^2 \mathbf{r}_t \quad . \tag{2.5.4}$$

Dropping the ε^2 term in (2.5.5) and setting $\varepsilon = 1$ to get a dimensional relation, we get

$$L^2 \mathbf{x} = -L N(\mathbf{x}) \quad , \tag{2.5.5}$$

to $O(\varepsilon^2)$ and L. is given by (2.1.4). As mentioned in Browning et al., an iterative procedure can be used to solve (2.5.5) and it is also shown that it is related to the classical nonlinar balance equation. Our interest will however be to directly compare the preceding result (2.5.5) with the balancing constraint from Machenhauer's scheme in the context of implicit NMI as described in section II.7.

II.6 Initialization schemes in normal mode space

This class of initialization schemes operates in normal mode space. They differ from the bounded derivative method in that they are usually applied to a subset of the fast modes only. In some sense they are more controllable and may be more usefull for systems for which the scaling assumptions are not very well satisfied. As a preliminary requirement for applying the initialization schemes to be presented in later sections of this chapter, we introduce the formal procedure to get the normal mode form of the governing dynamical equations. The selected finite set of basis functions (e.g. spherical harmonics, finite-elements, etc.) or grid points discretization, transform the shallow-water equations into a set of ordinary differential equations in time usually called a **dynamical system.** The model equations may then be written as

$$\frac{d\mathbf{y}}{dt} + i\,\mathbf{A}\mathbf{y} = \mathbf{N}(\mathbf{y}) \quad .$$

The *state vector* **y** will represents the vector of predictive variables, **A** is a (real and symmetric) constant coefficient matrix. The normal mode decomposition of **A** is

$$\mathbf{A} = \mathbf{E}\Lambda\mathbf{E}^T \tag{2.6.1}$$

where Λ is a diagonal matrix of eigenvalues of **A**, **E** is an orthogonal matrix whose columns corresponds to the eigenvectors of **A**. We further assume that the eigenvalues of **A** which are real, can be split into slow (Rossby) or fast (gravity) parts. This splitting can be represented for Λ as:

$$\Lambda = \begin{bmatrix} \Lambda_R & 0 \\ 0 & \Lambda_G \end{bmatrix} .$$

The normal mode decomposition of the dynamical system is given by

$$\mathbf{E}^T \frac{d\mathbf{y}}{dt} + i\,\mathbf{E}^T\mathbf{A}\mathbf{y} = \mathbf{E}^T\,\mathbf{N}(\mathbf{y}) \tag{2.6.2}$$

or

$$\frac{d\mathbf{z}}{dt} + i\,\Lambda\,\mathbf{z} = \mathbf{Q} \tag{2.6.3}$$

where $\mathbf{z} = (\mathbf{R},\mathbf{G})$ i.e. the state vector in normal mode space having both Rossby and gravity-inertia components, while **Q** now represents the projected nonlinear terms. **R** and **G** are vector amplitudes of the slow and fast mode of the system respectively.

a. Baer and Tribbia's initialization method

The starting point for this scheme (first derived by Baer and Tribbia 1977 using a somewhat different approach where fast *and* slow time scales were introduced) is the nondimensional form of the shallow-water equations (2.3.7) which is written on a slow (advective) timescale. Noting that the time tendncy and nonlinear terms in (2.3.7) are O(1) and that we may write Λ_G is O(1) and Λ_R is O(ε), we get

$$\frac{d\mathbf{R}}{dt} + i\Lambda_R\,\mathbf{R} = \mathbf{Q}_R(\mathbf{R},\mathbf{G}) \tag{2.6.4}$$

$$\varepsilon\frac{d\mathbf{G}}{dt} + i\Lambda_G\,\mathbf{G} = \varepsilon\,\mathbf{Q}_G(\mathbf{R},\mathbf{G}) \tag{2.6.5}$$

Provided the Rossby number ε is smaller than unity, **G** may be developped in a power series in terms of ε as:

$$\mathbf{G} = \sum_{n=0}^{N} \varepsilon^n \mathbf{G}^{(n)} \quad . \tag{2.6.6}$$

Directly inserting (2.6.6) into (2.6.5) leads to the appropriate approximation to order ε desired for slow behaviour of the **G**'s. As an example, we get for the lower order approximations, the sequence

$$\mathbf{G}^{(0)}(t) \equiv 0 \tag{2.6.7}$$

$$\mathbf{G}^{(1)}(t) = -i \Lambda_G^{-1} \mathbf{Q}_G(\mathbf{R},0) \quad . \tag{2.6.8}$$

By **R** here we mean the slow mode amplitudes given at time t. The nonlinear terms in (2.6.8) represents interactions between the slow modes only, due to our first order estimate (2.6.7).

b. Lorenz 's scheme

We now exibit a second method of normal mode initialization which involves an *iteration* process in order to specify the desired balancing amplitude of the fast mode components of the system at a given time. The first iterative nonlinear NMI scheme was introduced by Machenhauer (1977). This scheme can be seen to be a particular case of a more general higher-order initialization scheme introduced by Lorenz (1980). His condition of "superbalance" is given by

$$\frac{d^n \mathbf{G}}{dt^n} = 0$$

that is, the n^{th} order time tendency of the fast mode components are required to be zero, where " n " should be a large number. Lorenz's method is therefore a special case of BDI to order ¨n¨ (Lorenz 1986). It should be borne in mind that for a given order ¨n¨, one may encounter non existence or multiple branch solutions. Thaning (1984) gave simple examples of this for the first order scheme (i.e. Machenhauer's scheme). Lorenz applied his scheme in the context of a 9 component f-plane shallow-water model including forcing, diffusion and bottom topography. Machenhauer's (1977) and Leith's (1980) iterative schemes are particular cases of Lorenz's scheme when n = 1, 2 respectively. Tribbia

(1984) proposed an extension of Machenhauer's scheme (see also Machenhauer 1982) by expanding the nonlinear terms of the fast mode equations (2.6.5) in a Taylor series about the initial time. His method of successive approximation is shown to be asymptotically equivalent to Baer and Tribbia's method but has the advantage of eliminating the need to compute Frechet derivatives (see also Tribbia 1979). We review in the following the standard derivation of Machenhauer's scheme.

Due to the nonlinear nature of the dynamical system (2.6.3), simply setting to zero the initial amplitude of the fast mode components of the initial data is not sufficient to adequately control the large amplitude oscillations of these components during a model integration (Williamson 1976).

Machenhauer's (1977) initialization scheme is based on the observation that for the noise components, their nonlinear forcing term varies "slowly". Assuming that $(Q_G) \cong$ cst in time, then for a given gravity mode component "n"

$$G_n(t) = \left[G_n(0) + i \frac{(Q_G)_n}{\sigma_n} \right] \exp(i\sigma_n t) + i \frac{(Q_G)_n}{\sigma_n} \qquad (2.6.9)$$

for the *fast* mode components with frequency σ_n. We may then consider the nonlinear initialization scheme where one attempts to set

$$G_n(0) = -i \frac{(Q_G)_n}{\sigma_n}$$

in order to eliminate the fast oscillations in the system. Since $(Q_G)_n$ depends on the amplitudes G_n, one defines an iterative procedure where

$$G_n^{(\mu+1)}(0) = -i \frac{(Q_G)_n^{(\mu)}}{\sigma_n} \qquad (2.6.10)$$

where μ is the iteration number. Equation (2.6.10) describes Machenhauer's nonlinear normal mode initialization scheme. Combining (2.6.3) with (2.6.10) gives the standard form of Machenhauer's iterative scheme as

$$\Delta G_n = - i \frac{(\delta_t G_n)^{(\mu)}}{\sigma_n} \qquad (2.6.11)$$

where

$$\Delta G_n = G_n^{(\mu+1)}(0) - G_n^{(\mu)}(0) \qquad (2.6.12)$$

$$\text{and} \quad \delta_t G_n \equiv \frac{dG_n}{dt} \quad .$$

Further references to ΔG_n and $(\delta_t G_n)$ in the text should be understood respectively as the changes that have to be made to the coefficient of the fast mode of index "n" and the tendency of the fast mode as evaluated by running the model for one forward timestep. The following derivation should illustrate clearly the iterative nature of higher order schemes.

Let's consider the dimensional form of (2.6.5) (i.e. $\varepsilon = 1$)

$$\frac{dG}{dt} + i \Lambda_G G = Q_G(R,G) \quad . \qquad (2.6.13)$$

The time derivative of (2.6.13) gives

$$\frac{d^2G}{dt^2} + i \Lambda_G \frac{dG}{dt} = \frac{dQ_G}{dt} \quad . \qquad (2.6.14)$$

Setting the second order time derivative to zero in (2.6.14) and substituting (2.6.13) into (2.6.14), we obtain

$$G = - \frac{1}{(i\Lambda_G)^2} \left[\frac{dQ_G}{dt} - i \Lambda_G Q_G \right] \qquad (2.6.15)$$

or, using an analogous technique used previously for Machenhauer's scheme, this latter result may be written as

$$\Delta G = - \frac{1}{(i\Lambda_G)^2} \frac{d^2G^{(\mu)}}{dt^2} \qquad (2.6.16)$$

where μ is the iteration index as in (2.6.12). The straightforward extension to a n^{th} order Lorenz's scheme is given by

$$\Delta G = \frac{(-1)^{n-1}}{(i\Lambda_G)^n} \frac{d^n G^{(\mu)}}{dt^n} \quad . \tag{2.6.17}$$

The solution technique for (2.6.17) may be to start with a linear balance $G^{(0)} = 0$, then determining a second estimate i.e. $G^{(1)}$ by iterating the first order scheme (or Machenhauer's scheme), then from this new estimate a third estimate is obtained by iterating the second order scheme (2.6.16). This process was used by Lorenz (1980) with n = 3. Temperton (1988, 1989) showed the benefit of a second order scheme over first order for the refinement of the balance (especially in the tropics, see also Browning et al., section 6) in the context of shallow-water models with a large number of degrees of freedom in the absence of external forcing.

II.7 The Implicit nonlinear Normal Mode Initialization method

a. Basic formulation

Following Temperton 1988, one iteration of Machenhauer's initialization scheme described by (2.6.11) may be written as:

$$\Delta y = i\, E_G \Lambda_G^{-1} E_G^T (\delta_t y)^{(\mu)} \tag{2.7.1}$$

where $\Delta y = y(o)^{(\mu+1)} - y(o)^{(\mu)}$, and the observed model state tendency $(\delta_t y)$ are expressed in physical space. Also

$$E = [E_R \vdots E_G] \quad ; \quad \Lambda = \begin{bmatrix} \Lambda_R & 0 \\ 0 & \Lambda_G \end{bmatrix}$$

where the columns of E_R and E_G corresponds to the slow (Rossby) and fast (gravity) modes respectively, and Λ_R , Λ_G are the low and high frequencies. The time tendency of the state vector is obtained by running the model for one forward timestep. Temperton's formulation of a physical space nonlinear NMI scheme (INMI) is based on the following simple observation. If equation (2.7.1) is multiplied by the matrix A and use is made of the similarity transformation for A, i.e.

$$\mathbf{A} = \mathbf{E}\Lambda\mathbf{E}^{\mathrm{T}} \tag{2.7.2}$$

then it follows that

$$\mathbf{A}\,\Delta\mathbf{y} = \mathrm{i}\,\mathbf{E}_G\mathbf{E}_G^{\mathrm{T}}\,(\delta_t\mathbf{y}) \tag{2.7.3}$$

or

$$\mathbf{A}\,\Delta\mathbf{y} = \mathrm{i}\,\mathbf{P}_G\,(\delta_t\mathbf{y}) \tag{2.7.4}$$

where $\mathbf{P}_G$ acts as a projection matrix which extracts the fast mode component of $(\delta_t\mathbf{y})$ and is the only unknown in (2.7.4) if no attempt is made to explicitly calculate the normal modes of matrix **A**. In practice, there are a number of ways to solve (2.7.4) (see Temperton 1989, hereinafter noted T89) in physical space. All of which have one thing in common, that is, the linearization is such that the dynamical system has slow modes which are all stationary.

Temperton (1988) (hereinafter noted T88) stressed that due to this particular choice of linear terms forming the matrix **A**, the latter will be singular and as is well known the linear system (2.7.4) will have a solution provided that the right-hand-side is orthogonal to the null space of **A**. This is guaranteed by the orthogonality of the eigenmodes of **A** and the fact that the term $\mathbf{P}_G(\delta_t\mathbf{y})$ if properly computed, has no component on the slow subspace spanned by the slow modes. In practical applications, the gravity mode "projector" $\mathbf{P}_G$ is not determined explicitly but rather certain basic properties of the slow and fast modes (i.e. stationary-nondivergent slow modes and zero linearized potential vorticity for the fast modes) are exploited to ensure unique solutions of (2.7.4). We may also note that since matrix **A** is representable as a sum of gravity mode projectors only (spectral decomposition theorem), we obtain by left multiplying (2.7.4) by **A**

$$\mathbf{A}^2(\Delta\mathbf{y}) = \mathrm{i}\,\mathbf{A}\,(\delta_t\mathbf{y}) \quad . \tag{2.7.5}$$

Thus (2.7.5) is directly solvable for $\Delta\mathbf{y}$ (belonging to the fast subspace) in terms of known quantities (see also Juvanon du Vachat, 1989, p.52). The system (2.7.5) being degenerate

(remember that A is singular here) one has to use the same supplementary condition as the one used by T88, and T89, that is, the fast gravity modes have zero linearized potential vorticity in order to close the system of linear equations (i.e. the usual condition to remove geostrophic degeneracy). Although (2.7.5) has no practical advantages over the methods of T88, T89 it is helpful when considering the relation between the implicit NMI method and the real-space bounded derivative method of Browning et al. (1980).

b. A simple illustration of the method

To illustrate the application of (2.7.5), we consider the f-plane shallow-water equations on a polar stereographic projection true at latitude θ_o in differentiated form (ref. T88) :

$$\frac{\partial \zeta}{\partial t} = - f D + N_\zeta \qquad (2.7.6)$$

$$\frac{\partial D}{\partial t} = f \zeta - \nabla^2 \phi + N_D \qquad (2.7.7)$$

$$\frac{\partial \phi}{\partial t} = -m^2 \Phi D + N_\phi \qquad (2.7.8)$$

where ζ, D are the vorticity and divergence fields, ϕ is a perturbation geopotential, f is the Coriolis parameter (constant here), and "m" is the map scale factor given by:

$$m = \frac{1 + \sin(\theta_o)}{1 + \sin(\theta)} \quad .$$

The nonlinear terms are represented by N_ζ , N_D , N_ϕ . Equation (2.7.4) corresponds to

$$\begin{bmatrix} 0 & f & 0 \\ -f & 0 & \nabla^2 \\ 0 & m^2 \Phi & 0 \end{bmatrix} \begin{bmatrix} \Delta\zeta \\ \Delta D \\ \Delta\phi \end{bmatrix} = \begin{bmatrix} \delta_t \zeta \\ \delta_t D \\ \delta_t \phi \end{bmatrix}_G \qquad (2.7.9)$$

where $\Delta\zeta$, ΔD, $\Delta\phi$ are changes to be made to the vorticity, divergence, geopotential fields in physical space and $\delta_t = \partial_t$. The index G refers to gravitational mode components. The

straightforward application of (2.7.5) to the present model in order to avoid the necessity of projecting each vectors appearing in (2.7.9) onto the fast subspace gives:

$$\begin{bmatrix} 0 & f & 0 \\ -f & 0 & \nabla^2 \\ 0 & m^2\Phi & 0 \end{bmatrix}\begin{bmatrix} 0 & f & 0 \\ -f & 0 & \nabla^2 \\ 0 & m^2\Phi & 0 \end{bmatrix}\begin{bmatrix} \Delta\zeta \\ \Delta D \\ \Delta\phi \end{bmatrix} = \begin{bmatrix} 0 & f & 0 \\ -f & 0 & \nabla^2 \\ 0 & m^2\Phi & 0 \end{bmatrix}\begin{bmatrix} \delta_t\zeta \\ \delta_t D \\ \delta_t\phi \end{bmatrix}$$

$$\begin{bmatrix} -f^2 & 0 & f\nabla^2 \\ 0 & -f^2+\nabla^2(\Phi m^2) & 0 \\ -f\Phi m^2 & 0 & m^2\Phi\nabla^2 \end{bmatrix}\begin{bmatrix} \Delta\zeta \\ \Delta D \\ \Delta\phi \end{bmatrix} = \begin{bmatrix} f\,\delta_t D \\ -f\,\delta_t\zeta + \delta_t\nabla^2\phi \\ m^2\Phi\,\delta_t D \end{bmatrix}$$

the first row gives:

$$\nabla^2(\Delta\phi) - f\Delta\zeta = \delta_t D \qquad (2.7.10)$$

the second row gives:

$$-f^2(\Delta D) + \nabla^2(\Phi m^2(\Delta D)) = \nabla^2(\delta_t\phi) - f(\delta_t\zeta) \quad . \qquad (2.7.11)$$

From the last matrix expression, it is clear that the first and third row are linearly dependent (this was to be expected from the dimension of the original operator **A**). To close system (2.7.10) and (2.7.11) we recall from section II.2 that since fast modes are characterized in the f-plane approximation by zero linearized potential vorticity , so are the changes on these modes, i.e.

$$m^2\Phi\Delta\zeta = f\Delta\phi \quad . \qquad (2.7.12)$$

System (2.7.10) to (2.7.12) is identical to the implicit scheme described in section 4 of T88 and obtained in a slightly different manner.

c. The relationship between INMI and BDI methods

In order to bring the BDI and INMI methods closer together, we make use first of the continuous form of the shallow-water equations (differentiated form or not) as :

$$\frac{\partial x}{\partial t} + L x + N(x) = 0 \quad . \tag{2.7.13}$$

We now consider the more general case where f is not considered as constant (i.e. models on the sphere or projected grids for example). For such a case, L will posses slow modes that are not stationary. However, let us use the following splitting of the linear operator L

$$L \rightarrow \tilde{L} + L' \tag{2.7.14}$$

such that $\tilde{L}$ has stationary Rossby modes. Using the continuous form of (2.7.5), one iteration of Machenhauer's scheme may be written as :

$$\tilde{L}^2 (\Delta x) = \tilde{L} (\delta_t x)_0 \tag{2.7.15}$$

where the δ_t operator includes nonlinear-terms. Using (2.7.14), (2.7.13) becomes (starting from an initial guess x_0):

$$\frac{\partial x_0}{\partial t} + \tilde{L} x_0 = - N (x_0) - L'(x_0) \tag{2.7.16}$$

take $\tilde{L}$(2.7.16):

$$\tilde{L} \left(\frac{\partial x_0}{\partial t} \right) + \tilde{L}^2 (x_0) = - \tilde{L} N (x_0) - \tilde{L} L'(x_0) \quad . \tag{2.7.17}$$

Now use (2.7.15) into (2.7.17) to get

$$\tilde{L}^2 (\Delta x) + \tilde{L}^2 (x_0) = - \tilde{L} N (x_0) - \tilde{L} L'(x_0) \tag{2.7.18}$$

thus

$$\tilde{L}^2 (x_1) = - \tilde{L} N (x_0) - \tilde{L} L'(x_0) \tag{2.7.19}$$

where

$$x_1 \equiv x_0 + \Delta x \quad .$$

It is also clear that the iteration process involved in Machenhauer's scheme means that the new estimate x_1 may be inserted on the right-hand-side of (2.7.19) to get x_2 and so on.

On an f-plane, $L' = 0$ and iterating (2.7.19) is equivalent to using the bounded derivative approach described in section II.5. The comparison between these two methods for the more general case of a variable Coriolis parameter will not be examined here. It is

important to note however that the way INMI proceeds in characterizing slow and fast mode components in its diagnostic equations (T88, T89 and Juvanon du Vachat 1988) has given a concrete advantage for extending INMI to a variational version in a form first suggested by Daley (1978) for explicit nonlinear NMI. This aspect will be considered fully in subsequent chapters and forms the essential part of the thesis.

Chapter III

Variational implicit normal mode initialization for a regional finite-element shallow-water model

III.1 Presentation of article 1

We now attack the question of the feasibility of Daley's (1978) approximation (see chapter I) in the context of impicit nonlinear NMI. Temperton's formulation (1988) of INMI was successfully applied to the Canadian regional finite-element shallow-water model. The use by the model of a projected variable resolution grid renders the computation of the model's normal modes intractable. On the other hand, the model has well posed boundary conditions and Temperton succeeded in formulating well posed boundary conditions for the nonlinear initialization problem in physical space.

Daley's approximation was first formulated in normal mode space and required the specific knowledge of the structure of the fast gravity modes to be initialized. Each of these modal structures enters in the form of an integral constraint (scalar product) imposed as a strong constraint (Sasaki 1958). The problem here is then clear: " How can we characterize these constraints on the slow and fast mode corrections in physical space while at the same time satisfying a minimization criteria on the total correction fields" . The essential ingredients will be shown to be clearly identified in the implicit context and putting the pieces together by the application of the calculus of variations (see the appendix at the end of chapter V, p101), will lead to a well posed variational problem which is referred to as VINMI.

Before proceeding to the details of the method, it is stressed immediatly here that based on the close relationship between INMI and BDI, a variational extension of the BDI method would proceed by a similar reasoning as the one presented hereafter. The clear

advantage of INMI here is its natural use of the concepts of slow and fast components of the flow, thus facilitating the application of Daley's variational procedure.

III.2 Article 1

Variational implicit normal mode initialization

Monthly Weather Review, 117, 2219-2229.

Reprinted from MONTHLY WEATHER REVIEW, Vol. 117, No. 10, October 1989
American Meteorological Society

Variational Implicit Normal Mode Initialization

LUC FILLION

Université McGill, Montréal, Québec, Canada.

CLIVE TEMPERTON

Recherche en prévision numérique, Atmospheric Environment Service, Dorval, Québec, Canada.

(Manuscript received 28 November 1988, in final form 8 May 1989)

ABSTRACT

It is shown that implicit normal mode initialization can be combined with a variational technique, in order to control the relative magnitudes of the changes to the analyzed mass and wind fields. Since the initialization procedure is expressed entirely in physical space, the use of locally varying weights in the variational integral becomes more straightforward than in previous efforts to combine variational methods with normal mode initialization.

We present details of the application to a finite-element model of the shallow water equations on a stereographic projection. It is demonstrated that the use of variational initialization can change the slowly evolving component of the subsequent forecast, as well as eliminate the unrealistic fast component.

1. Introduction

Most data assimilation systems consist of three components: an analysis step, a nonlinear normal mode initialization (NMI) procedure, and a short forecast (typically 6 hours) to provide the background field for the next analysis. The initialization step ensures a correct dynamical balance by allowing mutual adjustment of the analyzed mass and wind fields. Although these adjustments are generally small (Hollingsworth et al. 1986), the relative magnitudes of the changes to the mass and wind fields are governed essentially by geostrophic adjustment theory rather than by the reliability of the mass and wind analyses. In particular, the initialization step may result in changes to the surface pressure analysis which are locally larger than seems reasonable in view of the expected analysis error.

Daley (1978) first suggested a variational form of nonlinear normal mode initialization for a barotropic spectral model. In this scheme the balance implied by nonlinear NMI was achieved, while simultaneously minimizing a variational integral incorporating weights based on the presumed accuracy of the mass and wind analyses. Daley and Puri (1980) studied the impact of the variational procedure on a simulated data assimilation scheme. Puri (1983) extended the variational NMI technique to a multilevel spectral model. Temperton (1984) developed a variational NMI scheme for the ECMWF multilevel gridpoint model, and proposed a computationally more efficient approach to the problem. In all these applications, the horizontal (and vertical) distribution of the weights were artificially simple, depending at most on latitude. It was recognized that a more general specification of the weights, although clearly desirable in a realistic setting, would lead to a computationally difficult problem. The essence of this difficulty is that the weights are naturally expressed in physical space, while the initialization is performed in normal mode space. Tribbia (1982) successfully performed variational NMI with a general horizontal specification of the weights, but only for a low-resolution barotropic spectral model; the extension to models with more degrees of freedom remained apparently intractable.

Recently, an implicit form of normal mode initialization has been introduced (Temperton 1988). This scheme was developed in order to allow nonlinear NMI to be applied in models whose normal modes could not readily be found, for example, because of nonseparability of the underlying linear equations. Although equivalent to conventional NMI, the implicit form is expressed entirely in physical space, and thus offers a way out of the difficulty discussed above in solving the corresponding variational problem.

In this paper we present a variational form of the implicit normal mode initialization procedure, which allows a general specification of the weights in the variational integral. The scheme is formulated in section 2. Experimental results are presented in section 3 for

Corresponding author address: Dr. Luc Fillion, Recherche en prévision numérique, 2121, voie de service nord, Porte 508, Route Trans-canadienne, Dorval, Québec H9P 1J3, Canada.

the same barotropic finite-element regional model as was used by Temperton (1988). Section 4 contains a discussion and summary.

2. Theoretical aspects

a. Daley's approximation

Daley's (1978) procedure will serve as the starting point for the formulation of the variational algorithm. His basic idea can be summarized in symbolic form as seen below. Suppose we have preconstructed the whole set of model (linear) orthonormal modes denoted by $\{H_i\}_{i=1,\ldots,N}$ where N is the dimension of the linear space H spanned by these modes. Assuming completeness of that space (within the model) any state vector can be uniquely represented. Given a scalar product on that space, a correction vector will be represented by

$$\Delta \mathbf{x} = \sum_{i=1}^{N} \Delta x_i \mathbf{H}_i, \tag{2.1}$$

where

$$\Delta x_i = \langle \Delta \mathbf{x}, \mathbf{H}_i \rangle. \tag{2.2}$$

Now on the basis of some selection criteria, suppose we can identify each eigenmode as belonging to a slow or a fast subspace of H. Then for each element of the fast subspace H_G, nonlinear NMI will ascribe to it a specific amplitude. We thus have

$$(\Delta \mathbf{x})_G = \sum_{i=1}^{M} (\Delta x_i)_G \mathbf{H}_i, \tag{2.3}$$

where M is the dimension of H_G.

These changes to the initial fields may be uncomfortably large in certain geographical regions compared with the expected analysis error known a priori for these fields. It is then possible to take into account this expected error by defining a "fidelity metric" J which is a function of these correction fields, and which when minimized permits us to stay closer to the observations.

Daley (1978) first introduced the idea that it should be possible to control the adjustment process for achieving the balance between Rossby and gravity modes in such a way as to minimize J. Given a "slave" relation of the form

$$\mathbf{x}_G = G(x_R), \tag{2.4}$$

which expresses the total dependence of the amplitude of the gravity modes upon Rossby modes in a nonlinear initialization scheme, we may express formally our function in terms of these two sets of modes by

$$J = J((\Delta \mathbf{x})_G, (\Delta \mathbf{x})_R). \tag{2.5}$$

In fact, because of the relation (2.4) we must recognize that the tuning amplitudes are ultimately those of Rossby modes. However, the nonlinear nature of the relation (2.4), and the need to build a variational algorithm that is efficient for practical applications, suggests an iterative procedure for approximating the solution. This procedure can be summarized as follows:

(i) Given an initial state $\mathbf{x}^0$, generate a correction vector $(\Delta \mathbf{x})_G$ where the new state

$$\mathbf{x}^1 = \mathbf{x}^0 + (\Delta \mathbf{x})_G$$

is in approximate balance.

(ii) Alter the Rossby part of the initial state to minimize J in (2.5).

Relation (2.4) then indicates a need to readjust the amplitude of the gravity modes. The iteration loop with steps 1 and 2 can thus be used to approximate the original problem. As Tribbia (1982) indicated, this procedure is not completely equivalent to directly minimizing (2.5) under the strong constraint (2.4). Each iteration of Daley's scheme solves a new minimization problem, with $(\Delta \mathbf{x})_G$ and $(\Delta \mathbf{x})_R$ in (2.5) relating only to the increments during that iteration, rather than to the accumulated changes made to the original fields. However, the results of Tribbia (1982) suggested that Daley's procedure provides a very good approximation to the solution of the full minimization problem.

Thus Daley determines

$$\Delta \mathbf{x} = (\Delta \mathbf{x})_R + (\Delta x)_G$$

by a constrained variational problem where each of these constraints is of the isoperimetric (integral) type and can be written in symbolic form as (2.2) for each element of H_G. This variational problem is iterated until stopping criteria (to be discussed later) are achieved.

Daley's approach (using integral constraints) has the undesirable effect of producing a large linear system of equations to be solved if the variational problem is attacked by Euler–Lagrange equations, since the dimension of this linear system is directly proportional to the number of Lagrange multipliers (or constraints) present. This dimensionality can be reduced by certain restrictions on the spatial variability of the control parameters present in the variational problem but there is a clear limitation for applications.

b. The implicit form of variational NMI

It is the purpose of this paper to reconsider Daley's approach in relation to the newly developed implicit normal mode initialization scheme presented by Temperton (1988), hereinafter referred to as T(88). We summarize here the salient features of this procedure and refer the reader to T(88) for further details.

First, if an appropriate choice is made for the linearized model equations which determine the normal modes, two fundamental properties come out concerning slow and fast eigenmodes:

Property 1. Slow modes are stationary and nondivergent, and

Property 2. Fast modes have zero linearized potential vorticity.

Let $\Delta\mathbf{x}$ represent a model correction vector. We can express Machenhauer's (1977) scheme, for example, with the previously mentioned special form of the linear operator, denoted by A, as

$$A\Delta\mathbf{x} = iP_G(\delta_i\mathbf{x})_o$$

where P_G is the projection operator onto H_G, and $(\delta_i\mathbf{x})_o$ is an observed model-state tendency. It is then possible with property 1 and 2, and also requiring that $\Delta\mathbf{x}\epsilon H_G$, to specify $\Delta\mathbf{x}$ uniquely without having to know the structure of the eigenmodes of A. In the case of Daley's variational approach, we need to know the structure of each gravity mode being initialized since they appear explicitly in each constraint of the form (2.2). The following will show how to circumvent this difficulty using the implicit form of nonlinear NMI.

Since variational initialization is intended to be a special procedure for controlling changes being made during initialization, the controlled variables taken here will be the same as those used for the implementation of the unconstrained initialization scheme. In order to introduce our variational algorithm, we have chosen the regional finite-element barotropic model of Staniforth and Mitchell (1977, 1978) as a reference. Consequently, the vector of controlled variables will be

$$\Delta\mathbf{x} = (\Delta\psi, \Delta\chi, \Delta\phi)$$

where use has been made of Helmholtz's decomposition theorem for the wind field

$$\Delta U = \frac{\partial(\Delta\psi)}{\partial y} + \frac{\partial(\Delta\chi)}{\partial x}, \quad \Delta V = \frac{\partial(\Delta\psi)}{\Delta x} + \frac{\partial(\Delta\chi)}{\partial y}.$$

Here U, and V are wind images related to the u and v components of the wind vector by

$$U = \frac{u}{m}, \quad V = \frac{v}{m}$$

where "m" is the map-scale factor and is given at latitude "θ" by

$$m = \frac{1 + \sin(60°)}{1 + \sin(\theta)}$$

for a polar stereographic projection, true at 60°N.

Property 1 thus takes the simple form for each correction vector

$$(\Delta\chi)_R = 0, \quad (\Delta\chi) = (\Delta\chi)_G;$$

that is, the gravitational part of the divergent windfield correction is governed by unconstrained initialization. The geopotential and streamfunction are then the only fields to be controlled by the following function,

$$J = \iint_D \frac{w_\phi}{m^2}(\Delta\phi)^2 + \Phi w_\psi(\nabla(\Delta\psi))^2 dxdy \quad (2.6)$$

where "D" is the physical domain, "m" is the map scale factor for projected grids and Φ represents the domain average of geopotential. This function is quadratic in the geopotential and the rotational part of the wind field. It incorporates weighting factors for each of these two fields, which are allowed to vary in both spatial directions. Setting these factors between zero and 1 will permit a mutual adjustment between ϕ and ψ. They must also be specified in such a way as to produce acceptable changes of the controlled fields where we know something about the expected amplitude of these changes. This information can, in principle, be extracted from an objective analysis scheme. At the same time, it should be borne in mind that certain choices of the weights may lead to unrealistic results; for example, setting w_ψ too close to zero in the tropics may lead to ellipticity problems (Daley 1978; Tribbia 1981; Temperton 1984).

It should be mentioned here that the functional given in (2.6) is simply a choice and is not the only one possible. Other forms are possible including, for example, the gradient of geopotential which has been used previously by Daley (1978). Now, from the unconstrained scheme we know

$$(\Delta\mathbf{x})_G = \{(\Delta\phi)_G, (\Delta\psi)_G, (\Delta\chi)_G\}.$$

During step 2, we look for a correction vector on slow modes,

$$(\Delta\mathbf{x})_R = \{(\Delta\phi)_R, (\Delta\psi)_R, 0\}$$

that will make J a minimum. To be consistent with the implicit scheme [T(88)], property 1 must be the required contraint on $(\Delta\phi)_R, (\Delta\psi)_R$

$$\nabla^2(\Delta\phi)_R = \mathcal{F}(\Delta\psi)_R \quad (2.7)$$

where

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$$

$$\mathcal{F} = \frac{\partial}{\partial x}\left(f\frac{\partial}{\partial x}\right) + \frac{\partial}{\partial y}\left(f\frac{\partial}{\partial y}\right),$$

together with its appropriate boundary condition

$$\frac{\partial}{\partial n}(\Delta\phi)_R = f\frac{\partial}{\partial n}(\Delta\psi)_R \quad \text{on} \quad \Gamma \quad (2.8)$$

where Γ represents the boundary of D, $f(x, y)$ is the Coriolis parameter, and the derivative with respect to "n" is an outward normal derivation on Γ.

Following T(88), the gravitational part of the controlled correction fields satisfies a linear relation of the same form as (2.7) given by

$$\nabla^2(\Delta\phi)_G - \mathcal{F}(\Delta\psi)_G = (\delta_i\nabla^2\chi)_o + \mathcal{B}(\Delta\chi)_G \quad (2.9)$$

where

$$\mathcal{B} = \frac{\partial}{\partial x}\left(f\frac{\partial}{\partial y}\right) - \frac{\partial}{\partial y}\left(f\frac{\partial}{\partial x}\right)$$

with boundary conditions

$$\frac{\partial}{\partial x}(\Delta\phi)_G = f\frac{\partial}{\partial x}(\Delta\psi)_G + f\frac{\partial}{\partial y}(\Delta\chi)_G \text{ on } \Gamma_y$$

$$\frac{\partial}{\partial y}(\Delta\phi)_G = f\frac{\partial}{\partial y}(\Delta\psi)_G - f\frac{\partial}{\partial x}(\Delta\chi)_G \text{ on } \Gamma_x \tag{2.10}$$

where

$$\Gamma = \Gamma_x U \Gamma_y.$$

Since the correction fields present in the function concern the Rossby part as well as the gravitational part, it is preferable to work with a constraint written in terms of these fields. Since (2.7) and (2.9) are linear equations, a unique constraint can be formed by adding (2.7) to (2.9) and forming its associated boundary condition by adding (2.8) to (2.10). The differential constraint is thus

$$M(\Delta\phi, \Delta\psi) = \nabla^2(\Delta\phi) - \mathcal{F}(\Delta\psi) - (\delta_t\nabla^2\chi)_o - \mathcal{B}(\Delta\chi)_G = 0$$

$$\frac{\partial}{\partial x}(\Delta\phi) = f\frac{\partial}{\partial x}(\Delta\psi) + f\frac{\partial}{\partial y}(\Delta\chi)_G \quad \text{on } \Gamma_y$$

$$\frac{\partial}{\partial x}(\Delta\phi) = f\frac{\partial}{\partial y}(\Delta\psi) - f\frac{\partial}{\partial x}(\Delta\chi)_G \quad \text{on } \Gamma_x. \tag{2.11}$$

In addition to constraint (2.11), conservation of mass over D is required which implies

$$\iint_D \frac{(\Delta\phi)}{m^2}\,dxdy = 0 \tag{2.12}$$

so that the mean value of the geopotential is not changed.

Thus our variational problem is to minimize J given by (2.6) under the differential constraint (2.11), and the integral constraint (2.12). Using classical theorems of the calculus of variations (Smirnov 1964) our problem is equivalent to the minimization of the new function given by

$$J = \iint_D \left\{\frac{w_\phi}{m^2}(\Delta\phi)^2 + \Phi w_\psi(\nabla(\Delta\psi))^2 + 2\lambda M(\Delta\phi, \Delta\psi) + 2\lambda^*\frac{(\Delta\phi)}{m^2}\right\}dxdy$$

where λ is a variable Lagrange multiplier and λ^* is a constant multiplier.

This variational problem can be solved either by an optimization algorithm (see Le Dimet and Talagrand 1985; Navon and Legler 1987, for example) or with Euler–Lagrange equations. We chose the latter because the linear equations to be solved are rather simple, and to show the connection with unconstrained initialization. Since we use as a reference, the limited-area finite element barotropic model of Staniforth and Mitchell (1977, 1978), a wall boundary condition restricts the streamfunction on Γ to zero,

$$\psi = 0 \quad \text{on } \Gamma.$$

Thus, variations must be performed with

$$\Delta\psi = 0 \quad \text{on } \Gamma.$$

Taking into account this latter boundary condition, together with the boundary conditions specified by (2.11) when making variations on the limited domain D, this forces a natural boundary condition of a specific form for the variable multiplier λ. The reader will find the essentials of the required manipulations in Smirnov's (1964) book (chapter II).

It is thus clear that the following variational set of equations have been formulated to be fully compatible with the numerical model being initialized.

The Euler–Lagrange equations are as follows:

$$\nabla^2\lambda = \frac{w_\phi}{m^2}(\Delta\phi) + \frac{\lambda^*}{m^2}$$

$$\frac{\partial\lambda}{\partial n} = 0 \quad \text{on } \Gamma, \tag{2.13}$$

$$\nabla\cdot(\Phi w_\psi\nabla(\Delta\psi)) = \mathcal{F}(\lambda)$$

$$\Delta\psi = 0 \quad \text{on } \Gamma, \tag{2.14}$$

$$\nabla^2(\Delta\phi) = \mathcal{F}(\Delta\psi) + \mathcal{B}(\Delta\chi)_G + (\delta_t\nabla^2\chi)_o$$

$$\frac{\partial}{\partial x}(\Delta\phi) = f\frac{\partial}{\partial x}(\Delta\psi) + f\frac{\partial}{\partial y}(\Delta\chi)_G \text{ on } \Gamma_y,$$

and

$$\frac{\partial}{\partial y}(\Delta\phi) = f\frac{\partial}{\partial y}(\Delta\psi) - f\frac{\partial}{\partial x}(\Delta\chi)_G \text{ on } \Gamma_x \tag{2.15}$$

$$\iint_D \frac{(\Delta\phi)}{m^2}\,dxdy = 0. \tag{2.16}$$

c. *Unconstrained implicit NMI*

An immediate consequence of this set of equations is that if we set $w_\phi = w_\psi = \text{const}$, then (2.13) and (2.16) together with the consistent solving of (2.13) imposes $\lambda^* = 0$ and we may write in symbolic form

$$\lambda = w_\phi\nabla_n^{-2}\left(\frac{\Delta\phi}{m^2}\right)$$

where ∇_n^{-2} is the linear operator which inverts the Laplacian with homogeneous Neumann boundary conditions.

Equation (2.14) gives

$$\Delta\psi = \frac{w_\phi}{\Phi w_\psi}\nabla_d^{-2}\mathcal{F}\nabla_n^{-2}\left(\frac{\Delta\phi}{m^2}\right) = \frac{1}{\Phi}\nabla_d^{-2}\mathcal{F}\nabla_n^{-2}\left(\frac{\Delta\phi}{m^2}\right) \quad (2.17)$$

where ∇_d^{-2} is the linear operator which inverts the Laplacian with homogeneous Dirichlet boundary conditions. Inserting this result into (2.15) gives

$$\nabla^2(\Delta\phi) - \frac{1}{\Phi}\mathcal{F}\nabla_d^{-2}\mathcal{F}\nabla_n^{-2}\left(\frac{\Delta\phi}{m^2}\right) = \mathcal{B}(\Delta\chi)_G + (\delta_t\nabla^2\chi)_o. \quad (2.18)$$

Equations (2.17) and (2.18) are exactly the same as those derived by T(88) for unconstrained initialization. Since we also use $\Delta\chi = (\Delta\chi)_G$, this special form of the variational problem reduces to unconstrained initialization (with Machenhauer's scheme), a result found previously by Daley (1978) and Temperton (1984) in different contexts.

This fact has also been used by Ballish (1980, chapter IV) for closing his system of equations in the context of the bounded derivative initialization method. Ignoring beta effects, it is easily shown that Ballish's scheme of non-normal mode initialization is a special case of implicit nonlinear NMI considered by Temperton (1989). As a consequence, the present variational scheme encompasses the variational form of non-normal mode initialization proposed by Ballish (1980, pp. 54–55) which has not been exploited (surprisingly enough) since its early formulation.

d. The numerical procedure

The numerical method used to solve the Euler–Lagrange equations (2.13)–(2.16) is very similar to that used in T(88) for the unconstrained case. The equations may be combined into a single equation for $\Delta\phi$:

$$\nabla^2(\Delta\phi) - \frac{1}{\Phi}\mathcal{F}\mathcal{L}_d^{-1}\mathcal{F}\nabla_n^{-2}\left(\frac{w_\phi}{m^2}\Delta\phi + \frac{\lambda^*}{m^2}\right) = \mathcal{B}(\Delta\chi) + (\delta_t\nabla^2\chi)_o \quad (2.19)$$

where the linear (but variable-coefficient) operator $\mathcal{L}$ is defined by

$$\mathcal{L} = \nabla\cdot w_\psi\nabla.$$

In (2.19), the operator $\mathcal{L}$ is to be inverted using homogeneous Dirichlet boundary conditions.

To solve (2.19), we use the algorithm of Concus and Golub (1973), accelerated by a conjugate-gradient technique (Concus et al. 1976). For the kernel of the Concus–Golub scheme we take the constant-coefficient Helmholtz operator $\nabla^2 - \lambda_0$, where λ_0 is a suitably chosen average value of

$$\frac{1}{\Phi}\frac{f^2 w_\phi}{m^2 w_\psi}.$$

Compared with the unconstrained case, the only additional problem is the inversion of the operator $\mathcal{L}$. This was achieved by an inner iteration, itself based on the Concus–Golub scheme with conjugate-gradient acceleration.

The whole variational algorithm may be summarized as follows for one iteration:

(i) Run part of the unconstrained initialization scheme to obtain $(\Delta\chi)_G$. This step is identical to step 1 and 2 of T(88) and does not interact with the $(\Delta\phi)$ and $(\Delta\psi)$ fields.

(ii) Given $(\Delta\chi)_G$, $(\delta_t\nabla^2\chi)$. Solve (2.13) to (2.16) to get $(\Delta\phi)$ and $(\Delta\psi)$. (This step replaces step (3) and (4) of T(88).) The next steps are identical to T(88) (steps 5 and 6) which are

(iii) Find (ΔU), (ΔV) from $(\Delta\psi)$, $(\Delta\chi)$.

(iv) Add increments (ΔU), (ΔV), $(\Delta\phi)$ to model fields.

3. Experimental results

In order to test the variational algorithm, we illustrate in this section an implementation into the regional finite element model of Staniforth and Mitchell (1977, 1978). A number of experiments were run to examine the effects of varying the weighting functions w_ϕ, w_ψ.

In the present case, the domain of initialization was chosen to be a square of side 20 000 km centered at the North Pole using a polar stereographic projection true at 60°N. This domain is illustrated in Fig. 1. The boundary is a solid wall in the vicinity of the equator. The grid has a uniform resolution of 100 km in the region of interest (61 × 61 points) and relaxes uniformly outside that region until the whole grid gets completed by a set of 101 × 101 points.

As compared to previous experiments on variational initialization for shallow-water models, the problem faced here can be considered a fairly large one, since we deal with 30 000 variables to be adjusted for the streamfunction, velocity potential and geopotential fields.

All experiments were run with the same initial conditions as those used by T(88). The mean depth for the uninitialized height field shown in Fig. 2 was 5600 m. Unless otherwise stated, all other features of the model to be prescribed before starting the integration are implicitly assumed to be the same as those specified in T(88).

For each experiment to be presented below, we look at the behavior of the three following values for both unconstrained and variational schemes:

DOMAIN AND NONUNIFORM GRID

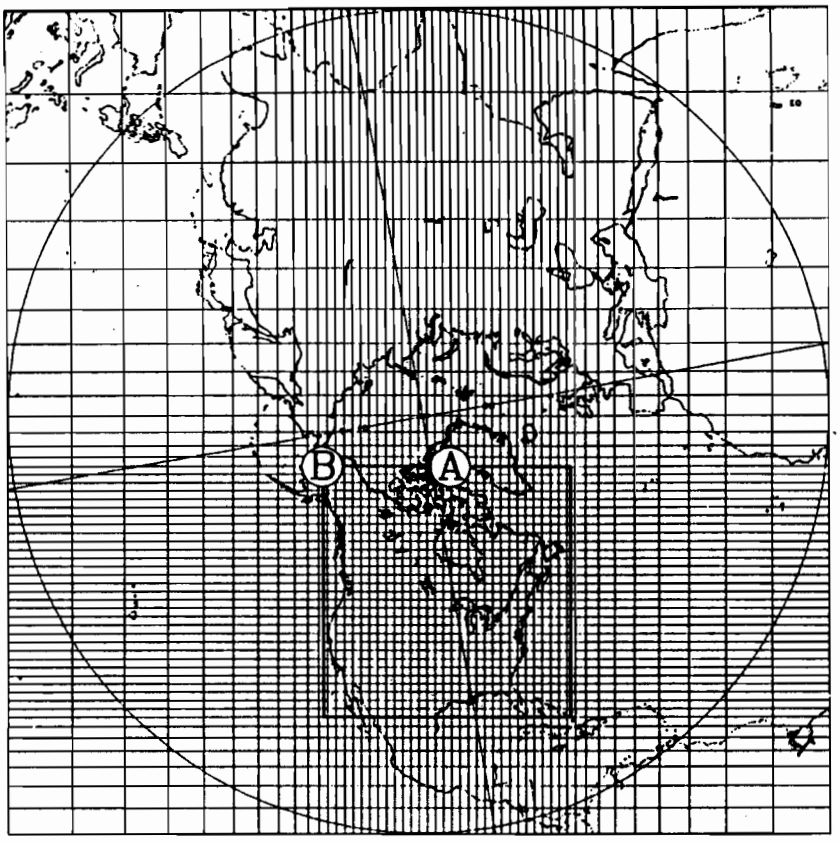


FIG. 1. The experimental domain and the nonuniform grid. Time traces in height field given by Figs. 8 and 10 are associated with gridpoints denoted A and B respectively.

$$J = \iint_D \frac{w_\phi}{m^2}(\Delta\phi)^2 + \Phi w_\psi(\nabla(\Delta\psi)^2)dxdy$$

where $\Delta\phi$, $\Delta\psi$ are the instantaneous correction fields computed at a given iteration for both schemes;

$$J_T = \iint_D \frac{w_\phi}{m^2}(\Delta\phi)_T^2 + \Phi w_\psi(\nabla(\Delta\psi)_T)^2 dxdy$$

where the index "T" refers to the total correction applied to the uninitialized original fields at a given iteration for both schemes; and

$$\mathrm{BAL}_1 = \iint_D \left\{\frac{(\delta_t\phi)_G^2}{m^2} + \Phi[(\delta_t U)_G^2 + (\delta_t V)_G^2]\right\} dxdy.$$

This value can be computed entirely in physical space and is a measure of the convergence of the unconstrained (Machenhauer) scheme. The reader will find the details of the computation of BAL_1 in T(88).

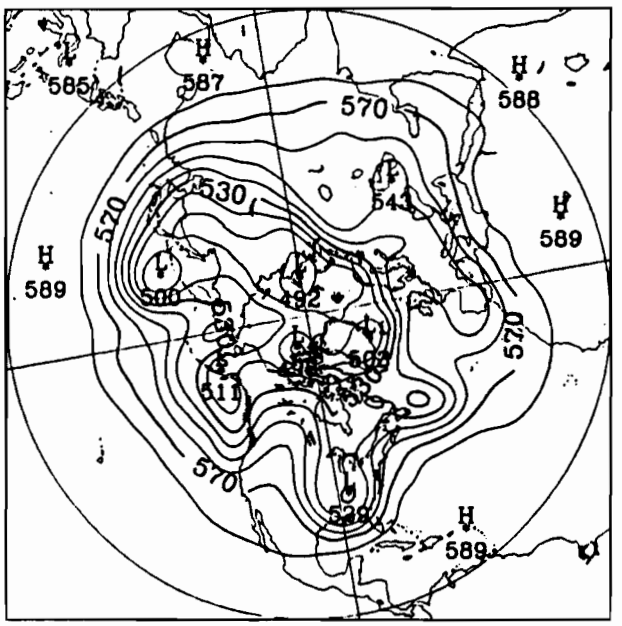


FIG. 2. The uninitialized height field (contour interval 10 dam).

In the present context, BAL_I will simply serve as a measure of gravity wave activity and will be given for comparison of unconstrained versus constrained initialization schemes. Note that the way that our variational problem has been formulated, we have no guarantee a priori that J_T will be smaller for the variationally adjusted field when compared to its value in the unconstrained scheme. Each scheme follows a different path as they attempt to reach a dynamical balance.

a. *Latitudinally varying weights*

This first experiment concerns latitudinally varying weights w_ϕ, w_ψ. The structure of these weights is taken from Daley (1978) and takes the simple form for a latitude θ

$$w_\psi = (1 - \mu^2)^4 \quad \text{where} \quad \mu = \sin\theta$$

$$w_\phi = 1 - w_\psi.$$

Once projected onto the polar-stereographic grid these weights are variable in both "x" and "y" directions. Figure 3 shows w_ϕ on the calculation grid. From the chosen form of the function $J(2.6)$, the variational scheme will tend to adjust the rotational part of the wind field to the initial mass field at higher latitudes and conversely at lower latitudes.

In the following, changes due to unconstrained or constrained initialization are represented by subtracting the uninitialized fields from the adjusted fields. Figures 4 and 5 show these changes for the height and streamfunction fields for the case of unconstrained initialization. Comparing these to the results obtained by variational adjustment (given by Figs. 6 and 7 for $\Delta\phi$ and $\Delta\psi$ respectively), it is apparent that the major differences are latitudinal in character. The increasing weight on mass with latitude strongly forces the retention of mass at higher latitudes. For both unconstrained and variational schemes the structure of the height field corrections are somewhat similar although with different amplitude at lower latitudes.

FIG. 4. Difference in height field between unconstrained initialization and no initialization (contour interval 7 m).

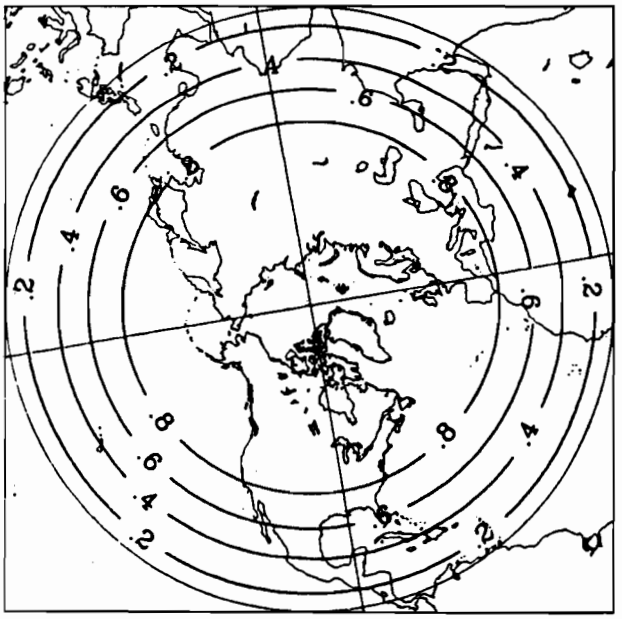


FIG. 3. Spatial structure of weighting function w_ϕ for experiment in section 3a.

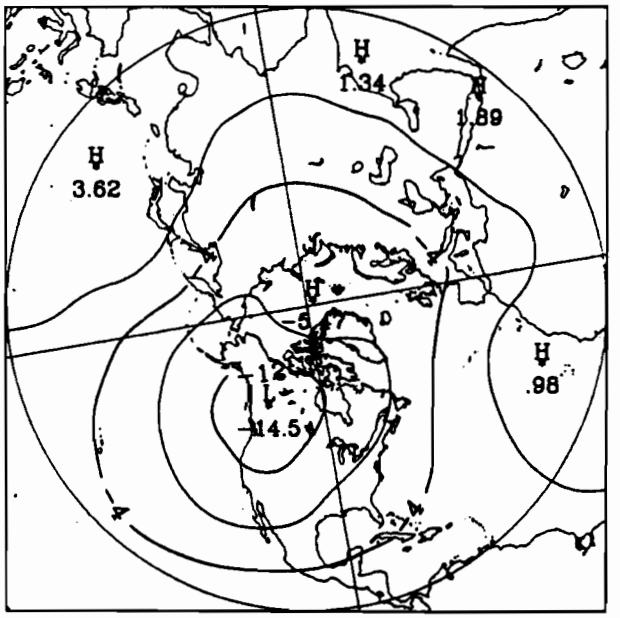


FIG. 5. Difference in streamfunction field between unconstrained initialization and no initialization (contour interval $4 \times 10^5\ m^2\ s^{-1}$).

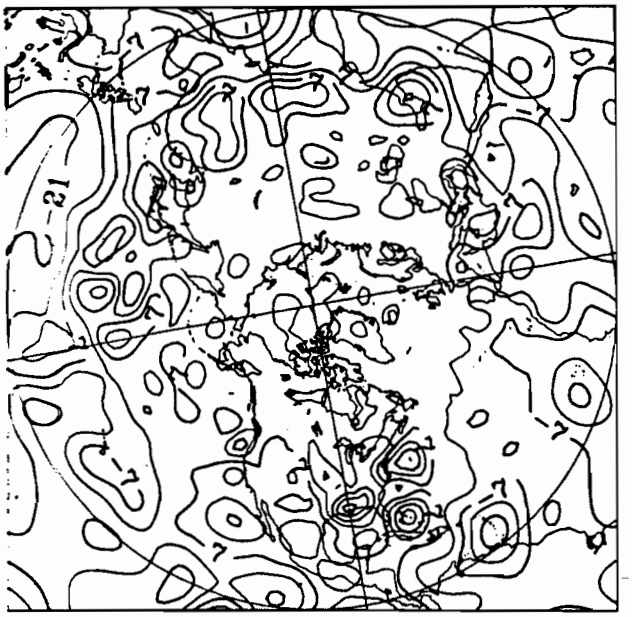


FIG. 6. Difference in height field between variational adjustment using latitudinal weights w_ϕ, w_ψ and no initialization (contour interval 7 m).

Changes made to the streamfunction field are completely different for the variational and unconstrained cases. The variational scheme presents a correction field which acts in the desired way as regards mutual adjustment with the mass field previously described. On the other hand, unconstrained initialization has produced a smoothly varying correction field distributed over D.

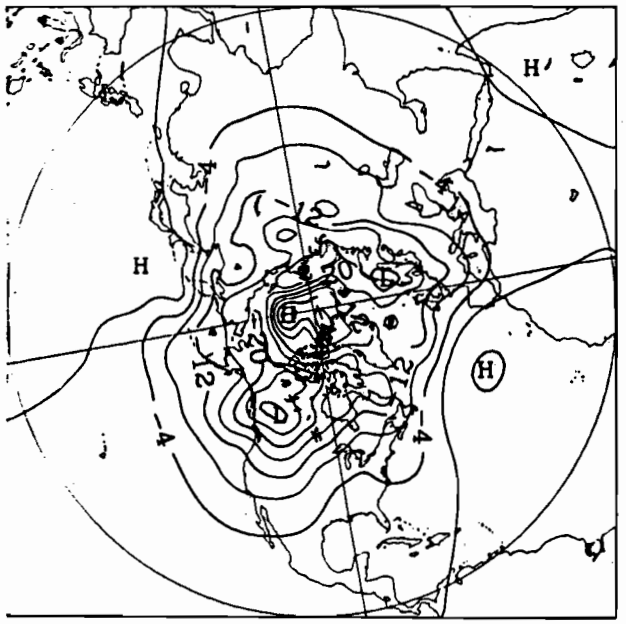


FIG. 7. Difference in streamfunction field between variational adjustment using latitudinal weights w_ϕ, w_ψ and no initialization (contour interval 4×10^5 m^2 s^{-1}).

TABLE 1. Daley's weights.

Iteration	Value of functional J	Value of functional J_T	BAL_1
	Unconstrained		
0	—	—	2.8×10^{11}
1	1.136×10^{18}	1.136×10^{18}	2.0×10^{9}
2	1.282×10^{15}	1.144×10^{18}	7.7×10^{7}
	Variational		
0	—	—	2.8×10^{11}
1	4.264×10^{17}	4.264×10^{17}	1.0×10^{10}
2	1.086×10^{17}	6.016×10^{17}	2.0×10^{9}
3	1.161×10^{16}	5.975×10^{17}	1.5×10^{8}
4	2.986×10^{13}	5.979×10^{17}	7.2×10^{7}

Table 1 gives the behavior of J, J_T and BAL_1 at each iteration of the unconstrained and variational schemes. Note that in the unconstrained case, J and J_T are purely diagnostic quantities computed for the sake of comparison with the variational experiment. The solution of the unconstrained problem is of course totally independent of w_ϕ and w_ψ. It is seen that the value of BAL_1 decreases more rapidly than the variational case as a function of iteration. The convergence of the variational scheme can be judged by the gradual decrease of J and a stabilizing value of J_T as a function of iteration. Note that at the final stage of the initialization process, the value of J_T for the variational case is effectively lower than that obtained with the unconstrained scheme.

It is worth mentioning that in the present context of a shallow-water model, it has been observed that iterating Machenhauer's scheme more than twice (starting with the uninitialized fields) gives no improvement on the balance of the resulting fields [see also T(88), p. 1025]. Further reference to unconstrained initialization will always refer though to two iterations of this scheme. For both variational and unconstrained experiments, we present in Fig. 8 the time trace of the height field for point (A) of Fig. 1. The model was integrated for 48 hours. Initial conditions from the variational scheme are obtained using three iterations. For the uninitialized case, high frequency gravity-inertia waves are present at the initial stage of integration of the model (heavy line) with large initial amplitudes that diminish as the integration proceeds. In both nonlinearly adjusted initial-condition cases, the degree of balance as judged from the smoothness of the associated height traces are comparable. Although the initial states (in the height field) of the model at this particular grid point are separated by a small difference, the variational scheme has clearly changed the slow Rossby modes present initially and forced a significant departure from its unconstrained analogue.

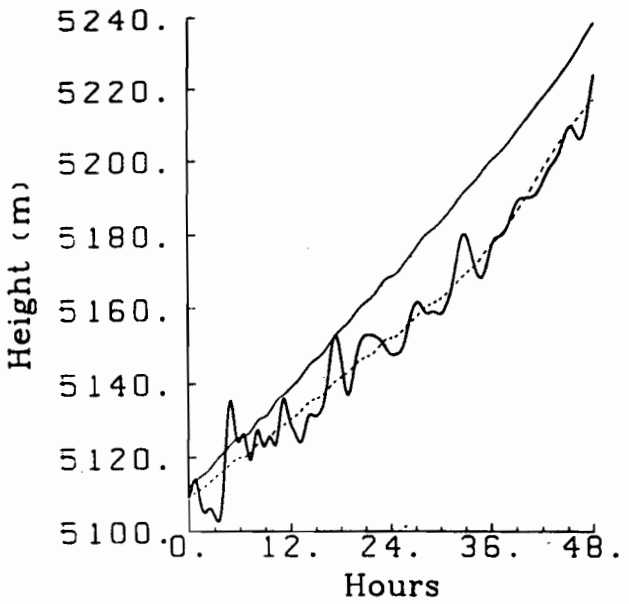


FIG. 8. Time trace of height field for point A shown in Fig. 1. Heavy line—no initialization, dashed line—unconstrained initialization (2 iterations), light line—variational initialization using latitudinal weights (3 iterations).

b. Latitude–longitude varying weights

We now describe another experiment dealing with both latitude and longitude variability of the weighting functions w_ϕ, w_ψ. Clearly this longitudinal variability is crucial for discriminating between the reliability of oceanic and continental analysis data within the initialization process. As mentioned previously, the optimal determination of these weights can be obtained in conjunction with the objective analysis scheme.

For illustrative purposes however, we will examine the behavior of the variational scheme under the following longitudinal structure. We simply cut the domain D by straight lines originating from the North Pole so as to isolate oceans (an abrupt version of Tribbia's, 1982, longitudinal structure). The lines are located by the following angles and with them the associated ocean domains:

Atlantic Ocean (15°W, 65°W).
Pacific Ocean (125°W, 135°E).

For comparison, the latitudinal structure of w_ϕ, w_ψ are given the same form as Daley's weights discussed in the preceding experiment.

Over both oceanic sectors, a reducing factor of 10^{-2} was applied to w_ϕ and a factor of 10^{-1} to w_ψ.

As we have seen previously, over the continents (in the extratropical regions) the pure latitudinal structure of w_ϕ, w_ψ will force the rotational part of the wind field to adjust strongly to the height. In the present case, the chosen reducing factors on both weights simulate the effect of higher confidence in wind field data relative to the mass field over oceanic regions. Consequently, as one goes from continents to oceans, the wind field becomes less strongly forced to adjust to the original mass field.

Subtracting the adjusted height field obtained from Daley's weights from the height field obtained with the present set of weights (after three iterations), the longitudinal effect is clearly seen in Fig. 9 which displays the resulting height-field changes. The sharp longitudinal delimitations are clearly represented here and a more elongated north–south pattern could be obtained by a judicious choice of the weighting functions.

Table 2 summarizes the behavior of the variational scheme for the present set of weights. Values of J and J_T have also been recomputed for the unconstrained initialization scheme using the new set of weights. Once again, the value of J_T for the variational case is less that of the unconstrained case, and three iterations are found to be sufficient for providing an acceptable degree of balance. To show this we choose point (B) on the calculation grid of Fig. 1 to monitor the height trace during integration of the model. This point lies precisely where most of the difference in the height field is observed relative to the first experiment.

Heavy, light, and dashed lines shown in Fig. 10 are associated once again with uninitialized, variationally adjusted, and unconstrained initial conditions. Judging from the light curve, the addition of the longitudinal effect in the weighting functions has not inhibited the ability of the scheme to achieve a dynamical balance initially, while still taking into account the hypothetical oceanic weaknesses of the initial data given by the objective analysis.

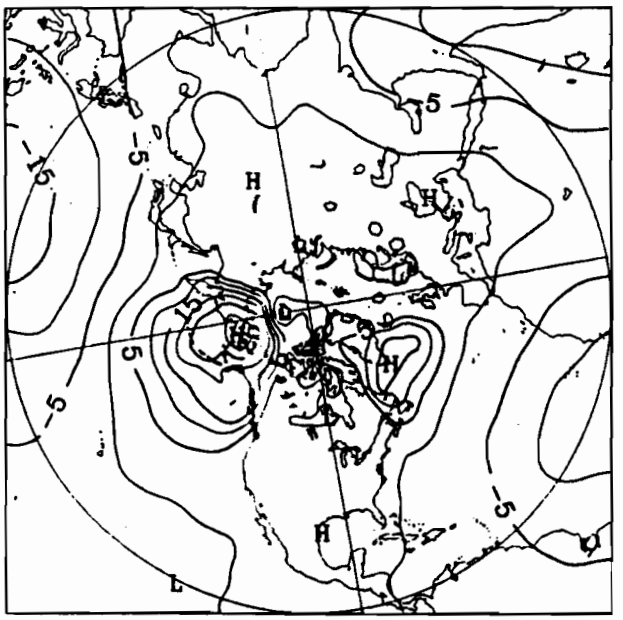


FIG. 9. Difference in height field between variational initialization using latitude–longitude weights and variational initialization using latitudinal weights (contour interval 5 m).

TABLE 2. Latitude-Longitude weights.

Iteration	Value of functional J	Value of functional J_T	BAL_I
	Unconstrained		
0	—	—	2.8×10^{11}
1	7.358×10^{17}	7.358×10^{17}	2.0×10^{9}
2	1.563×10^{13}	7.439×10^{17}	7.7×10^{7}
	Variational		
0	—	—	2.8×10^{11}
1	3.809×10^{17}	3.809×10^{17}	8.0×10^{9}
2	2.238×10^{15}	3.758×10^{17}	5.8×10^{8}
3	3.308×10^{14}	3.831×10^{17}	3.2×10^{8}
4	1.584×10^{14}	3.811×10^{17}	2.7×10^{8}

Finally, an interesting point to consider is whether this variational scheme has the potential ability to significantly change a model forecast as compared to standard unconstrained initialization. To consider this point, a two-day integration of the model using unconstrained initialization served as a reference for comparison. Using the latter specification of the weighting functions w_ϕ, w_ψ, the variational initialization scheme (using three iterations) provided the initial conditions for a second forecast. Fig. 11 presents the initial height departure at intervals of 5 m with a maximum of 22 m. The oceanic departures are clearly represented and the emphasis is on the possible amplification of this pattern as time goes on. After running the model for two days, the resulting pattern of height-field departures is given in Fig. 12. Differences are given at intervals of 10 m and in particular there has been a steady increase in the height field difference between the two forecasts associated with the southward movement of the low center (see Fig. 2), reaching a maximum absolute value of 67 m. We might thus expect that for initial departures in the initial data which ex-

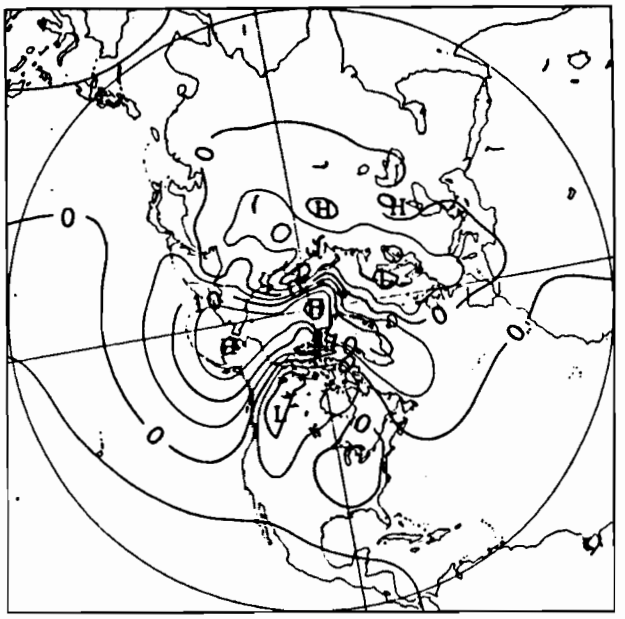


FIG. 11. Difference in height field between variational initialization using Latitude-Longitude weights and unconstrained initialization. Day 0, contour interval 5 m, maximum amplitude 22 m.

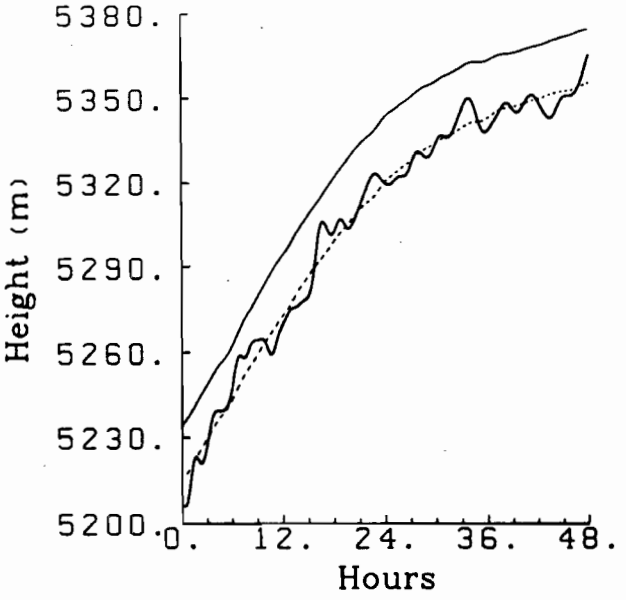


FIG. 10. Time trace of height field for point B shown in Fig. 1. Heavy line—no initialization, dashed line—unconstrained initialization, light line—variational initialization using latitude-longitude weights.

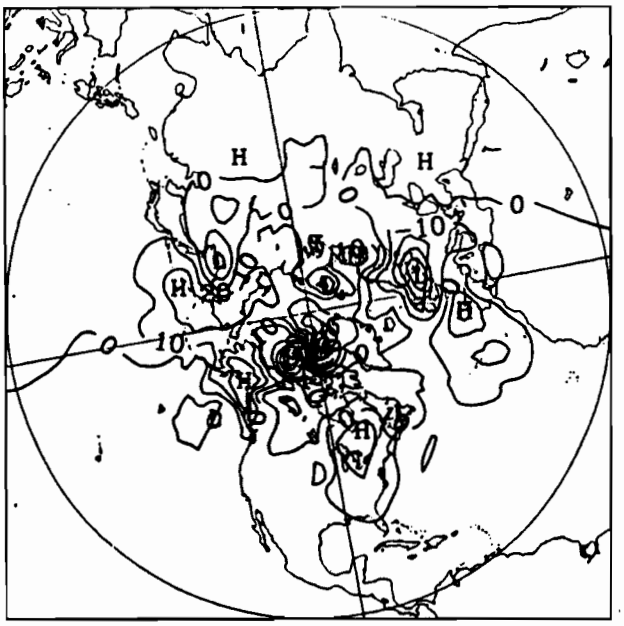


FIG. 12. As in Fig. 11, but after two days of model integration. Contour interval 10 m, maximum amplitude 67 m.

ceed those used here (generated by using suitable weights w_ϕ, w_ψ with respect to expected analysis error) we would obtain even larger differences in forecasts as compared to standard unconstrained initialization.

4. Conclusions

In this paper, we have shown that the implicit normal mode initialization technique can be generalized by combining it with a variational procedure. The resulting fields satisfy the balance condition required by nonlinear normal-mode initialization, while the changes to the analyzed fields are controlled through minimization of a variational integral. Since the whole procedure is expressed in physical space, the weights in this integral can be varied in the horizontal in a more or less arbitrary manner, to reflect local variations in the presumed reliability of the mass and wind field analyses. Some of the difficulties encountered in previous efforts to combine normal-mode initialization with variational techniques are thereby avoided.

We have also demonstrated the potential of variational normal-mode initialization to change the slowly evolving component of the resulting forecast. The next stage of this study will clearly be to extend the technique to a multilevel model, and to test it in the context of a full-fledged data assimilation scheme.

Acknowledgments. The authors wish to thank Dr. Andrew Staniforth who contributed a model and an internal review of this paper. We also gratefully acknowledge Diane Lespérance for typing the manuscript. The first author received support from NSERC.

REFERENCES

Ballish, B., 1980: Initialization, theory and application to the NMC spectral model. Ph.D. thesis, University of Maryland, 151 pp. [Dept. of Meteorology, University of Maryland, College Park, MD 20742.]

Concus, P., and G. H. Golub, 1973: Use of fast direct methods for the efficient numerical solution of non-separable elliptic equations, *SIAM J. Numer. Anal.*, **10**, 1103–1120.

——, ——, and D. P. O'Leary, 1976: A generalized conjugate gradient method for the numerical solution of elliptic partial differential equations, *Sparse Matrix Computations*, J. R. Bunch and D. J. Rose, Eds. Academic Press, 309–332.

Daley, R., 1978: Variational nonlinear normal mode initialization. *Tellus*, **30**, 201–218.

——, and K. Puri, 1980: Four-dimensional data assimilation and the slow manifold. *Mon. Wea. Rev.*, **108**, 85–99.

Hollingsworth, A., D. B. Shaw, P. Lönnberg, L. Illari, K. Arpe and A. J. Simmons, 1986: Monitoring of observation and analysis quality by a data assimilation system. *Mon. Wea. Rev.*, **114**, 861–879.

Le-Dimet, F. X., and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus*, **38A**, 97–110.

Machenhaueur, B., 1977: On the dynamics of gravity oscillations in a shallow water model, with application to normal mode initialization. *Contrib. Atmos. Phys.*, **50**, 253–271.

Navon, I. M., and D. M. Legler, 1987: Conjugate-gradient methods for large scale minimization in meteorology. *Mon. Wea. Rev.*, **115**, 1479–1502.

Puri, K., 1983: Some experiments in variational normal mode initialization in data assimilation. *Mon. Wea. Rev.*, **111**, 1208–1218.

Smirnov, V. I., 1964: *A Course of Higher Mathematics, Vol. IV*, 4th ed. Pergamon Press, 811 pp.

Staniforth, A. N., and H. L. Mitchell, 1977: A semi-implicit finite-element barotropic model. *Mon. Wea. Rev.*, **105**, 154–169.

——, and ——, 1978: A variable-resolution finite-element technique for regional forecasting with the primitive equations. *Mon. Wea. Rev.*, **106**, 439–447.

Temperton, C., 1984: Variational normal mode initialization for a multilevel model. *Mon. Wea. Rev.*, **112**, 2303–2316.

——, 1988: Implicit normal mode initialization. *Mon. Wea. Rev.*, **116**, 1013–1031.

——, 1989: Implicit normal mode initialization for spectral models. *Mon. Wea. Rev.*, **117**, 432–447.

Tribbia, J. J., 1981: Nonlinear normal-mode balancing and the ellipticity condition. *Mon. Wea. Rev.*, **109**, 1751–1761.

——, 1982: On variational normal mode initialization. *Mon. Wea. Rev.*, **110**, 455–470.

Chapter IV

Variational implicit normal mode initialization for a global spectral shallow-water model

IV.1 Presentation of article 2.

Recently, Temperton (1989) showed that the concepts of implicit nonlinear NMI could be of some use even in the context of spectral models. The degree of departure of the implicit technique with respect to the standard "explicit" nonlinear NMI technique is more easily examined in the spectral context due to the separability of the linear system for each zonal wavenumber (Kasahara 1976). It turns out that for a mean equivalent depth of 5600 m. (i.e. between the external and first internal depth values), the differences between the "de luxe" version of the implicit NMI scheme (i.e. scheme B of Temperton 1989) and the conventional explicit NMI scheme, are limited to the very largest horizontal scales. Further studies on this point are yet to come. There is no doubt however that for the next generation of *variable resolution models*, the implicit nonlinear NMI as described in Temperton (1989) will impose itself as much as the implicit NMI for regional models.

The usefulness of VINMI for spectral models is however highly interesting even for the current applications of this model. The variational form of nonlinear NMI was first considered by Daley (1978) in the spectral context. The size of the nonseparable linear system involved in his algorithm was a limiting factor for its operational use in a multilevel context.

We present in the following, the application of VINMI to a global spectral shallow-water model. Similar tests as those considered in chapter III indicate the robustness of the scheme when operated at triangular truncation 63.

IV.2 Article 2

Variational implicit normal mode initialization on the sphere

Monthly Weather Review , 119, 631-652

Variational Implicit Normal Mode Initialization on the Sphere

LUC FILLION

Recherche en prévision numérique, Atmospheric Environment Service, Dorval, Québec, Canada

(Manuscript received 19 January 1990, in final form 4 September 1990)

ABSTRACT

Due to it's prohibitive computational cost, variational nonlinear normal mode initialization as first suggested by Daley has received little interest during the last 10 years. Recently, soon after the introduction of the framework now called *implicit nonlinear normal mode initialization*, an efficient reformulation of variational nonlinear normal mode initialization using the implicit technique was demonstrated by Fillion and Temperton in the context of a barotropic finite-element regional model. This scheme allowed full variation of the variational weights at a low computational cost. To complement this previous work, the same variational approach for a global spectral shallow-water model is presented here. Similar results regarding the controlling and balancing aspects of the scheme are illustrated. The special form taken by the variational scheme in the context of *height-constrained* initialization is reconsidered after establishing the relationship between the implicit schemes and quasi-geostrophic theory on an f plane. Possible extensions of the method are mentioned at the end of the paper.

1. Introduction

The nonlinear normal mode initialization technique (nonlinear NMI) as a mean of controlling the amplitude of the high-frequency gravity-inertia waves in a primitive equation model has benefited from several improvements during the last decade. Stimulated by the development of an initialization method in physical space called the "bounded derivative initialization method" (BDI) due to Browning et al. (1980) (based on the bounded derivative principle of Kreiss 1979, 1980), a physical space formulation of nonlinear NMI has gradually emerged. Ballish's (1980) results were gradually set into a rigorous framework during the 1980s and culminated in what is now called "implicit nonlinear NMI" (Temperton 1988, 1989) or "non-normal mode initialization" (Ballish 1980; Juvanon du Vachat 1988). As in Fillion and Temperton (1989) (hereafter referred to as FT89) this new method will be referred to as *Implicit nonlinear normal mode initialization technique* (INMI). Implicit NMI schemes are formulated in physical space but are still characterized by slow and fast mode components of the flow. This characterization turns out to be important for the formulation of a variational INMI (VINMI) in the spirit of Daley's work (1978), and has recently been exploited by FT89 in the context of a barotropic finite-element regional model. Their variational procedure was shown to efficiently minimize changes to the analyzed fields while at the same time maintaining a degree of balance comparable to the unconstrained implicit scheme. Based on Kasahara's work (1982a), it is expected that a variational extension of the BDI method would be very similar to VINMI.

Although implicit nonlinear normal mode initialization was largely stimulated by the desire to apply the normal mode technique to limited area models, it is also recognized to be a more useful approach than the standard explicit technique for initializing very high resolution spectral models. This paper presents what may prove to be a still greater benefit of applying the implicit initialization technique to global spectral models; that is, the feasibility of variationally initializing such models. Indeed, to date the problem of variational normal mode initialization when formulated in normal mode space has been recognized to be a problem of gargantuan size and consequently not viable for use in an operational data assimilation scheme. It is the purpose of this paper to show that a variational extension of INMI is feasible at a reasonable cost with present day computers in the context of a global spectral shallow-water model. The technique to be presented is based on the formulation of VINMI by FT89 and incorporates full horizontal variations of the weights.

Section 2 gives a setting of the dynamical model together with a quick review of the "implicit scheme B" developed by Temperton (1989). The variational form of INMI is presented in section 3 and results of its application are discussed in section 4. Section 5 briefly addresses the relationship between INMI and quasi-geostrophic theory (for an f-plane model) which is related to the problem of the divergence of the *height-*

Corresponding author address: Dr. Luc Fillion, Recherche en prévision numérique, Atmospheric Environment Service, 2121 Trans-Canada Highway, Dorval, Quebec H9P 1J2, Canada.

constrained normal mode initialization iterative scheme. Finally, section 6 gives a summary and possible future extensions of the method.

2. Implicit nonlinear normal mode initialization

In order to describe the variational scheme, the following section presents the essential ingredients that are needed. Following FT89, these are the models to be initialized, and the differential constraints characterizing the specific implicit NMI scheme considered in this paper.

a. The model

The governing equations are the shallow-water equations on the sphere, in their differentiated form as used by Temperton (1989) (hereafter referred to as T89) and Daley (1978):

$$\frac{\partial \zeta}{\partial t} = -2\Omega \sin\theta D - \frac{2\Omega}{a^2}\left(\frac{\partial \psi}{\partial \lambda} + \cos\theta \frac{\partial \chi}{\partial \theta}\right) + Q_\zeta \quad (2.1)$$

$$\frac{\partial D}{\partial t} = 2\Omega \sin\theta \zeta - \frac{2\Omega}{a^2}\left(-\cos\theta \frac{\partial \psi}{\partial \theta} + \frac{\partial \chi}{\partial \lambda}\right) - \nabla^2 \phi + Q_D \quad (2.2)$$

$$\frac{\partial \phi}{\partial t} = -\Phi D + Q_\phi \quad (2.3)$$

the quantities Q_ζ, Q_D, Q_ϕ are the nonlinear (advective and metric) terms. The symbol Φ refers to the mean geopotential depth of the original analyzed field, and the symbol ϕ is a geopotential departure from Φ. The other symbols have their standard meaning. Setting Q_ζ, Q_D, Q_ϕ to zero obtains the linear spectral form of the previous system by using a spherical harmonic expansion of the dependent variables, where for example,

$$\zeta(\lambda, \theta, t) = \sum_{m=-M}^{M} \sum_{n=|m|}^{M} \zeta_n^m(t)\, P_n^m(\sin\theta) \exp(im\lambda)$$

for triangular truncation, and

m = zonal (east–west) wavenumber
P_n^m = associated Legendre function of the first kind of order m and degree n
ζ_n^m = spectral component of the vorticity field.

The final system can be put in standard symmetric form by introducing the following change of variables

$$\hat{\zeta}_n^m = a[n(n+1)]^{-1/2}\zeta_n^m \quad (2.4)$$

$$\hat{D}_n^m = i[a[n(n+1)]^{-1/2}D_n^m \quad (2.5)$$

$$\hat{\phi} = \Phi^{-1/2}\phi_n^m \quad (2.6)$$

which gives

$$\frac{d\hat{\zeta}_n^m}{dt} = i(b_n^m \hat{\zeta}_n^m + f_n^m \hat{D}_{n-1}^m + f_{n+1}^m \hat{D}_{n+1}^m) \quad (2.7)$$

$$\frac{d\hat{D}_n^m}{dt} = i(b_n^m \hat{D}_n^m + f_n^m \hat{\zeta}_{n-1}^m + f_{n+1}^m \hat{\zeta}_{n+1}^m + c_n^m \hat{\phi}_n^m) \quad (2.8)$$

$$\frac{d\hat{\phi}_n^m}{dt} = ic_n^m \hat{D}_n^m \quad (2.9)$$

where

$$b_n^m = \frac{2\Omega m}{n(n+1)} \quad (2.10)$$

$$f_n^m = \frac{2\Omega}{n}(n^2-1)^{1/2}\epsilon_n^m \quad (2.11)$$

$$c_n^m = \frac{\Phi^{1/2}}{a}[n(n+1)]^{1/2} \quad (2.12)$$

and

$$\epsilon_n^m = \left(\frac{n^2-m^2}{4n^2-1}\right)^{1/2}.$$

In matrix form, this system is written as

$$\frac{d\mathbf{X}}{dt} = i\mathbf{A}\mathbf{X}$$

where the state vector $\mathbf{X}$ for a given zonal wavenumber m is

$$\mathbf{X} = \begin{bmatrix} \hat{\zeta} \\ \hat{\mathbf{D}} \\ \hat{\phi} \end{bmatrix}$$

and each component of the state vector contains its spectral components; for example

$$\hat{\zeta} = (\zeta_m^m, \cdots, \zeta_M^m)^T.$$

The preceding derivation is common to both explicit and implicit nonlinear NMI schemes but they depart from each other at this point.

b. INMI constraints

Recently, T89 has examined various possible perturbations of the matrix $\mathbf{A}$ (which defines the normal modes of the system) all having the effect of producing stationary nondivergent Rossby modes. His conclusion about these possible variants was that the "*B* scheme" was the most natural to choose since it is based on the least drastic modification of the linear system and is closest to explicit NMI. As a consequence, the present study is based on the variational extension of T89's "scheme *B*." Given here is a resumé of the essential elements of this scheme and the interested reader is referred to T89 for further details and basic terminology.

In the context of Machenhauer's (1977) initialization, the following algebraic system must be solved for the correction state vector $\Delta\mathbf{X}$,

$$\mathbf{A}(\Delta\mathbf{X}) = i\mathbf{E}_G\mathbf{E}_G^T(\delta_t\mathbf{X})_0 \quad (2.13)$$

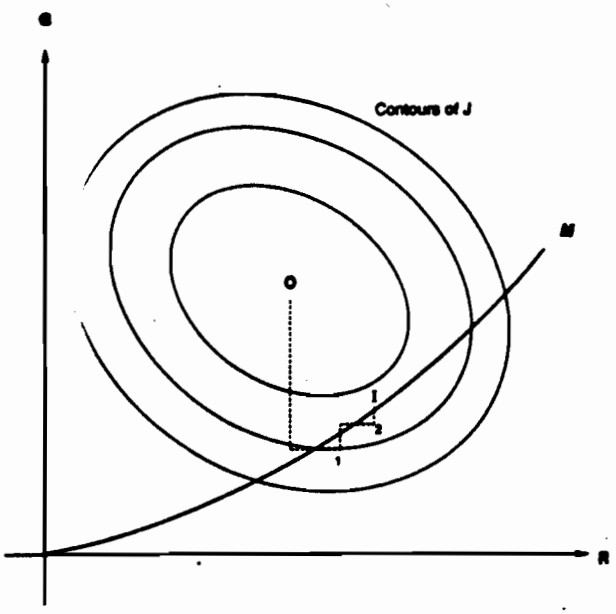


FIG. 1. Variational initialization in phase space.

where $\mathbf{E}_G$ is the matrix whose columns are the eigenvectors of $\mathbf{A}$ associated with a chosen set of gravity modes to be initialized. These matrices act on $(\delta_t \mathbf{X})_0$ which is the "observed" time rate of change of the state vector computed from a single forward time step. In matrix form this system is written as

$$\begin{bmatrix} \mathbf{B} & \mathbf{F} & \mathbf{O} \\ \mathbf{F} & \mathbf{B} & \mathbf{C} \\ \mathbf{O} & \mathbf{C} & \mathbf{O} \end{bmatrix} \begin{bmatrix} \Delta\hat{\zeta} \\ \Delta\hat{D} \\ \Delta\hat{\phi} \end{bmatrix} = i \begin{bmatrix} (\delta_t\hat{\zeta})_G \\ (\delta_t\hat{D})_G \\ (\delta_t\hat{\phi})_G \end{bmatrix} \qquad (2.14)$$

where the submatrices **B**, **F**, and **C** have matrix elements defined by (2.10) to (2.12) and their precise structure may be found in T89's Appendix. The matrix **Q** has only zero elements.

Scheme *B* is characterized by constricting the slow modes to be stationary and nondivergent [in order to be able to solve the system of equations (2.14) for $\Delta\mathbf{X}$ belonging to the fast subspace]. This constriction is achieved by neglecting the $b_n^m\hat{\zeta}_n^m$ term (beta term) appearing in (2.7) for the vorticity tendency. Thus, **A** is replaced by $\mathbf{A}_0$, i.e.,

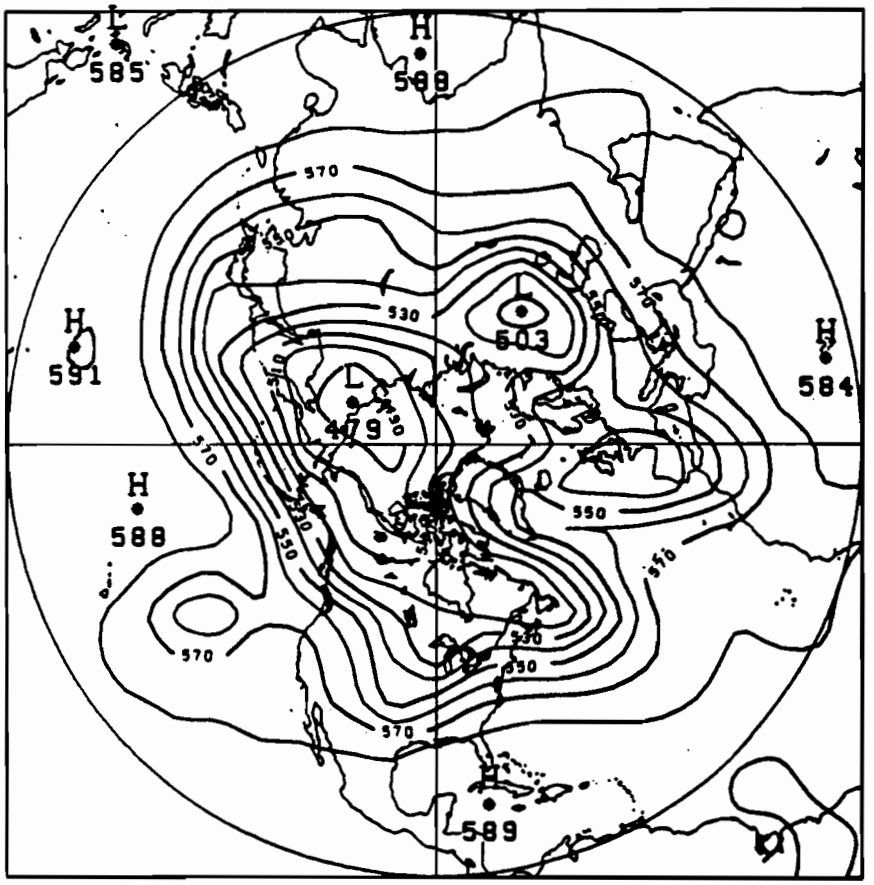


FIG. 2. The uninitialized height field (contour interval = 10 dam).

$$\begin{bmatrix} \mathbf{B} & \mathbf{F} & \mathbf{O} \\ \mathbf{F} & \mathbf{B} & \mathbf{C} \\ \mathbf{O} & \mathbf{C} & \mathbf{O} \end{bmatrix} \rightarrow \begin{bmatrix} \mathbf{O} & \mathbf{F} & \mathbf{O} \\ \mathbf{F} & \mathbf{B} & \mathbf{C} \\ \mathbf{O} & \mathbf{C} & \mathbf{O} \end{bmatrix}$$

where the matrices **B**, **C**, **F** have matrix elements b_n^m, c_n^m, f_n^m, respectively, [defined in (2.10)–(2.12)] and are the same as in the original system. The matrix $\mathbf{A}_0$ admits slow modes which are stationary, nondivergent, and characterized by the linear balance relationship

$$\hat{\mathbf{D}}_R = \mathbf{0}, \quad \mathbf{C}\hat{\phi}_R = -\mathbf{F}\hat{\zeta}_R \qquad (2.15)$$

where the lower index stands for Rossby modes.

Due to the approximation of matrix **A**, the physical space diagnostic equation for the divergence field correction is given by namely, Eq. (4.15) of T89;

$$(\mathbf{F}^2 + \mathbf{C}^2)\Delta\hat{\mathbf{D}} = i\mathbf{F}(\delta_t\hat{\zeta})_0 + i\mathbf{C}(\delta_t\hat{\phi})_0. \qquad (2.16)$$

Note that $\Delta\hat{\mathbf{D}}$ is completely determined from the model tendency of the vorticity and geopotential fields. This diagnostic form for $\Delta\hat{\mathbf{D}}$ can be shown to be related to quasi-geostrophic theory and this aspect is examined in section 5.

The last ingredient which is needed for the variational scheme comes from the second equation of the algebraic system (2.13) (using $\mathbf{A}_0$), i.e.,

$$\mathbf{F}\Delta\hat{\zeta} + \mathbf{B}\Delta\hat{\mathbf{D}} + \mathbf{C}\Delta\hat{\phi} = i(\delta_t\hat{\mathbf{D}})_G. \qquad (2.17)$$

Equations (2.15)–(2.17) constitute the desired constraints on slow and fast mode components in physical space needed by the variational scheme. Although higher-order nonlinear balancing could be envisaged within a variational framework, the present paper is concerned only with first-order nonlinear balancing. We now proceed to describe the specific form of the variational scheme.

3. Variational formulation

The implicit form of the variational scheme presented here is the spectral analogue of the scheme introduced by FT89. The main features of this iterative approach are given by the following steps;

(i) Given an initial state vector $\mathbf{X}^0$, generate a correction state vector $(\Delta\mathbf{X})_G$ where the new state

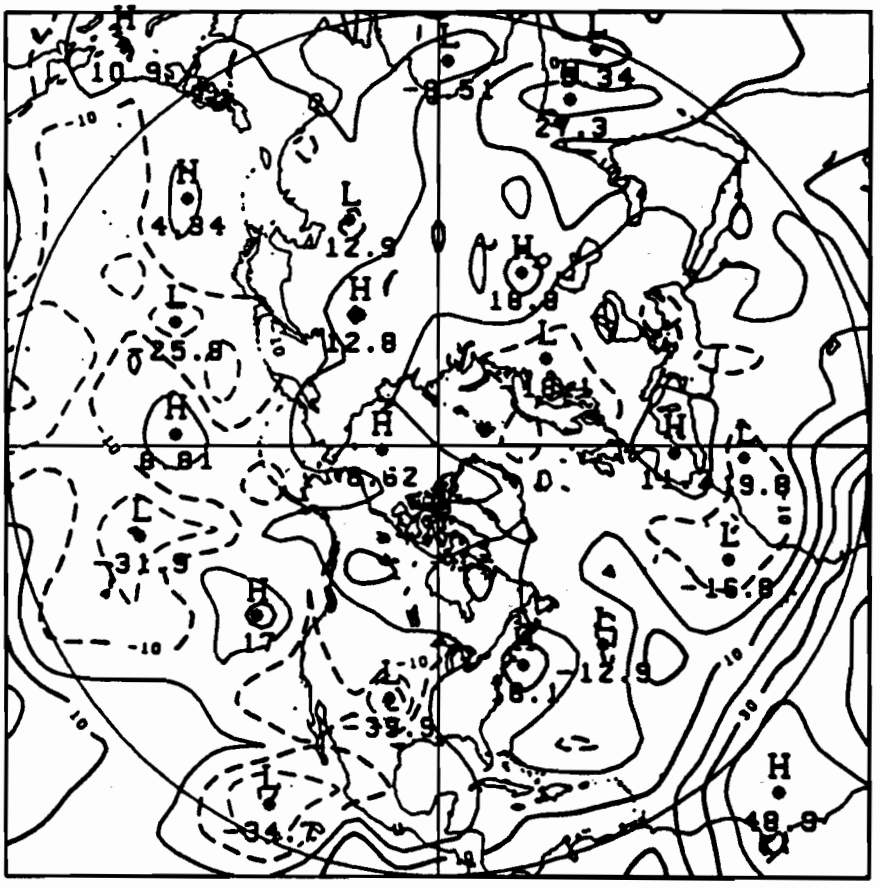


FIG. 3. Difference in height field between unconstrained initialization and no initialization (contour interval = 10 m).

$$\mathbf{X}^1 = \mathbf{X}^0 + (\Delta \mathbf{X})_G$$

is in approximate balance.

(ii) To return "closer" to $\mathbf{X}^0$ (a metric has to be defined here), alter the slow Rossby part of the state vector $\mathbf{X}^0$.

The nonlinear coupling between the slow and fast components of X can be taken into account by iterating steps (i) and (ii).

A phase space diagram is very helpful when discussing different constrained initialization algorithms. The present iterative scheme is illustrated in Fig. 1, and will later be used in the special context of height-constrained initialization. Each iteration of the combined steps (i) and (ii) are indicated, where points 1 and 2 indicate the first and second iterations, respectively. Point O indicates the original analyzed fields and point I indicates the balanced state obtained with VINML.

In practical situations, it is found (Daley 1978; FT89) that three iterations are sufficient to achieve a balanced state I. It is worth noting at this point the slight departure of steps (i) and (ii) from Daley's original suggestion. Machenhauer's initialization procedure involves an iterative scheme in order to get $(\mathbf{X})_G$ in balance with the slow modes of $\mathbf{X}^0$. Incrementing the new state X with $(\Delta \mathbf{X})_G$ generates (assuming convergence of the scheme) a state vector X such that $(\delta_t \mathbf{X})_G \to \mathbf{O}$. Thus (2.13) ensures that $(\Delta \mathbf{X})^n \to \mathbf{O}$ (where n is the iteration count), i.e., a first-order balanced state has been achieved. In the formulation and implementation of the variational algorithm, the approximate balancing correction vector $(\Delta \mathbf{X})_G$ mentioned in step (i) is given by (2.13) for each iteration of steps (i) and (ii). This approach has been taken previously by Temperton (1984) and FT89. Tribbia (1982) however used five iterations of Machenhauer's scheme (2.13) to obtain $(\Delta \mathbf{X})_G$ for each cycle of his variational initialization scheme. Figure 1 illustrates the approach where the nonlinear manifold M represents states where $(\delta_t \mathbf{X})_G = \mathbf{O}$. However in operational practice using analyzed data, only a few iterations of the unconstrained scheme are made; this generally results in a significant reduction of $(\delta_t \mathbf{X})_G$. Further con-

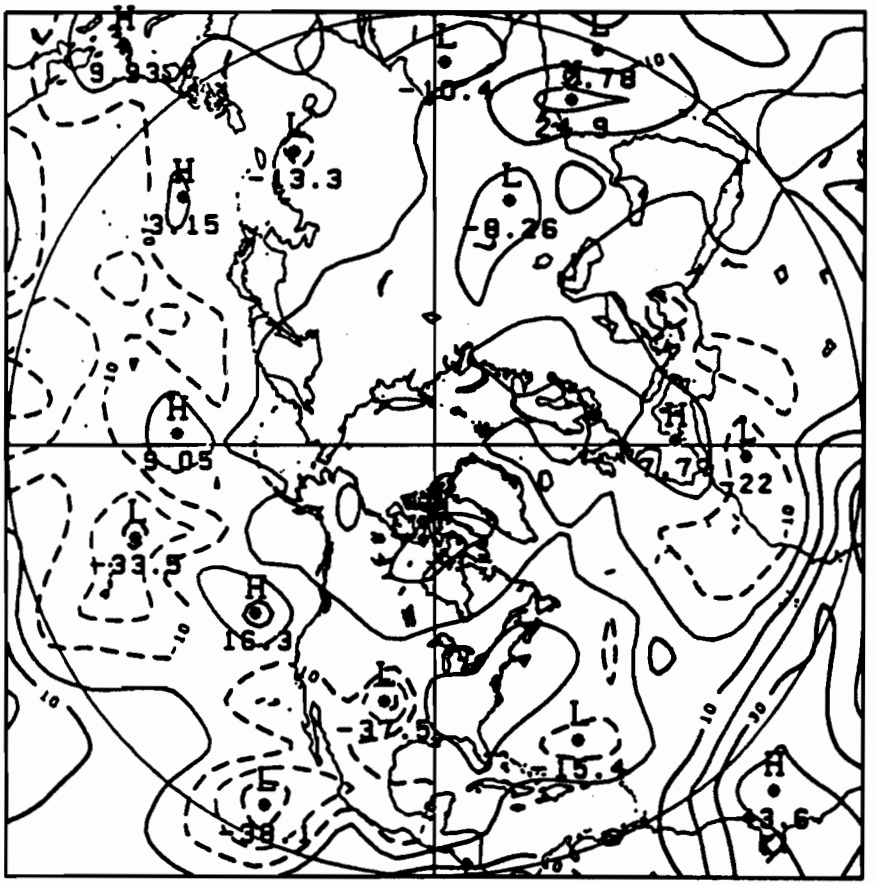


FIG. 4. Difference in height field between variational adjustment (using latitudinal weights ω_ϕ, ω_ϕ and no initialization (contour interval 10 m).

sideration on this point will be given in section 4c. Although not essential, the approach taken here permits a direct coupling of steps (i) and (ii) in physical space and gives rise to a variational problem with a strong differential constraint in the sense of Sasaki (1958).

As in FT89, (2.15) expressed in terms of the correction state $\Delta\mathbf{X}$ is

$$(\Delta\hat{\mathbf{D}})_R = 0, \quad \mathbf{C}(\Delta\hat{\phi})_R = -\mathbf{F}(\Delta\hat{\zeta})_R. \tag{3.1}$$

For an f plane shallow-water model, the characterization of the Rossby mode corrections with the linear balance equation (3.1) is exact and the variational scheme would be identical to Daley's (1978) variational scheme. On the sphere, however, this characterization of the slow modes ceases to be an exact relation due to the presence of the beta term in (2.2). In the light of previous studies (Burger 1958; Moura 1976; Wiin-Nielsen 1979), constraint (3.1) remains accurate except for the largest scale modes. Combining (2.17) with (3.1) gives the desired constraint on $(\Delta\mathbf{X})$:

$$M(\Delta\hat{\zeta}, \Delta\hat{\phi}) = \mathbf{F}\Delta\hat{\zeta} + B(\Delta\hat{\mathbf{D}})_G + C\Delta\hat{\phi} - i(\delta_t\hat{D})_G = 0. \tag{3.2}$$

Also, since $(\Delta\hat{D})_R = 0$ and considering what has been mentioned previously for $(\Delta\hat{D})_G$ in (2.16), the following functional is selected to be minimized on the sphere

$$J = \int_{-\pi/2}^{\pi/2}\int_0^{2\pi} \omega_\phi(\Delta\phi)^2 + \Phi\omega_\psi[\nabla(\Delta\psi)]^2 d\lambda\cos\theta d\theta \tag{3.3}$$

where ω_ϕ and ω_ψ are weights attributed to the reliability of the mass and rotational components of the wind field, respectively.

Functional J, for the special case where $\omega_\phi = \omega_\psi$ = constant, represents a truncated form of the linearized energy of the shallow-water equations in the sense that the divergent part of the wind field is absent. In continuous form, (2.17) can be shown to be

$$M(\Delta\psi, \Delta\phi) = \nabla\cdot[f\nabla(\Delta\psi)] - \nabla^2(\Delta\phi) - \frac{2\Omega}{a^2}\frac{\partial(\Delta\chi)_G}{\partial\lambda} + (\delta_t D)_0 = 0. \tag{3.4}$$

We still require that the mean depth of the fluid remain unchanged (mass conservation) which implies a further integral constraint

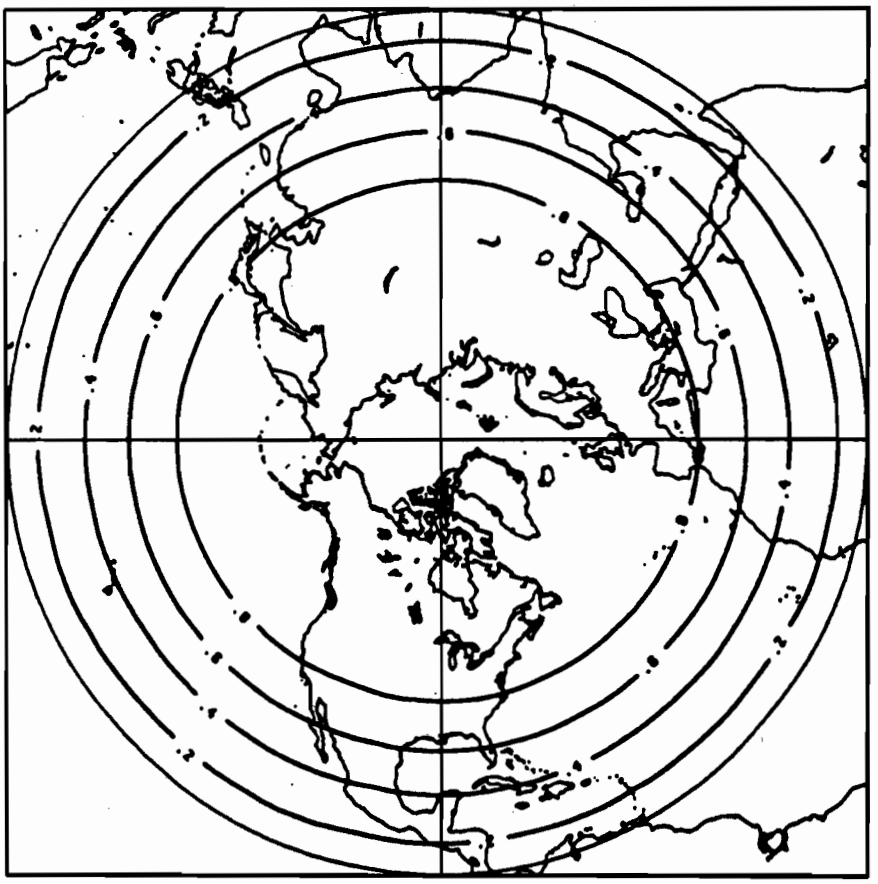


FIG. 5. Latitudinally varying weight ω_ϕ.

$$\int_{-\pi/2}^{\pi/2}\int_{0}^{2\pi}(\Delta\phi)d\lambda\cos\theta d\theta = 0. \qquad (3.5)$$

The variational problem is thus to minimize functional J under the strong constraints (3.4) and (3.5). As usual, the minimization problem thus posed is applied to the *current* fields in the initialization loop defined by steps (i) and (ii) (see also FT89). This aspect is not illustrated in Fig. 1.

The system of Euler–Lagrange equations to be solved are given by (3.5) and

$$\nabla^2\gamma = \omega_\phi(\Delta\phi) + \gamma^* \qquad (3.6)$$

$$\nabla\cdot[\Phi\omega_\psi\nabla(\Delta\psi)] = \nabla\cdot(f\nabla\gamma) \qquad (3.7)$$

$$M(\Delta\psi, \Delta\phi) = 0 \qquad (3.8)$$

where γ and γ^* are (as in FT89) Lagrange multipliers (γ^* is a constant to be determined) associated with the constraints (3.4) and (3.5), respectively.

As a particular case, this system admits the same solution as the unconstrained implicit scheme for a choice of weights given by $\omega_\phi = \omega_\psi$ = constant (see appendix A). On the basis of FT89 this result is to be expected.

The numerical procedure for solving (3.6)–(3.8) is essentially the same as the finite-element algorithm described in FT89. For one iteration, the following steps summarize the variational initialization scheme:

(1) Run part of the unconstrained implicit scheme to obtain $(\Delta\hat{D})_G$, following T89:

(i) Run the model for one forward time step to obtain the time tendencies of the spectral coefficients (ζ_n^m, D_n^m, ϕ_n^m).
(ii) Scale the tendencies according to (2.4) to (2.6).
(iii) Solve (2.16) for $(\Delta\hat{D})_G$.

(2) Compute the "forcing" term:

$$\frac{-2\Omega}{a^2}\frac{\partial}{\partial\lambda}(\Delta\chi)_G + (\delta_t D)_0$$

where $(\Delta\chi)_G$ is obtained from $(\Delta D)_G$ using (2.5) and the inverse Laplacian.

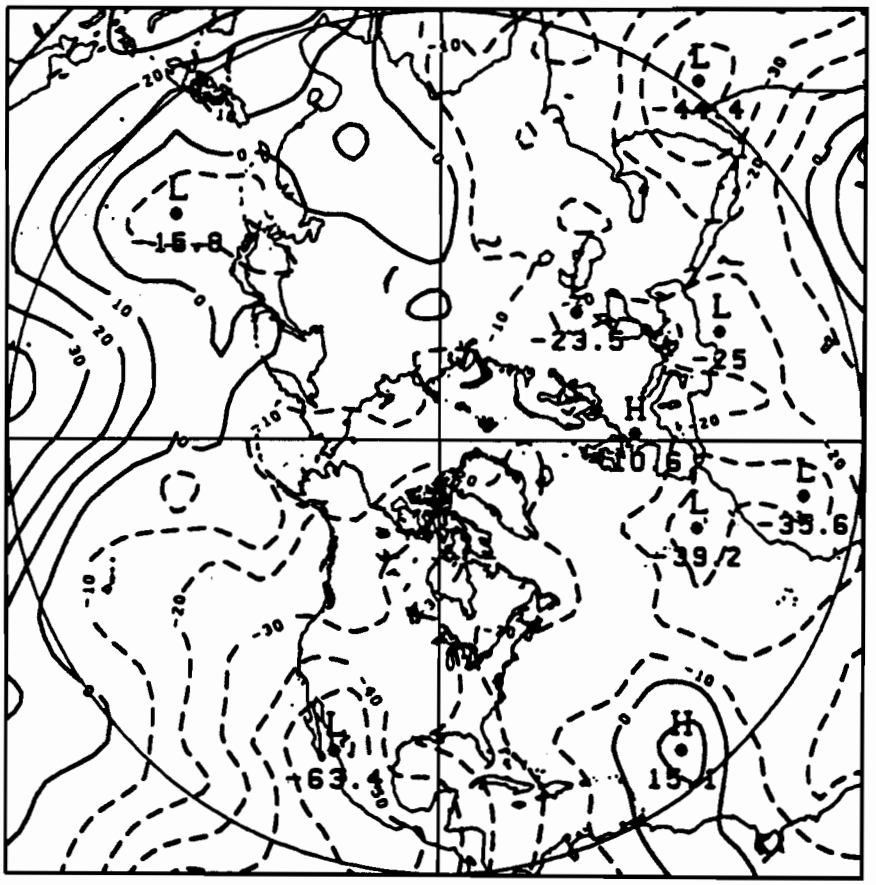


FIG. 6. Difference in potential function χ between unconstrained initialization and no initialization (contour interval 10×10^5 m^2 s^{-1}).

(3) Solve the Euler–Lagrange equations (3.6)–(3.8) for the spectral components of $(\Delta\psi)$, $(\Delta\phi)$.

(4) Add increments $(\Delta\psi)$, $(\Delta\chi)$, $(\Delta\phi)$ to obtain initialized spectral coefficients.

(5) Generate new values of (ζ, D, ϕ) on the Gaussian grid.

4. Results of VINMI

a. The model and data used

To facilitate the comparison between INMI (based on "scheme *B*" of T89) and VINMI described in section 3 this study uses the same model and data presented in T89. A global barotropic spectral model with triangular truncation at wavenumber 63 is also used. The initial data are from a FGGE analysis of heights and winds at 500 mb for 0000 UTC 21 December 1978. All other features of the model not mentioned here are implicitly assumed to be the same as in T89. Figure 2 shows the initial height field. The mean equivalent depth H, ($\Phi = gH$ where g is the acceleration due to gravity) used in this study is 5600 m.

b. The controlling aspect of the scheme

The interest in this subsection is the effective control of changes on the rotational part of the wind field and to the mass field by the variational scheme. The effect of varying the weights ω_θ and ω_ϕ is both spatial directions (λ, θ) are examined. For all experiments reported in this section, this study always refers to two iterations for the INMI case and three iterations for VINMI. However in subsection 4c, the effect by using more than three iterations is investigated. In all experiments the original fields are those previously described in subsection 4a. The typical CPU time for the VINMI results to be presented is approximately 80 s on a Cray-XMP.

Figure 3 shows the difference between INMI and the original uninitialized height fields. This difference field may be obtained by adding Figs. 6a,b of T89. Figure 3 is to be compared with Fig. 4, which shows the difference between the VINMI and original uninitialized height fields. The weights used for this experiment have only a latitudinal structure (see Fig. 5) and are given by (Daley 1978):

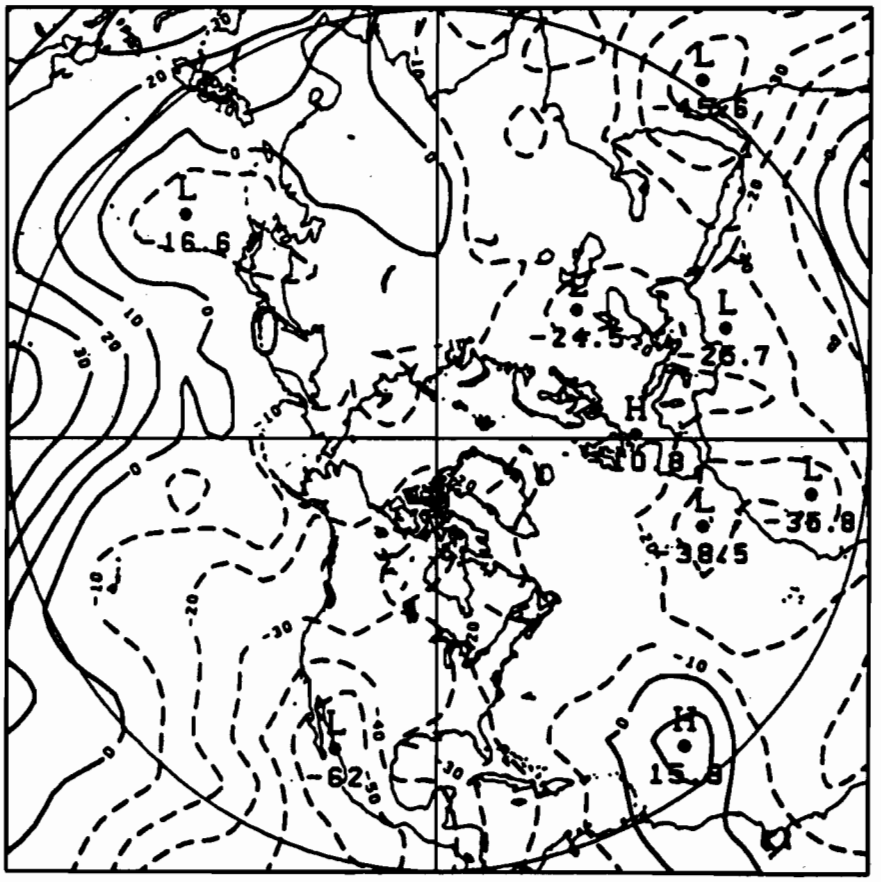


FIG. 7. Difference in potential function χ between variational adjustment using latitudinal weights ω_θ, ω_ϕ and no initialization (contour interval 10×10^5 m^2 s^{-1}).

$$\omega_\psi = (1 - \mu^2)^4; \quad \mu = \sin\theta \tag{4.1}$$

$$\omega_\phi = 1 - \omega_\psi. \tag{4.2}$$

This choice of weighting forces the wind field (rotational part) to adjust to the original height field analysis at high latitudes and vice versa at lower latitudes. The strong retention of mass information north of 40° latitude is a consequence of (4.1) and (4.2) and is clearly seen in Fig. 4. One feature of the variational scheme of section 3 is that the divergent part of the flow is not explicitly controlled within each iteration of steps 1 to 5.

Comparing Figs. 6 and 7 obtained by using VINMI (Daley's weights) and INMI, respectively, it is seen that both schemes lead to almost identical departures from the original uninitialized velocity potential function χ. The implication of these results is discussed in section 6. To show the effect of VINMI on the rotational part of the wind field, Fig. 8 shows the difference (vector) of VINMI and INMI wind fields. These differences are mainly to the rotational part of the flow as discussed previously. The structure of the wind field correction in Fig. 8 agrees with the latitudinal structure of ω_ψ specified previously; that is, retention of the rotational part of the original flow at low latitude and adjustment to the original height field at higher latitudes. The maximum wind modulus of the difference vectors is 1.5 m s^{-1}.

We now examine the behavior of the variational scheme using a fully variable structure of the weights in the horizontal. The following choice of weights ω_ψ and ω_ϕ will simply serve to illustrate the ability of the scheme to discriminate between the reliability of the analyzed fields over continents and oceans. The specific form of the weights for this experiment are:

$$\omega = \omega_D \quad \text{for} \quad \lambda \in \Lambda$$

$$= \{[60°\text{W}, 130°\text{W}] \cup [135°\text{E}, 15°\text{W}]\}$$

$$\omega = 10^{-2}\,\omega_D{}^2 \quad \text{for} \quad \lambda \notin \Lambda \tag{4.3}$$

where ω stands for both ω_ϕ and ω_ψ, and ϕ_D is defined by (4.1) and (4.2). With such a choice of weights, one expects that the largest differences between two fields initialized using weights (4.1) and (4.2) and initialized using weights (4.3) should be found over the ocean sectors deduced from (4.3). One clearly sees (Fig. 9a)

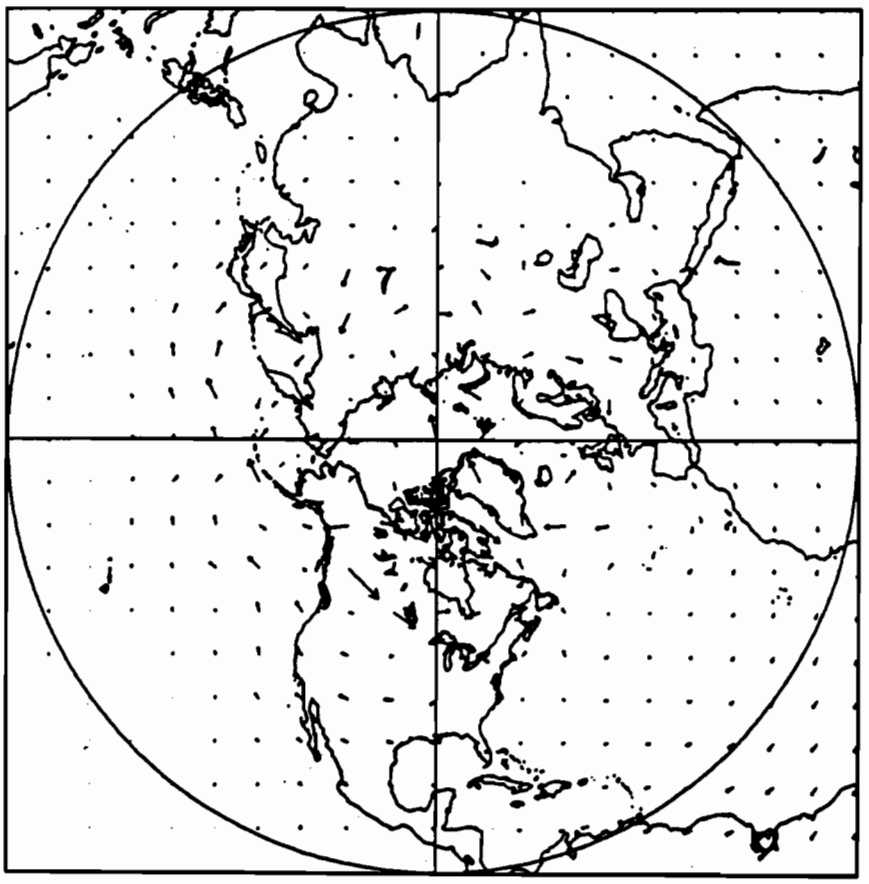


FIG. 8. Wind differences between variationally adjusted (using latitudinal weights) and unconstrained initialized fields. Maximum wind modulus = 1.5 m s^{-1}.

the effect of "ocean discrimination," the elongated north–south pattern of differences over the oceans being obtained due to the use of ω_D^2 in (4.3). This last choice was made for illustration purposes only in order to demonstrate the feasibility of such corrections. The overall effect of the variational scheme on the original analyzed height and wind fields is depicted in Figs. 9b,c. The European continent is the typical type of corrections obtained by allowing the weights to vary horizontally. Assuming higher reliability of the analyzed height and wind fields over the continent as compared to the adjacent North Atlantic Ocean, the result of Fig. 3 (i.e., unconstrained INMI) may be considered to be undesirable. Forcing the initialization scheme to recognize the weaknesses of the analyzed height *and* wind fields over the oceans would be greatly preferable. For the same region considered above, Figs. 9b,c illustrate clearly this aspect where the height and wind fields mutually adjusted over the ocean without damaging the analyzed fields over the European continent. It is also important to notice that this behavior of VINMI was feasible due to the possibility of using full horizontal variability of the weights. It is however clear that the proper specification of weights must be related to the estimation of the statistical error distribution of the analyzed fields. Since the model is global this study now considers the Southern Hemisphere results of VINMI. Note here that symmetric weights were used about the equator in (4.3), the reality of such specification being immaterial here. Figure 10 gives the differences between VINMI using (4.3) and the original uninitialized analysis. In these quadrants, one first identifies the southern equivalent of the North American continental sector where retention of mass is maximal (southwest of Argentina for example). The other continental sector located southwest of Australia, is associated (south of 40°) with a strong increase in the retention of the mass field. In Fig. 11 (which is the analogue of Fig. 10 but using INMI), the previously described southern "land" regions [i.e., as defined by (4.3)] bear important height corrections reaching 53 meters in one case. The variational scheme has eliminated these strong corrections, a result which illustrates the robustness of the scheme.

Finally, the effect of constrained versus unconstrained initialization schemes on a 2 day forecast may

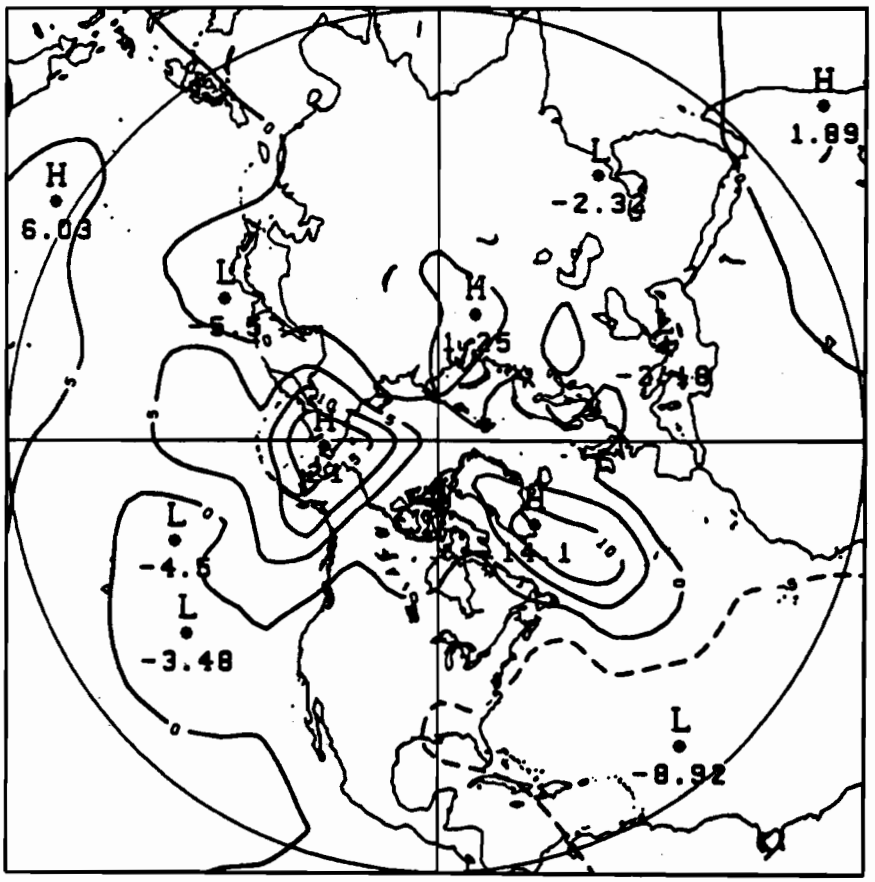


FIG. 9a. Difference in height between variational initialization using latitude–longitude weights and variational initialization using latitudinal weights (contour interval = 5 m).

be examined by using a more realistic set of weights. This set may be obtained by extracting typical standard deviations of the estimated height and wind analysis errors at 500 mb produced by the operational objective analysis scheme at CMC (Centre Météorologique Canadien) (Mitchell et al. 1990) for a winter case. Figure 12a shows the estimated 500-mb height analysis error (m). We first note that the amplitude of ~3 m over the northern continents increases to approximately 15 m over the oceans and the tropics. The continents are clearly distinguishable by a strong gradient in the field of standard deviations along the coasts. This effect was considered in the previous experiment by using a reducing factor for the ocean sectors (see also FT89). Also note that no attempt was made to smooth the local minima over the Pacific Ocean where upper-air data were available to the analysis (e.g., Hawaii). Similar features can be observed in Fig. 12b for the standard deviation of the zonal component of the wind analysis error field (the meridional component is not shown but has a similar structure) with a minimum deviation of 2.5 m s^{-1} over the continents (extratropics) increasing to ~5 m s^{-1} elsewhere. The model is now integrated for 48 h using variationally initialized fields with

$$\Phi\omega_\psi = 1/\sigma_\psi^2; \quad \omega_\phi = 1/\sigma_\phi^2$$

where σ_ψ and σ_ϕ are the standard deviations as previously described. Figure 12c shows the difference in the height field between VINMI (using the latter specification of weights) and INMI after 48 h. As in FT89 (Fig. 12), most of the differences are found in midlatitudes where the RMS value is approximately 20 m. It is possible that other cases would show larger differences, particularly after five days for example. VINMI might be expected to produce a cumulative improvement of the analysis when employed in a data assimilation cycle by improving the quality of the 6-h forecast used as a first guess in the data assimilation scheme.

c. *The balancing aspect of the scheme*

As previously shown, the variational scheme has the flexibility of allowing full variability in the weighting functions ω_ψ and ω_ϕ, which may represent the initial state of the atmosphere more realistically than the un-

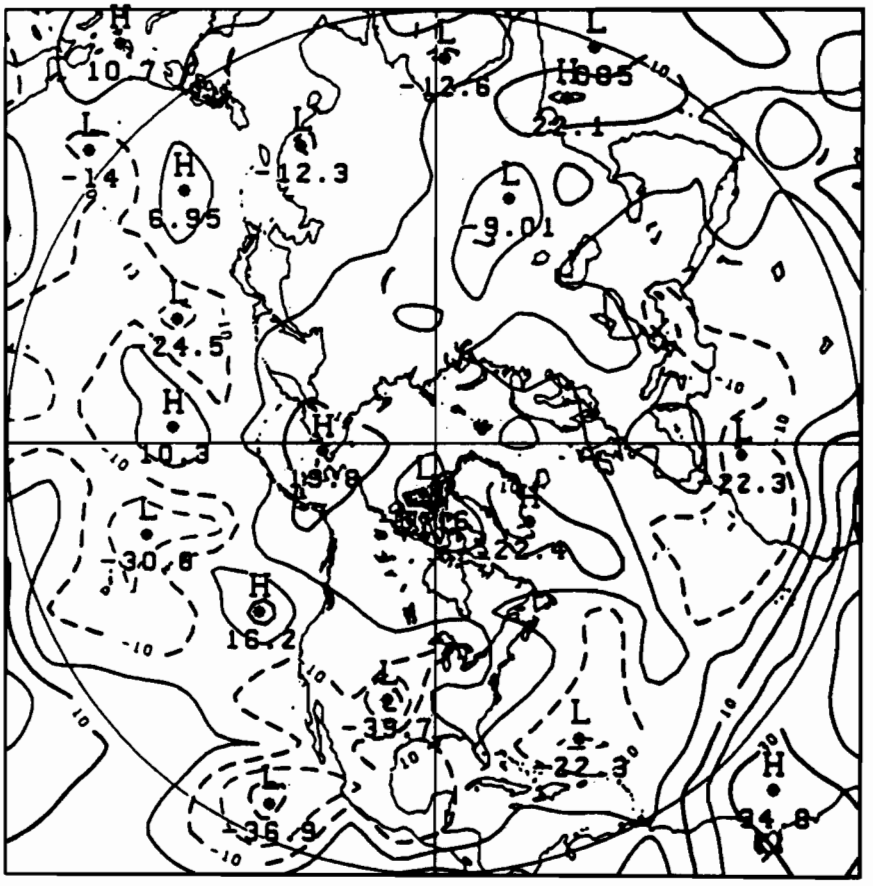


FIG. 9b. Difference in height field between variational initialization using latitude–longitude weights and original uninitialized fields (contour interval = 10 m).

constrained case. The main benefits of the latter being to efficiently control the excitation of high-frequency gravity–inertia waves, this aspect of the variational scheme should also be observed in order to be a viable generalization. This point will be considered here as a final comparison between INMI and VINMI schemes.

Standard monitoring of Machenhauer's initialization scheme is based on the evaluation of the BAL_t value, for each iteration of the scheme, which is defined as

$$BAL_t = \sum_j \left| \frac{dG_j}{dt} \right|^2$$

where dG_j/dt is the time tendency of a particular gravity mode. We proceed in the same way as T89 for the evaluation of BAL_t in physical space. The remarks given in FT89 also apply here, in the sense that the usefulness of monitoring the value of BAL_t in unconstrained NMI is based on the fact that Machenhauer's scheme attempts to set BAL_t to zero. VINMI proceeds in a different way since the amplitude of the slow modes are changed for each iteration of the scheme. The retention of information on the initial data in physical space is a potential barrier for the vanishing of BAL_t, a consequence which is absent in unconstrained schemes. This aspect of VINMI is clearly represented in Fig. 13 where both variational and unconstrained schemes have been tested using five iterations. As noted in T89, scheme *B* (which is represented by the unconstrained scheme here) has a monotonic decrease of BAL_t even after 10 iterations (see his Fig. 3). Temperton stressed that "this latter convergence may however be towards a state which is not appropriate as an initial condition for the model." This latter remark, together with the previous discussion characterizes the saturation value of BAL_t. Further, once the variational scheme has attained it's saturation value of BAL_t which is associated with three iterations of the scheme (see Fig. 13), a comparison of model integrations can be done using this initial state and the initial state obtained with the unconstrained scheme using two iterations. We note immediately that these two initial states have approximately the same value of BAL_t.

The variational scheme used the latitude–longitude specification of weights which are described in section 4b. In each integration, the model was run for 48 h

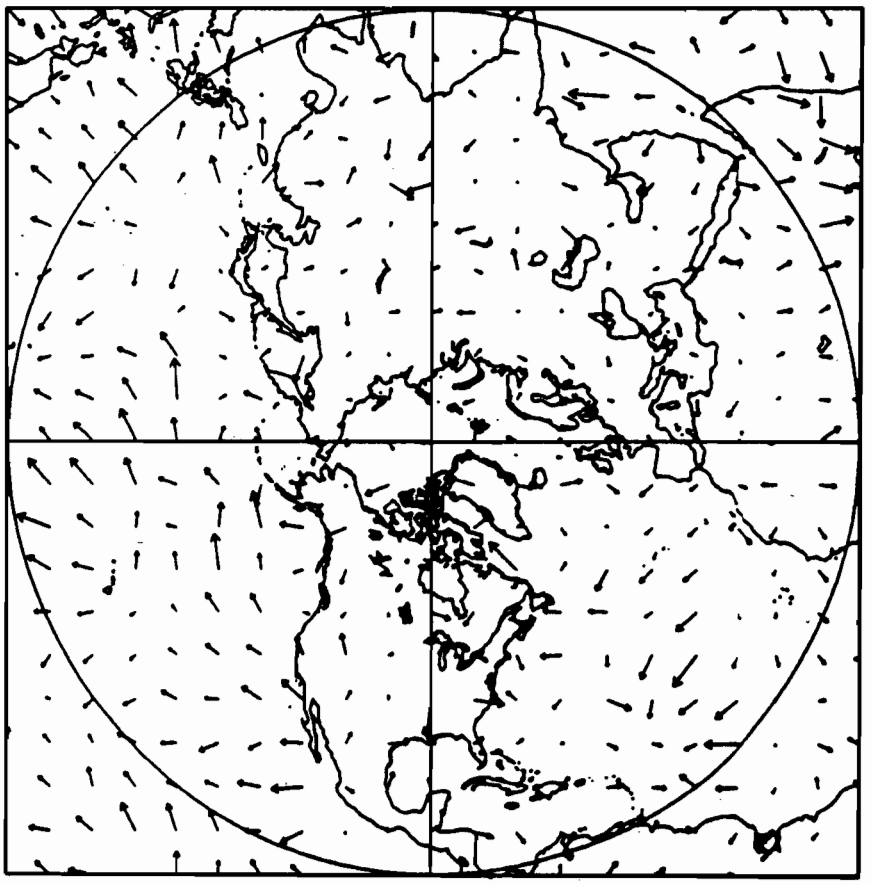


FIG. 9c. Wind difference between variationally adjusted (using latitude–longitude weights) and original uninitialized fields (maximum modulus = 1.5 m s^{-1}).

and the height- traces were stroed for three different locations on the sphere. These traces are shown in Figs. 14–16 corresponding to the points (45.70°N, 180°), (0.93°N, 180°), (45.70°S, 180°), respectively, and these points were chosen to represent the typical degree of balance achieved at those latitudes. The quality of the balance attained can be judged by further comparison with a model integration using uninitialized initial conditions. This latter integration shows high-frequency oscillations with an initial amplitude of approximately 50 m for the Northern Hemisphere case. For both initialized model integrations, these oscillations are considerably reduced and a comparable balance is observed. It is found that further iterations of the unconstrained scheme does not produce a better balancing even though the corresponding values of BAL_1 are dramatically different; second-order initialization (T89) is therefore needed for a refinement of the balance. Figures 14 and 16 clearly illustrate the effect of the variational scheme on the slow mode behavior. In one case (46°S) a small departure in the height field between unconstrained and variational run

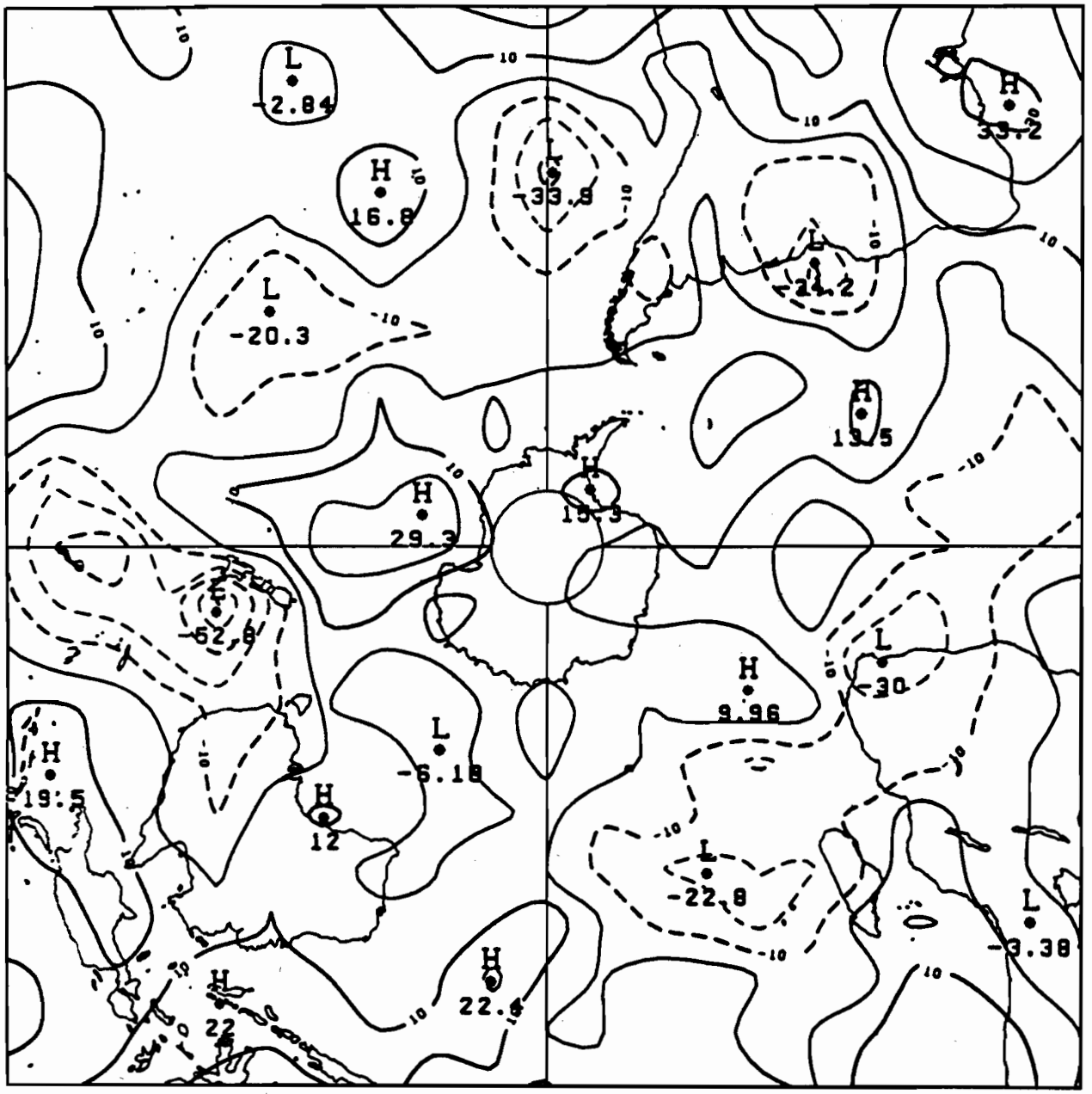


FIG. 10. Same as Fig. 9b but for the Southern Hemisphere (contour interval = 10 m).

is maintained for the first 15 h but *significantly* diverges in the ensuing forecast reaching a 50 m difference at 48 h.

It is concluded from these last experiments that the variational extension of implicit NMI does maintain a balance in the initial data to a degree comparable to the unconstrained scheme and allows the possibility of significantly altering the slow evolving part of the signal. It is therefore quite clear that a proper specification of the weights ω_ϕ and ω_ψ is crucial for a refinement in the analysis and forecast of these slow modes.

5. VINMI and the ellipticity condition

In the early applications of variational normal mode initialization in the context of realistic data and model of the atmosphere, it was realized that in some cases, Machenhauer's initialization scheme failed to specify a balanced initial wind field given a particular distribution of mass field. Daley (1978) first observed an approximate correspondence of "nonelliptic" regions and regions where the NMI balancing process was diverging; that is, one could not find a flow which bal-

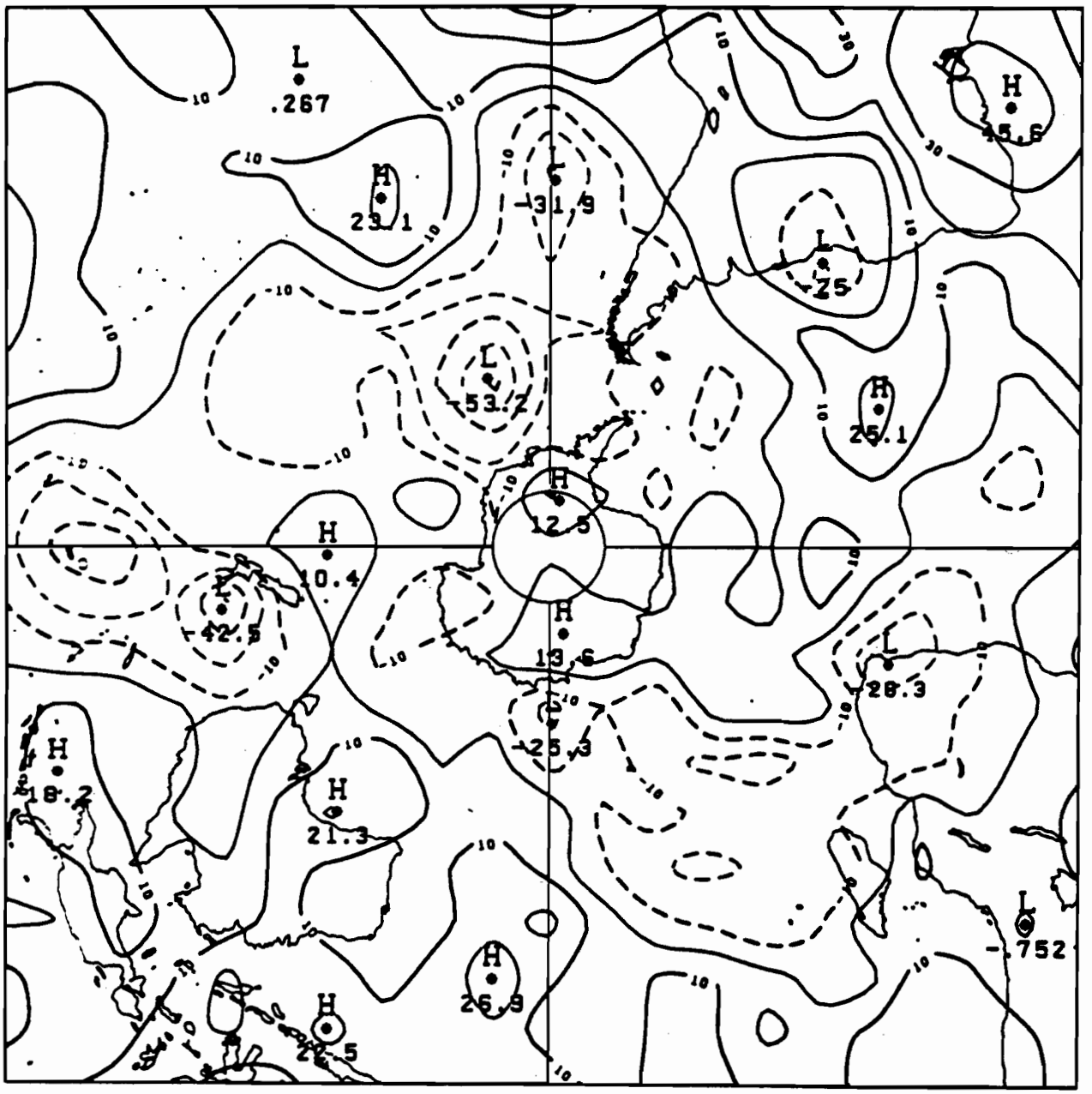


FIG. 11. Difference height field between unconstrained initialization and no initialization for the Southern Hemisphere (contour interval = 10 m).

anced the mass field over those regions. Tribbia (1981) clarified this point by analyzing the relationship between the results of a local f-plane model (where the analytic balanced flow is known) and a global spectral model with rhomboidal truncation 20. Once again a strong correlation was noted between the appearence of nonelliptic geopotential regions and the breakdown of the iteration scheme used in nonlinear normal mode initialization.

Before discussing the "ellipticity problem" in the context of height constrained initialization, it will be sufficient for this purpose to relate INMI with quasi-geostrophic theory in the context of shallow-water models on an f plane.

a. *INMI and the classical nonlinear balance equation*

Leith (1980) (hereinafter noted L80) showed that, on a periodic f-plane Boussinesq model, the first-order Baer–Tribbia (1977) estimate or one "iterate" of Machenhauer's scheme (starting from geostrophic initial conditions) was equivalent to solving the quasi-geostrophic balance equations. In terms of total geopotential and streamfunction ϕ_1, ψ_1, respectively, L80 found a related version of the classical nonlinear balance equation given by:

$$\nabla^2\phi_1 = f\nabla^2\psi_1 - M(\phi_0) \qquad (5.1)$$

where

$$M(\phi_0) = \frac{2}{f^2}\left[\left(\frac{\partial^2\phi_0}{\partial x\partial y}\right)^2 - \left(\frac{\partial^2\phi_0}{\partial x^2}\right)\left(\frac{\partial^2\phi_0}{\partial y^2}\right)\right].$$

Here the index "0" concerns the rotational part of the initial geopotential field while index "1" indicates the resulting fields after one iteration of Machenhauer's scheme (not to be confused with L80's notation). The same results as (5.1) may be obtained by using the implicit form of nonlinear NMI. This process starts

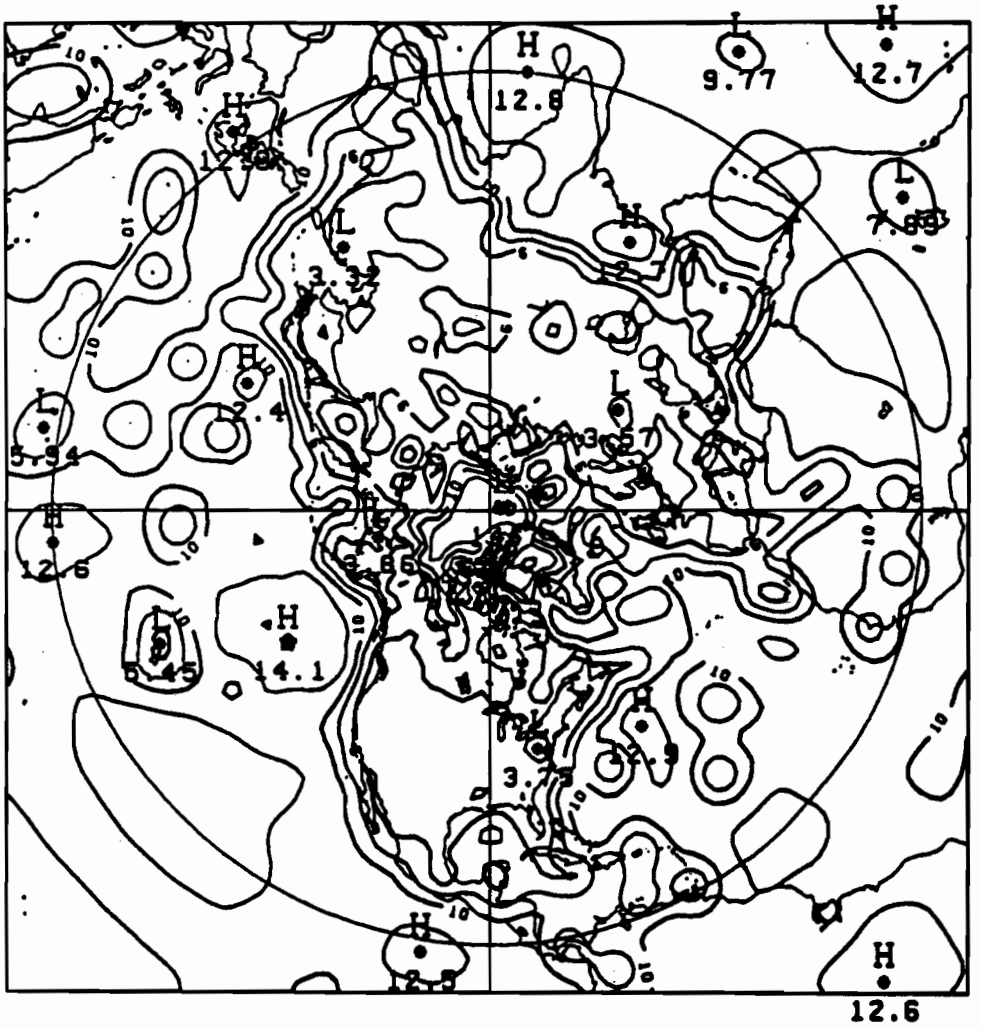


FIG. 12a. Typical standard deviation of the estimated errors of the analyzed 500-mb height field for a winter case. Contour interval = 2 m.

first with a linear initialization step which (due to the f-plane approximation) is the requirement of geostrophy. Note here that all the implicit schemes discussed by Temperton (1988, 1989) reduce to the same implicit scheme when dealing with an f-plane model. In that case, implicit and explicit NMI are exactly the same and the corresponding slow modes are all stationary and nondivergent. The result of the first implicit step is thus

$$Y_0 = \begin{bmatrix} \psi_0 \\ \chi_0 \\ \phi_0 \end{bmatrix}; \quad \psi_0 = \phi_0/f; \quad \chi_0 = 0. \qquad (5.2)$$

Second, starting from Y_0, integrate the nonlinear model for one time step to get the value of $(\delta_t D)$. Note that in order to avoid confusion, the index "o" accompanying the δ_t operator [as defined in (2.13)] will be omitted in the following derivations. This latter quantity is required in order to solve one of the balancing constraints (analogous to the one used as a strong constraint in section 3) on $(\Delta\phi)$, $(\Delta\psi)$ given by

$$\nabla^2(\Delta\phi) = f\nabla^2(\Delta\psi) + \delta_t D. \qquad (5.3)$$

The time tendency of the divergence field appearing in (5.3), using as initial conditions the previous linear step, is given by

$$(\delta_t D) = -\nabla^2\phi_0 + f\zeta_0 - \nabla\cdot(V_0\cdot\nabla V_0) \qquad (5.4)$$

where

$$V_0 = k\times\nabla\psi_0.$$

Using Helmholtz's decomposition of the wind field together with (5.2), the last term on the right-hand side of (5.4) may be rewritten as $-M(\phi_0)$. Putting the pieces together a "balance equation" can be written in terms of total geopotential and streamfunction (given by $\phi_1 = \phi_0 + \Delta\phi$, $\psi_1 = \psi_0 + \Delta\psi$) as

$$\nabla^2\phi_1 = f\nabla^2\psi_1 - M(\phi_0). \qquad (5.5)$$

Thus L80's result (5.11) is recovered as expected from the exact equivalence between explicit and implicit NMI for an f-plane model. As mentioned by L80, if

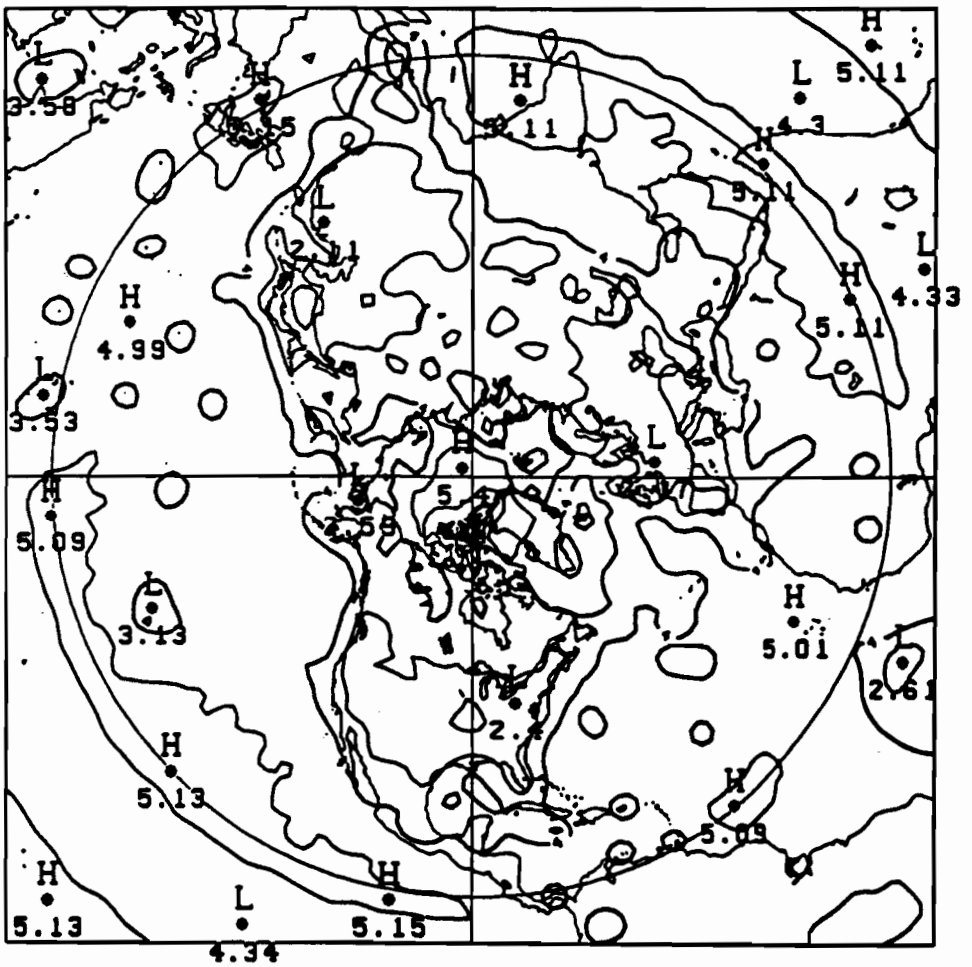


FIG. 12b. Typical standard deviation of the estimated errors of the analyzed 500-mb zonal component of the wind field for a winter case. Contour interval = 1 m s^{-1}.

term $\mathcal{M}$ were evaluated in terms of ψ_1, (5.5) would be the classical nonlinear balance equation of Charney (1955). Passing to the divergence field required by INMI, its connection with quasi-geostrophic theory is also straightforward to obtain. In the present context, Eq. (4.15) of T89 reduces to:

$$\nabla^2(\Delta D) - \frac{f^2}{\Delta}(\Delta D) = \frac{1}{\Phi}[-f\delta_t(\nabla^2\psi) + \nabla^2(\delta_t\phi)]. \tag{5.6}$$

For the nonlinear case, the nonlinear tendencies of ζ and ϕ are given by

$$(\delta_t\zeta) = -\mathbf{V}_0\cdot\nabla\zeta_0 - (f+\zeta_0)D_0$$

$$(\delta_t\phi) = -\mathbf{V}_0\cdot\nabla\phi_0 - \phi D_0.$$

Using (5.2) (i.e., a geostrophic initial step):

$$D_0 = 0, \quad \mathbf{V}_0\cdot\nabla\phi_0 = 0$$

and (5.6) reduces to

$$\nabla^2(\Delta D) - \frac{f^2}{\Phi}(\Delta D) = \frac{f}{\Phi}\mathbf{V}_0\cdot\nabla\zeta_0.$$

Using the geostrophic relation between the wind field and vorticity field with the geopotential ϕ_0, the quasi-geostrophic divergence equation is recovered on an f plane (Haltiner and Williams 1980, chapter 3) for the total divergence D (remember that $D_0 = 0$ so $\Delta D = D$ here):

$$\nabla^2(\Delta D) - \frac{f^2}{\Phi}(\Delta D) = \frac{1}{f\Phi}(\mathbf{k}x\nabla\phi_0)\cdot\nabla(\nabla^2\phi_0).$$

b. Height constrained initialization

From the preceding results, adjusting the rotational part of the wind field compulsively to the initial geopotential field via VINMI is susceptible to produce the

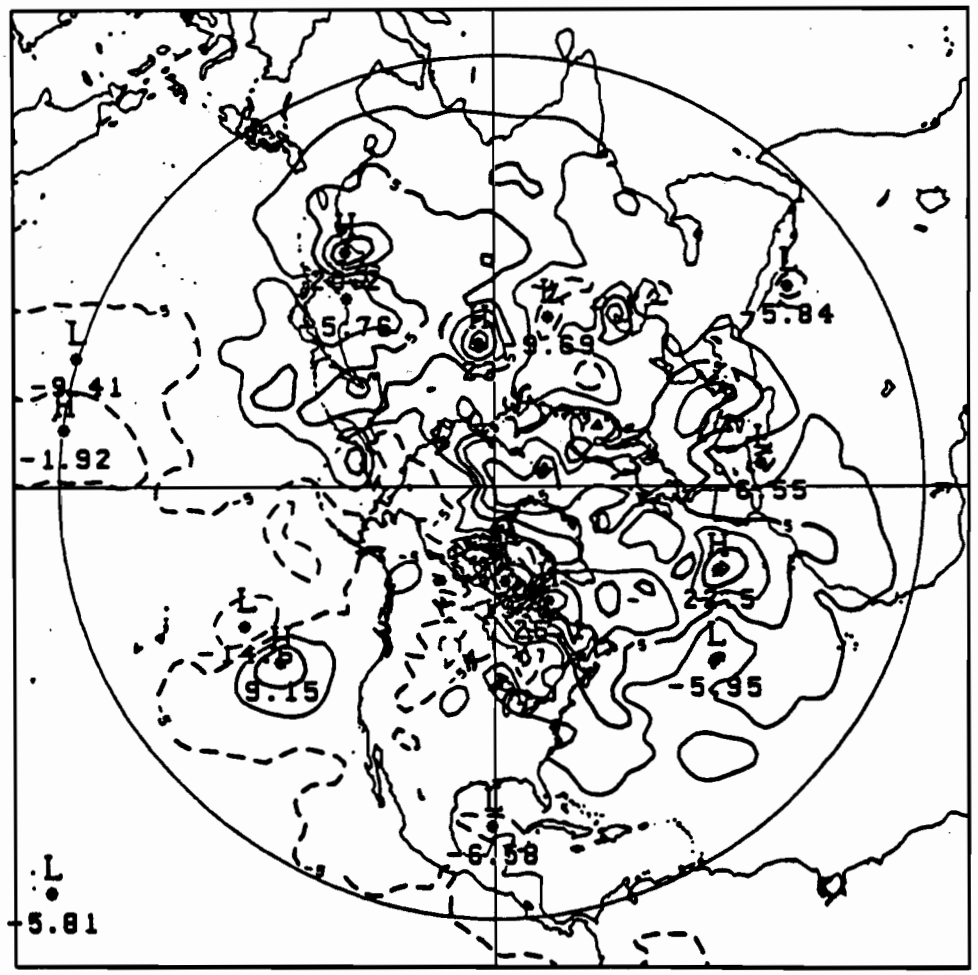


FIG. 12c. Difference in height field between variational initialization using known standard deviations of the estimated height and wind analysis errors, and unconstrained initialization for a 48-h forecast. Contour interval = 5 m. Maximum amplitude = 26 m.

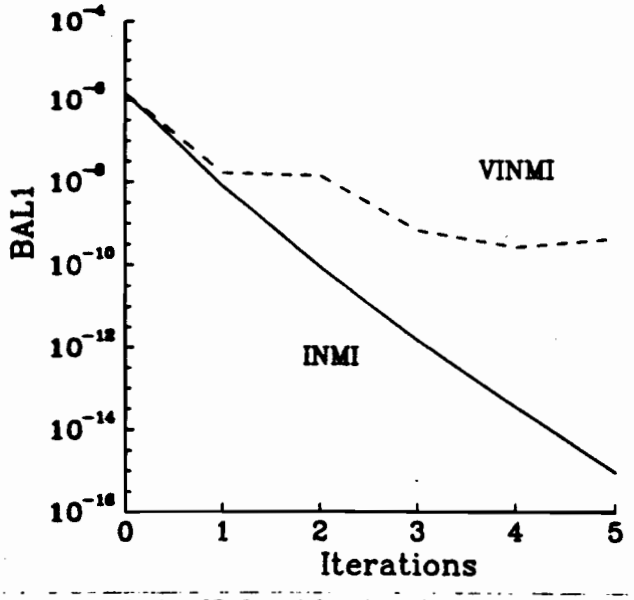


FIG. 13. Behavior of BAL_1 as defined in section 4c for INMI and VINMI (latitudinal weights) as a function of iteration number.

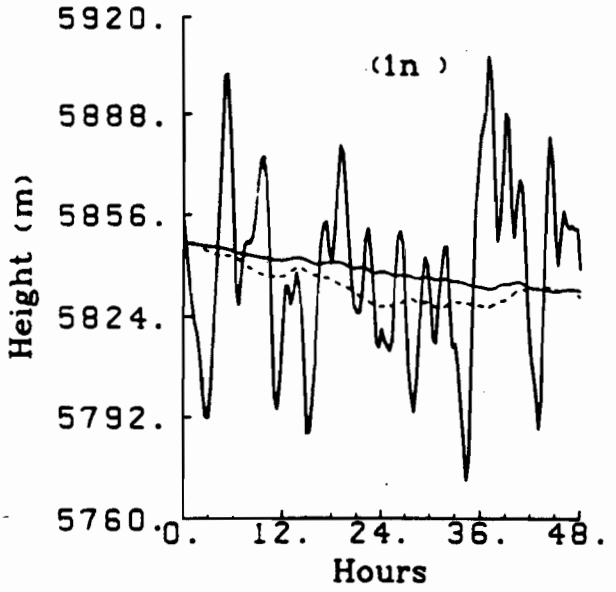


FIG. 15. Same as Fig. 14, but for point at (0.93°N, 180°).

same type of convergence problem as one faces with the iterative solution of the classical nonlinear balance equation. This process of height-constrained initialization can be represented in phase space as in Fig. 17. By this procedure, one attempts to force the wind field to balance a given height field without restrictions on the representativeness of the latter. This can be accomplished in the framework of section 3 by setting $\omega_\phi = 1$ and ω_ψ a "small" value. Strictly setting $\omega_\psi = 0$ creates a degeneracy of the Euler–Lagrange equations. For this extreme case, as independent formulation must be adopted. It's formulation is given in appendix B. It should be mentioned here that what is really meaningful is the *relative* value of these weights (refer to the functional being minimized) and in this sense the vanishing of weights can always be avoided. This study

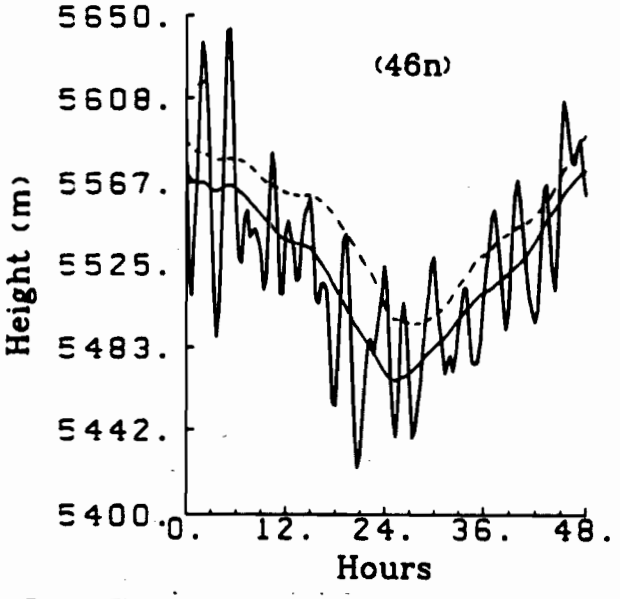


FIG. 14. Time trace of height field for point located at (45.70°N, 180°). Heavy line—no initialization, dashed line—unconstrained initialization, light line—variational initialization using latitude-longitude weights.

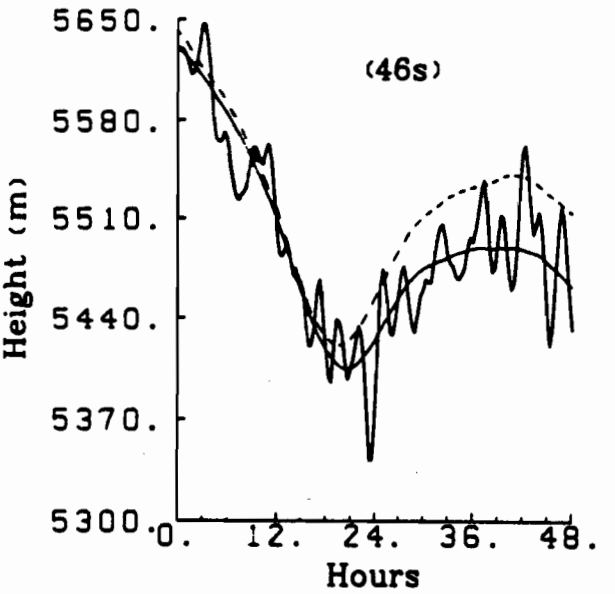


FIG. 16. Same as Fig. 14, but for point at (45.70°S, 180°).

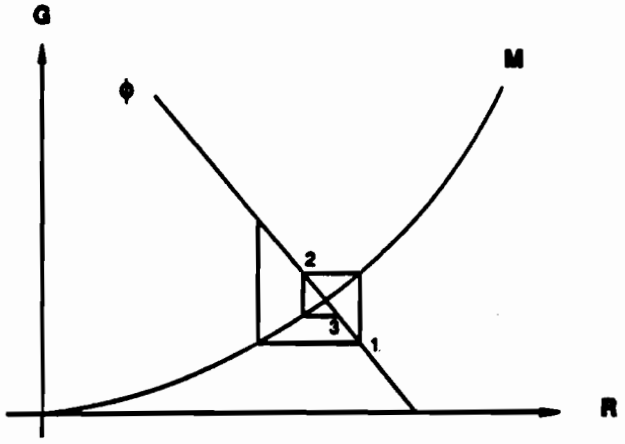


FIG. 17. Height-constrained initialization in phase space.

now examines the applicability of the variational scheme described in section 3 for small values of ω_ψ and it's relation with the ellipticity condition for the nonlinear balance equation. Houghton (1968) has given the correct form of the ellipticity condition for spherical coordinates and it's approximate form is given by

$$M = \nabla^2\phi + \frac{f^2}{2} - \frac{\beta}{a}\frac{\partial\psi}{\partial\theta} > 0 \qquad (5.7)$$

where

$$\beta = df/d\theta.$$

Condition (5.7) must be satisfied over the sphere for a solution of the nonlinear balance equation to exist. Kasahara (1982b) examined the observational evidence of nonelliptic regions in the atmosphere using the FGGE level IIIb global analyses of the ECMWF. He found that nonelliptic regions are widespread in the tropics. For the dataset mentioned in section 4a, Fig. 18 gives the regions where the ellipticity condition (5.7) is violated (i.e., regions of negative values of M). The tropics are indeed a typical region for failure of (5.7) and some negative regions at midlatitudes which are located in regions of anticyclonic flow are also noted. In order to test the behavior of the variational

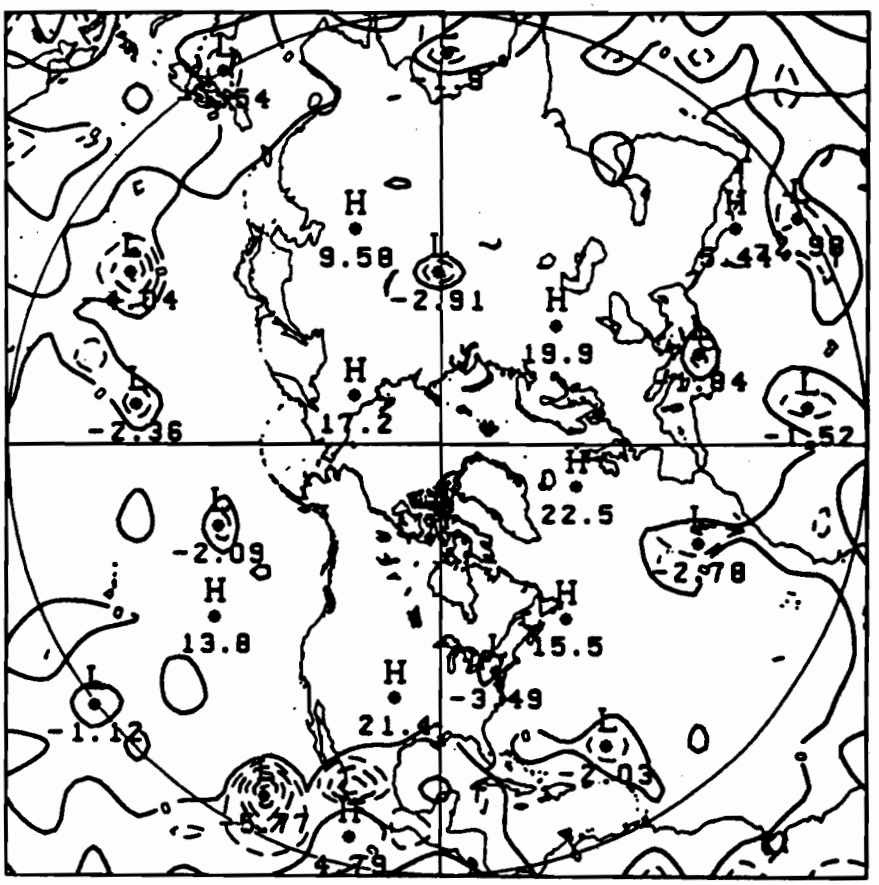


FIG. 18. Northern Hemisphere distribution of the ellipticity measure M as defined by (5.7). For clarity, only the location of the maximum values are indicated for positive values of M.

scheme in the context of height-constrained initialization, we choose the same initial conditions as in section 4a. It is noted there that Machenhauer's iterative scheme was converging and that the quantity BAL_1 could be reduced by 16 orders of magnitude after 10 iterations (see also T89). The unconstrained experiment will be used as a reference for comparison between the behavior of BAL_1 and height-constrained experiments using different values of ω_ψ. As previously noted in Fig. 18, setting ω_ψ sufficiently low in order to force the rotational part of the wind field to adjust to the geopotential field might cause convergence problems in the tropics. To show that this is indeed the case here, the following specification of weights on the sphere were tested:

$$\omega_\phi = 1, \quad \omega_\psi = \epsilon.$$

Figure 19 gives the behavior of BAL_1 as a function of iteration number for ϵ set to 0.1 and 0.01, respectively. It is clear that setting ω_ψ too low in the tropics does cause a divergence of the iterative scheme. The damaging effect on the wind field in the tropics as ϵ is reduced is not shown here but it can be said that a real "blow up" is observed in this region as the iteration progresses.

6. Summary and discussion

This paper has extended the application of variational implicit normal mode initialization (VINMI), developed by Fillion and Temperton (1989), to the context of a global spectral shallow-water model. The implicit "*B* scheme" considered by Temperton (1989) for the fast gravity modes was coupled as a strong constraint with a linear balance relationship for the slow modes. This scheme is the spectral analogue of the variational scheme developed by Fillion and Temperton (1989) for a barotropic finite-element regional model. Standard tests illustrated the robustness of the scheme to control the changes made to the analyzed fields using prescribed weights which are allowed to vary horizontally. The balance achieved by such a process was shown to be comparable to standard INMI. As illustrated in FT89, the VINMI method has the possibility of *significantly* altering the slowly evolving component of the forecast.

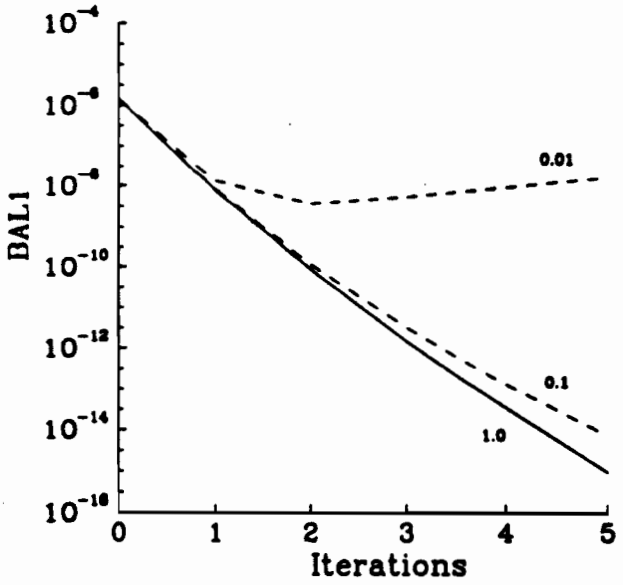


FIG. 19. Behavior of the quantity BAL_1 as a function of iteration for different values of the weight ω_ψ in the context of height-constrained initialization.

Present implicit NMI techniques impose a fundamental requirement, the stationarity of the slow modes in the retained linear system. This restriction on the slow modes was maintained in the variational framework developed here (see also FT89) and was shown to result in very similar changes on the divergent part of the analyzed flow during the balancing process. The question of whether or not a controlling procedure for the divergent part of the flow is feasible was not examined in this paper. This aspect of VINMI (if desirable) would depend on the feasibility of implicit NMI schemes in the case of nonstationary slow modes. Juvanon du Vachat (1988) examined this latter aspect and the characterization in physical space of the slow modes by the linear balance relationship [i.e., (2.15)] appears unavoidable for implicit schemes to be possible, thus, excluding the possibility to control explicitly the divergence field.

The question of height constrained adjustment using a special form of VINMI was considered here on the basis of the relationship between INMI and quasi-geostrophic theory on an *f* plane. The VINMI scheme proposed in this study was shown to diverge when the height field was given too strong a reliability over the wind field. For such a case, large changes to the wind field in the tropics are observed, which is related to the ellipticity condition for the nonlinear balance equation (see Tribbia 1981).

The present variational initialization technique may also be useful for variable resolution global spectral models (Courtier and Geleyn 1988). A future extension of the present work is to apply an analogous approach to multilevel models. These results will be presented in a subsequent paper.

Acknowledgments. The author wishes to thank Dr. Harold Ritchie for providing the basic spectral model used in this study, and Dr. Jean Côté for his review of the paper. Thanks are also due to the anonymous reviewers for their helpful suggestions which resulted in an improvement of the presentation of the paper. The author also thanks Diane Lespérance for typing part of the manuscript.

APPENDIX A

Unconstrained Initialization

Let $\omega_\phi = \omega_\psi$ = constant in the system (3.6)–(3.8), then write

$$\nabla^2(\Delta\psi) = \Delta\zeta = \frac{1}{\Phi\omega_\psi}\nabla\cdot(f\nabla\gamma).$$

Since $\nabla^2\gamma = \omega_\phi(\Delta\phi)$, [$\gamma^* = 0$ due to (3.5)] it follows that for every spectral component,

$$\Delta\zeta_n^m = \frac{2\Omega\omega_\phi}{\Phi\omega_\psi}\left[\frac{(n+1)}{n}\epsilon_n^m(\Delta\phi)_{n-1}^m + \frac{n}{(n+1)}\epsilon_{n+1}^m(\Delta\phi)_{n+1}^m\right].$$

Used scaled variables (2.4)–(2.6), it follows that

$$\begin{aligned}\Delta\hat{\zeta}_n^m &= \frac{a}{[n(n+1)]^{1/2}}\left(\frac{\omega_\phi}{\omega_\psi}\right)\left(\frac{2\Omega}{n}\epsilon_n^m\right)(n+1)\Phi^{-1/2} \\ &\quad\times(\Delta\hat{\phi})_{n-1}^m + \frac{\omega_\phi}{\omega_\psi}\left[\frac{2\Omega}{(n+1)}\epsilon_{n+1}^m\right] \\ &\quad\times\frac{na}{[n(n+1)]^{1/2}}\Phi^{-1/2}(\Delta\hat{\phi})_{n+1}^m \\ &= \frac{\omega_\phi}{\omega_\psi}\left\{\frac{2\Omega}{n}\epsilon_n^m(n^2-1)^{1/2}\left(\frac{a\Phi^{-1/2}}{[n(n-1)]^{1/2}}\right)\right. \\ &\quad\times(\Delta\hat{\phi})_{n-1}^m + \frac{2\Omega}{(n+1)}\epsilon_{n+1}^m[(n+1)^2-1]^{1/2} \\ &\quad\left.\frac{a}{[(n+1)(n+2)]^{1/2}}\Phi^{-1/2}(\Delta\hat{\phi})_{n+1}^m\right\}.\end{aligned}$$

Thus,

$$(\Delta\hat{\zeta})_n^m = \frac{\omega_\phi}{\omega_\psi}[f_n^m(C_{n-1}^m)^{-1}(\Delta\hat{\phi})_{n-1}^m + f_{n+1}^m(C_{n+1}^m)^{-1}(\Delta\hat{\phi})_{n+1}^m] \quad (A.1)$$

where the matrix elements f_n^m and c_n^m are given by (2.11) and (2.12), respectively. In matrix form (A.1) is given by

$$\Delta\hat{\zeta} = \frac{\omega_\phi}{\omega_\psi}\mathbf{F}\mathbf{C}^{-1}(\Delta\hat{\phi}). \quad (A.2)$$

If $\omega_\phi = \omega_\psi$ is used, (A.2) is identical to (4.18) of T89 of unconstrained initialization. Further, the differential constraint (3.2) has exactly the same form as (4.16) of T89. Summarizing, whenever $\omega_\phi = \omega_\psi$ = constant, the present variational scheme and the implicit (unconstrained) *B* scheme of T89 are identical.

APPENDIX B

Height-Constrained Initialization with $\omega_\psi = 0$, $\omega_\phi = 1$

For this special choice, the basic steps represented on Fig. 17 are summarized by:

1) Obtaining the balancing correction state vector on fast modes

$$(\Delta\mathbf{X})_G = [(\Delta\psi)_G, (\Delta\chi)_G, (\Delta\phi)_G].$$

2) Altering the slow mode amplitudes so as to reset ϕ to ϕ_I, where ϕ_I is the initial uninitialized height field.

The starting point of the scheme may be the initial state given by an analysis scheme or it's projection on the rotational part of the analyzed fields. In the framework of section 3, steps 1) and 2) can be done by the following procedure. First, step 1) generates a $(\Delta\phi)_G$ that step 2) must cancel exactly to reset ϕ to ϕ_I:

$$(\Delta\phi)_R = -(\Delta\phi)_G. \quad (B.1)$$

Here the slow modes satisfy a linear balance relationship

$$\nabla^2(\Delta\phi)_R = \nabla\cdot[f\nabla(\Delta\psi)_R] \quad (B.2)$$

and the following relation for fast mode components (i.e., the constraint from INMI) is shown:

$$\nabla^2(\Delta\phi)_G = \nabla\cdot[f\nabla(\Delta\psi)_G] - \frac{2\Omega}{a^2}\frac{\partial(\Delta\chi)_G}{\partial\lambda} + (\delta_t D)_0. \quad (B.3)$$

Equations (B.2), (B.3) correspond to their discretized versions (3.1), (3.2). Combining (B.1) to (B.3) gives for $(\Delta\psi)$:

$$\nabla\cdot[f\nabla(\Delta\psi)] = \frac{2\Omega}{a^2}\frac{\partial}{\partial\lambda}(\Delta\chi)_G - (\delta_t D)_0. \quad (B.4)$$

The strict height constrained algorithm, for one iteration of the implicit NMI scheme may thus be summarized for the present spectral model as:

1) Operating steps 1 and 2 of the VINMI scheme described in section 3 to obtain $(\Delta\hat{D})$ and the right-hand side of (B.4).
2) Solving (B.4) for $(\Delta\psi)$.
3) Adding increments $(\Delta\psi)$, $(\Delta\chi)_G$ to get initialized spectral coefficients. Note that the mass field is left unchanged here.
4) Generating new values of (ζ, D, ϕ) on the Gaussian grid.

REFERENCES

Andersen, J. H., 1977: A routine for normal mode initialization with nonlinear correction for a multilevel spectral model with triangular truncation. ECMWF Int. Rep. No. 15, 41 pp. [ECMWF, Shinfield Park, Reading, Berkshire, RG2 9AX, England.]

Baer, F., and J. Tribbia, 1977: On complete filtering of gravity modes through nonlinear initialization. *Mon. Wea. Rev.*, **105**, 1536–1539.

Ballish, B., 1980: Initialization, theory, and application to the NMC spectral model. Ph.D. thesis, University of Maryland, 151 pp. [Dept. of Meteorology, University of Maryland, College Park, MD 20742.]

Browning, G., A. Kasahara and H. O. Kreiss, 1980: Initialization of the primitive equations by the bounded derivative method. *J. Atmos. Sci.*, **37**, 1424–1436.

Burger, A. P., 1958: Scale considerations of planetary motions of the atmosphere. *Tellus*, **10**, 195–205.

Charney, J., 1955: The use of primitive equations of motion in numerical prediction. *Tellus*, **7**, 22–26.

Courtier, P., and J. F. Geleyn, 1988: A global model with variable resolution—application to the shallow-water equations. *Quart. J. Roy. Meteor. Soc.*, **114**, 1321–1346.

Daley, R., 1978: Variational nonlinear normal mode initialization. *Tellus*, **30**, 201–218.

Fillion, L., and C. Temperton, 1989: Variational implicit normal mode initialization. *Mon. Wea. Rev.*, **117**, 2219–2229.

Haltiner, G. J., and R. T. Williams, 1980: *Numerical Weather Prediction and Dynamic Meteorology*. John Wiley, 477 pp.

Hinkelmann, K., 1951: Der Mechanismus des meteorologischen Lärmes. *Tellus*, **3**, 285–296.

Houghton, D. D., 1968: Derivation of the elliptic condition for the balance equation in spherical coordinates. *J. Atmos. Sci.*, **25**, 927–928.

Juvanon du Vachat, R., 1988: Nonnormal mode initialization: Formulation and application to the inclusion of the beta terms in the linearization. *Mon. Wea. Rev.*, **116**, 2013–2024.

Kasahara, A., 1982a: Nonlinear normal mode initialization and the bounded derivative method. *Rev. Geophys. Space Phys.*, **20**, 385–397.

——, 1982b: Significance of nonelliptic regions in balanced flows of the tropical atmosphere. *Mon. Wea. Rev.*, **110**, 1956–1967.

Kreiss, H.-O., 1979: Problems with different time scales for ordinary differential equations. *SIAM J. Num. Anal.* **16**, 980–998.

——, 1980: Problems with different time scales for partial differential equations. *Comm. Pure Appl. Math.* **33**, 399–439.

Leith, C. E., 1980: Nonlinear normal mode initialization and quasi-geostrophic theory. *J. Atmos. Sci.*, **37**, 954–964.

Machenhauer, B., 1977: On the dynamics of gravity oscillations in a shallow-water model with application to normal mode initialization. *Contrib. Atmos. Phys.*, **50**, 253–271.

Mitchell, H., C. Charette, C. Chouinard and B. Brasnett, 1990: Revised interpolation statistics for the Canadian data assimilation procedure: Their derivation and application.

Moura, A. D., 1976: The eigensolutions of the linearized balance equations over a sphere. *J. Atmos. Sci.*, **33**, 877–907.

Sasaki, Y., 1958: An objective analysis based on the variational method. *J. Meteor. Soc. Japan*, **36**, 77–78.

Staniforth, A. N., and H. Mitchell, 1978: A variable-resolution finite-element technique for regional forecasting with the primitive equations. *Mon. Wea. Rev.*, **106**, 439–447.

Temperton, C., 1984: Variational normal mode initialization for a multilevel model. *Mon. Wea. Rev.*, **112**, 2303–2316.

——, 1988: Implicit normal mode initialization. *Mon. Wea. Rev.*, **116**, 1013–1031.

——, 1989: Implicit normal mode initialization for spectral models. *Mon. Wea. Rev.*, **117**, 432–447.

Tribbia, J. J., 1981: Nonlinear normal mode balancing and the ellipticity condition. *Mon. Wea. Rev.*, **109**, 1751–1761.

——, 1982: On variational normal mode initialization. *Mon. Wea. Rev.*, **110**, 455–470.

——, 1984: A simple scheme for higher-order nonlinear and normal mode initialization. *Mon. Wea. Rev.*, **112**, 278–284.

Wiin-Nielsen, A., 1979: On normal mode linear initialization on the sphere. *J. Atmos. Sci.*, **36**, 2040–2048.

Chapter V

Variational Implicit Normal Mode Initialization for a multilevel model

III.1 Presentation of article 3.

The complete extension of variational nonlinear normal mode initialization to fully three dimensional weights has never been attained. Previous efforts to examine the problem were essentially those of Temperton (1984). Temperton's scheme was formulated for a global grid point primitive equation model and his formulation allowed latitudinal variation of the weights (longitudinal variation could be included but with much reduced efficiency). Due to these limitations, Temperton's conclusion on the usefulness of variational NMI are of limited utility. Puri's (1983) study was also seriously limited to latitudinally varying weights. His choice of vertically separable weights is inappropriate for a full three dimensional treatment of analysis errors during the balancing process. Consequently, no reliable conclusion could be made concerning the usefulness of variational NMI in a context of operational data assimilation on the basis of these previous works. Using the formulation of INMI, Temperton's (1984) work is reconsidered. The following chapter is the culminating point of this study where it is shown that VINMI is feasible for multilevel models and includes a full horizontal and vertical treatment of the analysis error in a consistent manner. We present the application of the technique to the Canadian operational regional finite-element model. The formulation is general enough to be applied to other models where the vertical structure differs from the one used here. The formulation of the initialization scheme and results of applications follow.

V.2 Article 3

Variational implicit normal mode initialization for a multilevel model

Submitted to *Monthly Weather Review* , March 1991.

VARIATIONAL IMPLICIT NORMAL MODE INITIALIZATION FOR A MULTILEVEL MODEL

Luc Fillion and Michel Roch

Recherche en prévision numérique
Atmospheric Environment Service
2121 Trans-Canada Highway
Dorval, Québec, Canada H9P 1J3

Submitted to Monthly Weather Review

March 1991

ABSTRACT

Recent studies have demonstrated that variational nonlinear normal mode initialization can be efficiently implemented in the context of shallow-water models provided one uses a physical space formulation. The Implicit nonlinear Normal Mode Initialization (INMI) technique provides essentially the same balancing benefit as standard " explicit " nonlinear NMI but does not require the explicit computation of the linear free modes of the model. This allows variational initialization with arbitrary horizontal variation of the weights which specify the changes to the analyzed fields during initialization. As a consequence, land-sea contrast in the reliability of analyzed fields can be taken into account in the initialization step, an advantage for data assimilation for weather prediction.

The purpose of this paper is to demonstrate the feasibility of Variational Implicit Normal Mode Initialization (VINMI) for multilevel models. This new scheme is illustrated on the presently operational Canadian baroclinic Regional Finite-Element (RFE) model. It is shown that the VINMI scheme efficiently controls the relative magnitude of the changes to the analyzed mass and wind fields during the balancing (initialization) process. A comparison is also made of the impact of the VINMI scheme versus that of the presently operational unconstrained version of the initialization scheme (INMI). Future development and applications of the method are discussed at the end of the paper.

1. Introduction

Nonlinear normal-mode initialization (nonlinear NMI) schemes are now widely used in operational practice to control the excitation of spurious gravity-inertial oscillations. The application of these techniques in assimilation schemes for weather prediction permitted better assimilation of new data on a regular basis without creating subsequent imbalance of the model. On the other hand, systematic deficiencies of these initialization techniques were gradually isolated during the last decade in an attempt to improve the schemes. In fact, it appears that the main difficulties encountered in operational practice when the analysis-initialization schemes are used in sequence may be identified as a two-fold problem. The first aspect pertains to the general idea of balance in a model. Machenhauer's (1977) initialization scheme for example must be restricted to a subset of vertical modes of the model in order to avoid false characterization of this kind of "balance" in a baroclinic model (Errico 1989, Errico and Williamson 1988, Ko et al 1989). The second aspect of the problem is that the inherent inaccuracies in the analysis schemes i.e., errors in the slow mode components, which feed the initialization should be taken into account.

One way to face the latter problem in the analysis-initialization scheme was considered by Daley (1978). The essence of his variational approach is to permit slow and fast mode alteration of the analyzed fields during the nonlinear normal mode balancing process. This approach renders the initialization scheme flexible with respect to analysis errors. Daley's technique (also considered by Tribbia 1982) suffers from computational problems related to its dimensionality (i.e. the dimension of the linear system to be solved) which renders the method inapplicable in an operational context. For multilevel models, Puri (1983) examined the application of Daley's variational initialization in the context of data assimilation. Temperton (1984) reconsidered the variational initialization approach in the context of a multilevel gridpoint model but still had to restrict the weights to vary only latitudinally in order to render the problem tractable. It was then realized that the variational

form of nonlinear NMI including fully three-dimensional variation of the weights in the context of explicit nonlinear NMI would remain problematical due essentially to the normal mode space formulation of the approach. Originally, Daley (1978) proposed his variational scheme in order to bound the changes made by the initialization using the expected analysis errors. A much simpler procedure of "initialization-insertion" was examined by Machenhauer (1977) and Temperton and Williamson (1979). A similar type of reinsertion of uninitialized fields was tested for surface pressure by Bourke and McGregor (1983). Machenhauer (1977) first suggested that it may be appropriate to iterate the analysis-initialization procedure until some convergence is attained in order to remain faithful to the observations. As an alternative to variational initialization, Machenhauer's iterative procedure was formulated and tested by Williamson and Daley (1983) in the context of a shallow-water model. Various methods of combining analysis and balancing schemes may also be found in Lorenc (1986).

Using Temperton's (1988) formulation of INMI (see also Juvanon du Vachat 1988), variational initialization becomes more attractive than before due to the physical space formulation of the balancing constraints on the fast modes and the approximate characterization of the slow modes. The variational extension of INMI schemes has been successfully implemented in the context of shallow-water models by Fillion and Temperton (1989) (hereinafter noted FT89) for a regional finite-element model and Fillion (1991) for a global spectral model. It became possible with VINMI to deal with fully variable weights in the variational algorithms at a reasonable cost which indicated a potential usefulness for operational practice. In the present paper, we basically reconsider Temperton's (1984) approach in order to extend the VINMI scheme of FT89 to the multilevel version of the Canadian Regional Finite-Element (RFE) model.

Section 2 gives a quick review of the INMI scheme for the RFE model as described in Temperton and Roch (1991) (hereinafter noted TR91). The formulation and special characteristics of the VINMI scheme are discussed in Section 3. The flexibility of the variational scheme as compared to INMI is illustrated in Section 4. Finally Section 5

includes a summary and an outlook regarding future use of the method in a data assimilation scheme .

2. A review of the implicit NMI scheme

The governing equations of the RFE model are based on the hydrostatic primitive equations on a polar stereographic projection. Details of the spatial discretizations may be found in Staniforth and Daley (1979) and more recently in Tanguay et al. (1989) which also includes a description of the semi-Lagrangian version of the model. The parameterization of physical processes (taken into account during the forecast integration) are described in Benoit et al. (1989). Figure 1 illustrates the horizontal extent of the calculation grid (125 X 101 points) and the variable horizontal resolution used in this study. The uniform resolution on the subdomain is 100 km. Also shown are selected grid points where the time-traces of specific fields are stored during the model integration. The RFE model has a wall boundary condition at the horizontal boundaries which forces the fluid to move inside a square region D_h of side 20,000 km centered at the North Pole. This solid wall Γ is placed in the vicinity of the equator and the condition of no normal flow across Γ may be satisfied by requiring

$$\psi = 0 \qquad \text{on } \Gamma$$

$$\mathbf{n}\cdot\nabla\chi = 0 \qquad \text{on } \Gamma$$

where ψ and χ are the Helmholtz streamfunction and velocity potential respectively.

In the vertical, the model uses σ-coordinates and has 19 levels as shown in Fig. 2. Throughout the rest of the paper we will refer to D_h and (D) as the horizontal (resp. three dimensional) domain of the model. The vertical boundary conditions are obtained by imposing the condition of no outflow at the top and bottom of the model. This dynamical requirement is characterized by

$$\dot{\sigma} = 0 \quad \text{at} \quad \sigma = \sigma_1 , 1 \quad \text{where} \quad \sigma_1 = 0.050 \quad . \qquad (2.1)$$

Following TR91, the differentiated form of the primitive equations (at each level "k" in the vertical) used for the formulation of the unconstrained implicit initialization scheme is

$$\frac{\partial}{\partial t}(\nabla^2\psi_k) = -F\chi_k + (Q_\psi)_k \qquad (2.2a)$$

$$\frac{\partial}{\partial t}(\nabla^2\chi_k) = F\psi_k + B\chi_k - \nabla^2 P_k + (Q_\chi)_k \qquad (2.2b)$$

$$\frac{\partial T_k}{\partial t} - \kappa T^* \left(\frac{W}{\sigma}\right)_k = (Q_T)_k \qquad (2.2c)$$

and

$$F = \frac{\partial}{\partial x}\left(f\frac{\partial}{\partial x}\right) + \frac{\partial}{\partial y}\left(f\frac{\partial}{\partial y}\right)$$

$$B = \frac{\partial}{\partial x}\left(f\frac{\partial}{\partial y}\right) - \frac{\partial}{\partial y}\left(f\frac{\partial}{\partial x}\right)$$

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \quad .$$

The generalized geopotential P_k is defined here as $\phi_k + RT^* \ln p_s$, where p_s is the surface pressure, and T^* is the isothermal reference temperature profile. The variable W is related to the divergence field (see Tanguay et al. 1989, equation (29)). The hydrostatic equation

$$\frac{\partial \phi}{\partial \sigma} = -\frac{RT}{\sigma}$$

is vertically discretized as

$$\phi_{k+1} - \phi_k = \frac{1}{2} R (T_{k+1} + T_k) \ln\left(\frac{\sigma_{k+1}}{\sigma_k}\right) \quad , \quad 1 \le k \le N-1, \qquad (2.3)$$

where $\phi_N = \phi_s$ and N is the number of levels of the model. Note that the $B\psi_k$ term normally appearing in the vorticity equation (2.2a) has been incorporated into the nonlinear terms $(Q_\psi)_k$ for the formulation of the implicit initialization scheme. We also introduce the integrated tendency equation for the surface pressure

$$\frac{\partial}{\partial t}(\ln p_s) - W_N = Q_{(\ln p_s)} \quad . \qquad (2.4)$$

The nonlinear terms are represented by the symbol Q, the other variables have their usual meaning (see Tanguay et al. 1989, and FT89).

In order to apply the implicit initialization scheme, it is still necessary to make use of a vertical mode decomposition. These modes are determined by diagonalizing the matrix **C** which appears in the matrix/vector form of the prognostic equation for the generalized geopotential (equation (2.16) of TR91), i.e.

$$\frac{\partial \mathbf{P}}{\partial t} + \mathbf{C}\,\mathbf{D} = \mathbf{Q}_P \quad . \tag{2.5}$$

The detailed structure of matrix **C** may be found in TR91. The decomposition of matrix **C** is

$$\mathbf{C} = \mathbf{E}\,\Phi\,\mathbf{E}^{-1} \tag{2.6}$$

where Φ is a diagonal matrix whose entries are the geopotential depths, and the columns of **E** are the eigenvectors of **C**. Using these vertical normal modes as a basis for vertical representation of the dependent variables, the prognostic system (2.2a-b), and (2.5) may be written as

$$\frac{\partial \hat{P}_n}{\partial t} = -\,m^2\,\Phi_n\,\nabla^2\hat{\chi}_n + \left(\hat{Q}_P\right)_n \tag{2.7a}$$

$$\frac{\partial}{\partial t}\left(\nabla^2\hat{\psi}_n\right) = -\,F\,\hat{\chi}_n + \left(\hat{Q}_\psi\right)_n \tag{2.7b}$$

$$\frac{\partial}{\partial t}\left(\nabla^2\hat{\chi}_n\right) = F\,\hat{\psi}_n + B\,\hat{\chi}_n - \nabla^2\hat{P}_n + \left(\hat{Q}_\chi\right)_n \tag{2.7c}$$

for each vertical mode " n ". System (2.7) has the same form as the shallow-water version of the RFE model if the geopotential ϕ is replaced with the generalized geopotential P, and the mean geopotential depth is associated with a particular eigenvalue of matrix **C**. This form of the equations is consistent with the splitting of the linear and nonlinear terms considered by Temperton (1988) for the derivation of his improved INMI scheme.

The way to implement the unconstrained implicit initialization scheme for the baroclinic RFE model is clear ; for each vertical mode, apply the INMI scheme as described in Temperton (1988) for the shallow-water version of the RFE model.

3. Variational formulation

Nonlinear NMI schemes for multilevel models are basically formulated in terms of the variable P, which couples the temperature and surface pressure fields. The variational formulation that follows will thus be considered within the framework of the initialization scheme itself which explicitly makes use of the generalized geopotential P.

Using the considerations of Section 2, we now introduce a variational extension of this scheme where the *constraint* and the *functional* to be minimized are successively introduced. The variational formulation of the implicit initialization technique for shallow-water models is as described in FT89 and Fillion (1991).

a. The constraint

The analog of the barotropic constraint on the fast modes (see (2.9) and (2.10) of FT89) is that, for each vertical mode, for one iteration of Machenhauer's scheme, the linear constraint on $(\Delta\hat{P})^n_G$ and $(\Delta\hat{\psi})^n_G$ is given by

$$\nabla^2(\Delta\hat{P})^n_G = F\,(\Delta\hat{\psi})^n_G + B\,(\Delta\hat{\chi})^n_G + (\delta_t\nabla^2\hat{\chi})^n_o \qquad (3.1a)$$

with boundary conditions:

$$\frac{\partial(\Delta\hat{P})^n_G}{\partial x} = f\frac{\partial(\Delta\hat{\psi})^n_G}{\partial x} + f\frac{\partial(\Delta\hat{\chi})^n_G}{\partial y} \quad \text{on } \Gamma_y \qquad (3.1b)$$

$$\frac{\partial(\Delta\hat{P})^n_G}{\partial y} = f\frac{\partial(\Delta\hat{\psi})^n_G}{\partial y} - f\frac{\partial(\Delta\hat{\chi})^n_G}{\partial x} \quad \text{on } \Gamma_x \qquad (3.1c)$$

where $\Gamma = \Gamma_x \cup \Gamma_y$, and the index "G" indicates a gravity mode component. The time tendency term appearing in (3.1a) is obtained as usual by integrating the model forward for one timestep.

Since we are concerned here with the extension of the INMI scheme of TR91, we restrict the application of (3.1) to the first three vertical modes of the model, say Ω, as is presently done operationally with the INMI scheme. As mentioned in Section 1, the

adjustment of the gravity modes using Machenhauer's scheme must be restricted to a subset of vertical modes.

In a manner similar to that in FT89, for a given vertical mode of the multilevel model and for one iteration of the initialization scheme, we make use of the linear balance relationship

$$\nabla^2(\Delta\hat{P})^n_R = F(\Delta\hat{\psi})^n_R \tag{3.2a}$$

with boundary conditions

$$\frac{\partial(\Delta\hat{P})^n_R}{\partial x} = f\frac{\partial(\Delta\hat{\psi})^n_R}{\partial x} \quad \text{on } \Gamma_y \tag{3.2b}$$

$$\frac{\partial(\Delta\hat{P})^n_R}{\partial y} = f\frac{\partial(\Delta\hat{\psi})^n_R}{\partial y} \quad \text{on } \Gamma_x \tag{3.2c}$$

to characterize the slow mode corrections. It is important to note at this point that the correction of the divergent part of the wind field appearing in (3.1), for each iteration, is diagnosed in the same way as for the unconstrained INMI scheme. This is due to the fact that the slow modes characterized by (3.2) are divergence free. Thus, for each iteration, $(\Delta\chi)$ projects only onto the restricted class of vertical modes Ω. While the same restriction is applied to $(\Delta\psi)$ and (ΔP) for the INMI scheme, this is no longer the case within the present variational framework. The physical space translation of the latter statement is obtained by coupling the linear equations (3.1) and (3.2) to form the constraint in three dimensional physical space, i.e.

$$M(\Delta P, \Delta\psi) = -\nabla^2(\Delta P) + F(\Delta\psi) + B(\Delta\chi)_G + (\delta_t\nabla^2\chi)_o = 0 \tag{3.3a}$$

with boundary conditions

$$\frac{\partial(\Delta P)}{\partial x} = f\frac{\partial(\Delta\psi)}{\partial x} + f\frac{\partial(\Delta\chi)_G}{\partial y} \quad \text{on } \Gamma_y \tag{3.3b}$$

$$\frac{\partial(\Delta P)}{\partial y} = f\frac{\partial(\Delta\psi)}{\partial y} - f\frac{\partial(\Delta\chi)_G}{\partial x} \quad \text{on } \Gamma_x \ . \tag{3.3c}$$

Note that the terms $(\Delta\chi)_G$ and $(\delta_t\nabla^2\chi)_o$ only have components in the class Ω of vertical modes which is used in unconstrained INMI. With this last proviso, (3.3) is taken as a strong constraint in physical space.

b. The functional

The choice of a cost-functional in the context of three-dimensional (3-D) VINMI proceeds as follows. It is known in the context of shallow-water theory (see Daley 1978, Temperton 1984, Fillion and Temperton 1989), that each iteration of Machenhauer's initialization scheme minimizes a functional J_n which is proportional to the energy of the linearized system, and given by

$$J_n = \int_{D_R} \left[(\nabla \Delta\hat{\psi}_n)^2 + (\nabla \Delta\hat{\chi}_n)^2 + \frac{1}{\Phi_n}(\Delta\hat{P})_n^2 \right] dx\, dy \qquad (3.4)$$

where Φ_n is an equivalent geopotential depth. The VINMI scheme as formulated by FT89, does not take into account the divergent part of the wind due to the use of nondivergent slow modes. For the multilevel model used in the present study, the energy functional (3.4) is (see Staniforth et al. 1985):

$$\Delta E_R = \int_D \left[\frac{1}{m^2} \left\{ \frac{\sigma}{R\gamma^*} \left(\frac{\partial(\Delta P)}{\partial\sigma} \right)^2 \right\} + (\nabla(\Delta\psi))^2 \right] dx\, dy\, d\sigma$$

$$+ \int_{D_R} \left[\frac{1}{RT^* m^2} (1 - \sigma_1)(\Delta P)^2 \right] \Big|_{\sigma=1} dx\, dy$$

where the geometry and boundary conditions of the RFE model were taken into account. The basic state static-stability γ^* is given by $\kappa T^*/\sigma$ and the subscript "R" indicates that only the rotational part of the kinetic energy changes are considered in the minimization process. For multilevel models, it is natural to choose as the functional to be minimized, the generalized energy-consistent functional associated with the linearized baroclinic primitive equations (see Temperton, 1984). By properly adjusting the energy functional ΔE_R, we

show in the Appendix that the correct functional to be minimized, which includes weighting functions and maintains (as a special case, see Section 3d) the natural property described previously, is

$$J = \int_D \left[\frac{\omega_T}{m^2} \left\{ \frac{\sigma}{R\gamma^*} \left(\frac{\partial(\Delta P)}{\partial \sigma} \right)^2 \right\} + \omega_\psi \, (\nabla(\Delta\psi))^2 + \frac{\omega_T}{RT^* m^2} (\Delta P)_s^{\,2} \right] dx\, dy\, d\sigma \quad . \quad (3.5)$$

The weighting functions ω_ψ and ω_T are associated with the rotational part of the wind and mass fields respectively and can vary horizontally and vertically within D. The functional J can also be written as

$$J = \int_D \left[\frac{\omega_T}{m^2} \left\{ \frac{c_P}{T^*} (\Delta T)^2 \right\} + \omega_\psi \, (\nabla(\Delta\psi))^2 \right] dx\, dy\, d\sigma$$

$$+ \int_{D_h} \frac{RT^* \omega_T}{m^2} (\Delta \ln p_s)^2 \, dx\, dy \quad - \int_{D_h} \frac{[\omega_T \sigma]|_{\sigma=\sigma_1}}{RT^* m^2} (\Delta P)_{\sigma_1}^2 \, dx\, dy \quad .$$

The solution of the variational problem is however formulated in terms of ΔP from which temperature and surface pressure changes are obtained (see Subsection 3c).

c. The variational problem

The variational problem to be solved thus becomes the minimization of functional J given by (3.5) subject to the constraint (3.3) together with the vertical boundary conditions

$$\frac{\partial(\Delta P)}{\partial \sigma} + \frac{\gamma^*(\sigma)}{T^*} (\Delta P)_s = 0 \quad \text{at } \sigma = \sigma_1, 1 \quad . \quad (3.6)$$

The vertical boundary conditions (3.6) at the top and bottom of the model, are *natural* boundary conditions for the minimization of (3.5) and correspond to the boundary conditions which are consistent with the determination of the vertical normal modes of the model. Using the standard procedure of the calculus of variations (see the Appendix), the Euler-Lagrange equations are given by :

$$\nabla^2 \lambda = \frac{-1}{m^2} \frac{\partial}{\partial \sigma} \left[\frac{\omega_T \sigma}{R \gamma^*} \frac{\partial(\Delta P)}{\partial \sigma} \right] \quad (3.7a)$$

$$\left(\frac{\partial \lambda}{\partial n}\right)_\Gamma = 0 \qquad (3.7b)$$

$$\nabla\cdot(\omega_\psi \nabla(\Delta\psi)) = F(\lambda) \qquad (3.8a)$$

$$(\Delta\psi)_\Gamma = 0 \qquad (3.8b)$$

together with (3.3) and (3.6).

d. Particular cases

We first note that if the weights ω_T and ω_ψ are assumed to be independent of the vertical coordinate σ, the right-hand side of (3.7a) can be written as

$$-\frac{\omega_T}{m^2}\frac{\partial}{\partial\sigma}\left[\frac{\sigma}{R\gamma^*}\frac{\partial(\Delta P)}{\partial\sigma}\right] \equiv -\frac{\omega_T}{m^2}\,\mathcal{V}(\Delta P) \qquad (3.9)$$

where the vertical structure operator $\mathcal{V}$ is defined by (A4). The expansion in terms of vertical normal modes for the correction state vector ΔX and the Lagrange multiplier λ is given by

$$\Delta X = \begin{Bmatrix} \Delta\psi \\ \Delta\chi \\ \Delta P \end{Bmatrix} = \sum_n \begin{Bmatrix} \Delta\hat{\psi}_n \\ \Delta\hat{\chi}_n \\ \Delta\hat{P}_n \end{Bmatrix} \xi_n(\sigma) \quad ; \quad \text{and} \quad \lambda = \sum_n \hat{\lambda}_n \xi_n \qquad (3.10)$$

where ξ_n is a given vertical mode of the model. Using the fact that the weights are independent of σ, (3.3), (3.6), (3.7) and (3.8) are separable for each vertical mode " n " and may be written as:

$$\nabla^2\hat{\lambda}_n = \frac{\omega_T}{m^2\,\Phi_n}(\Delta\hat{P})_n \qquad (3.11a)$$

$$\left(\frac{\partial\hat{\lambda}_n}{\partial n}\right)_\Gamma = 0 \qquad (3.11b)$$

$$\nabla\cdot(\omega_\psi \nabla(\Delta\hat{\psi})_n) = F(\hat{\lambda}_n) \qquad (3.12a)$$

$$((\Delta\hat{\psi})_n)_\Gamma = 0 \quad . \qquad (3.12b)$$

The natural boundary conditions (3.6) are automatically satisfied by the vertical mode decomposition. Note also that the solution of the Euler-Lagrange equations is restricted to the subset Ω due to the vertical mode restriction discussed in Section 3a. The preceding set of equations is analogous to the set of Euler-Lagrange equations obtained for the shallow-water version of the model. Equations (3.11) and (3.12) can formally be obtained by minimizing a variational integral of the form

$$J_n = \int_{D_h} \left[\frac{\omega_T}{\Phi_n m^2} (\Delta \hat{P})_n^2 + \omega_\psi \left[\nabla (\Delta \hat{\psi})_n \right]^2 \right] dx\, dy \quad .$$

By multiplying J_n by Φ_n (which does not affect the solution) we obtain the same type of functional used by FT89 and Fillion (1991) for shallow-water models. By further restricting the weights to be constant and equal, $\omega_T = \omega_\psi$ = constant, we obtain the INMI scheme for a multilevel model as described by TR91.

e. The numerical procedure

The numerical procedure characterizing the VINMI scheme in general, for one iteration, consists of the following steps,

1) A forward timestep is done to compute the tendencies

$$\frac{\partial(\nabla^2 \psi)}{\partial t} , \frac{\partial(\nabla^2 \chi)}{\partial t} , \frac{\partial T}{\partial t} , \frac{\partial(\ln p_s)}{\partial t} \quad .$$

2) For the restricted set Ω of vertical modes (see Section 3a), compute

$$\frac{\partial(\nabla^2 \hat{\psi}_n)}{\partial t} , \frac{\partial(\nabla^2 \hat{\chi}_n)}{\partial t} , \frac{\partial \hat{P}_n}{\partial t} \quad ; \quad n = 1, 2, 3 \quad .$$

3) For each vertical mode "n", determine $\Delta \hat{\chi}_n$ from Temperton's 1988 INMI scheme for the shallow-water version of the RFE model.

4) Use the inverse vertical transform to determine the vertically filtered fields $\Delta\chi$ and $\delta_t \nabla^2 \chi$, i.e.

$$\Delta\chi = \sum_{n=1}^{3} \Delta\hat{\chi}_n(x,y)\,\xi_n(\sigma) \quad ; \quad \delta_t\nabla^2\chi = \sum_{n=1}^{3} \delta_t\left(\nabla^2\hat{\chi}_n\right)\xi_n(\sigma) \quad .$$

5) Solve the Euler-Lagrange equations, i.e. (3.3), (3.6), (3.7) and (3.8) for $\Delta\psi$ and ΔP in physical space.

6) Compute ΔU and ΔV from $\Delta\psi$ and $\Delta\chi$ using the same procedure described in Section 2b of FT89.

7) Use the procedure described in TR91 to determine $\Delta(\ln p_s)$ and ΔT from ΔP.

We note that the VINMI scheme formulated here implicitly controls the changes to the temperature and pressure fields. Daley's (1979) variational inversion procedure to obtain ΔT and $\Delta \ln p_s$ from ΔP was not considered here but we rather used the procedure described by TR91 (where the last vertical mode is excluded in the inversion). This latter procedure as will be shown in Section 4, does not destroy the implicit control on T and $\ln p_s$ forced by the VINMI scheme. From (3.5), we note that the same weight ω_T is used for temperature and surface pressure changes at the lowest level of the model. This point will be discussed further in Section 4.

The numerical procedure for solving the Euler-Lagrange equations is based on the strategy adopted for shallow-water models. The idea is to form a single equation to be solved for (ΔP). This equation is obtained by combining (3.7), (3.8) and (3.6) in that order, and is written as

$$\nabla^2(\Delta P) - F\,L_d^{-1}\,F\,\nabla_n^{-2}\,I\,(\Delta P) = B(\Delta\chi)_G + (\delta_t\nabla^2\chi)_o \qquad (3.13)$$

where the vertical operator I is defined as

$$I \equiv -\frac{\partial}{\partial\sigma}\left(K\,\frac{\partial(\Delta P)}{\partial\sigma}\right) \quad .$$

The coefficient K is given by

$$K \equiv \frac{\omega_T\,\sigma}{R\,\gamma^*\,m^2}$$

or using the definition of γ^* (considering the chosen isothermal basic state temperature profile) as

$$K \equiv \frac{\omega_T \sigma^2}{m^2 R \kappa T^*} \quad .$$

The linear (but variable-coefficient) operator L is defined as in FT89, i.e.

$$L \equiv \nabla \cdot \omega_\psi \nabla$$

where the indices "n" and "d" appearing in (3.13) refer to an inversion using homogeneous Dirichlet or Neumann boundary conditions on Γ. The operators L and I have the form of horizontal and vertical diffusion operators with variable diffusion coefficients ω_ψ and K respectively. This is due to the fact that we attempt to minimize the square of the horizontal and vertical gradient of $\Delta\psi$ and ΔP respectively, based on the principle of minimization of the linearized "energy" functional (3.5). Generalizing the iterative approach taken by FT89, a suitable (vertically separable) kernel operator for (3.13) is

$$\nabla^2 + \lambda_o^2 \frac{\partial}{\partial\sigma}\left(\frac{\sigma}{R\gamma^*}\frac{\partial}{\partial\sigma}\right)$$

where

$$\lambda_o^{\,2} \equiv \frac{1}{2}\left[\left(\frac{f^2\omega_T}{m^2\omega_\psi}\right)_{Max} + \left(\frac{f^2\omega_T}{m^2\omega_\psi}\right)_{Min}\right] \quad .$$

The minimax value λ_o^2 is evaluated over *D*. Each iteration of the numerical algorithm used to solve (3.13) requires

a) the computation in finite-element space of

$$\nabla^2(\Delta P)^{(k)} - F L_d^{-1} F \nabla_n^{-2} I (\Delta P)^{(k)}$$

for a given estimate $(\Delta P)^{(k)}$ at iteration "k".

b) a vertical mode decomposition of the result of step (a).

c) the solution of the kernel problem by inverting the constant coefficient Helmholtz operator

$$\nabla^2 - \frac{\lambda_o^2}{\Phi_n}$$

for each vertical mode, excluding the last one since this mode is degenerate, i.e. $\Phi_n = 0$ for n = 19. In practice, due to the vertical mode projection (onto Ω) of the right-hand side terms of (3.13), it is found that using only the first 10 vertical modes to solve (3.13) is enough to approximate accurately the solution obtained using the full 18 vertical modes. The structure of the first 6 vertical modes of the model may be found in TR91. Figures 3a-b show the structure of vertical modes 7-10 of the model.

The iterative scheme for solving (3.13) is accelerated by a conjugate-gradient technique (Concus et al. 1976).

4. Results

We now illustrate the results obtained with the variational scheme described in Section 3. The model itself uses a semi-Lagrangian time integration scheme (Tanguay et al. 1989) with a timestep of 20 minutes while the initialization scheme uses a timestep of 2 minutes to compute the time tendencies required during the iteration cycle. The initialization is performed on the variable resolution grid shown in Fig. 1. The experiments of this section (i.e. INMI or VINMI) used three iterations and used three vertical modes to alter the *gravity* modes. The initialization schemes are adiabatic, i.e. without physical processes. In terms of the percentage of the computer time required to integrate the model for 48 hours, the INMI and VINMI schemes require approximately 2 % and 8 % respectively.

a. Initial conditions

The initial conditions for the following experiments came from the interpolation of analyses done on an hemispheric Gaussian grid to the variable resolution grid of the RFE model. We note that in the present version of the model (the immediate successor of the version used by TR91), no smoothing procedure is applied on the height and wind fields near the boundaries (i.e. south of 20° latitude). The condition of no outflow mentioned in Section 2 is still imposed on the analyzed wind. Figure 4 shows the mean sea level pressure for 00 UTC 22 January 1986 before initialization.

b. Spatial distribution of analysis errors

Altough the present study does not address the problem of the optimal use of the available error statistics (variances and covariances) of the analyzed fields within a VINMI procedure, it is instructive to examine the spatial structure of the root-mean-square expected analysis error. These fields indicate the expected level of accuracy of the analyses and serve as a basis of comparison for the changes done to the analysis by the initialization.

Figures 5a-f give the root-mean-square analysis error for the height, zonal wind component, and temperature over the domain D_h at 850 and 250 mb. These fields are available from the statistical interpolation procedure used at CMC (Centre Météorologique Canadien) (Mitchell et al. 1990) to do operational data assimilation. As expected, for all these fields, the land-sea contrast in the expected analysis error is clearly evident.

The following parts of this section will focus on the ability of the variational scheme to respond adequately to spatially varying weights even if this happens to be in a region where the INMI scheme already produces absolute changes below the expected analysis error variance. A coherent specification of the variational weights using analysis error statistics must be performed prior to the application of VINMI in a complete operational data assimilation scheme.

c. Horizontal control

We first examine the application of the variational scheme when the weights ω_T and ω_Ψ are independent of the vertical " σ " coordinate. It was shown in Section 3d that the VINMI scheme for that case reduces to the VINMI scheme of FT89 for each vertical mode considered. We will refer to this particular form as VINMI-2D in the following discussion. We begin by giving a simple illustration of the behaviour of the scheme when the same set of "latitude-longitude" weights as those used in FT89 in a shallow-water context are used for the multilevel version of the model. These weights are useful to illustrate without ambiguity the relative changes done to the mass and wind fields by the VINMI-2D scheme as compared to the INMI scheme.

These weights are expressed as

$$\omega_\Psi = H_\Psi^\lambda \cos^8(\theta)$$

$$\omega_T = H_T^{\lambda}\left[1 - \cos^8(\theta)\right]$$

where H^{λ}_T and H^{λ}_{ψ} are set to unity except when the longitude λ is over ocean where they change abruptly to 0.01 and 0.1 respectively. These delimiting longitudes are the same as in FT89, i.e.

Atlantic Ocean (15°W, 65°W)

Pacific Ocean (125°W, 135°E) .

Using these weights, we expect that the mass field will adjust to the wind field over the tropics and vice-versa over extratropical continental regions. The abrupt horizontal change will serve to test the robustness of the VINMI-2D scheme while at the same time force the initialization scheme to recognize the different level of accuracy of the analyzed fields over the continents and oceans. The need for such a distinction is evident from Figs. 5a-f.

We start by discussing the convergence characteristics of both unconstrained and constrained iterative initialization schemes, leaving the details of the controlling aspect of the VINMI-2D scheme to the end of this sub-section. To exhibit the reduction of the time-tendencies of the fast modes for both INMI and VINMI-2D schemes, we show in Figs. 6a and 6b the reduction in the value of BAL for each vertical modes of Ω as a function of the iteration number. The value of BAL (computed in physical space) is a sum of squares of the tendencies of the fast mode coefficients (ref. Temperton 1988). For the external and first internal modes, the reduction rate of BAL is essentially the same for both schemes. For the second internal mode however, the reduction rate is slightly slower for the constrained experiment. As will be shown later, this results in imperceptible imbalances during the ensuing forecast when compared to the balance resulting from the unconstrained INMI scheme. Another diagnostic measure of the convergence of both initialization schemes is now given in terms of root-mean-square increments computed for each iteration. Figures 7a-

b show these increments in terms of the variable P for both INMI and VINMI-2D schemes, for each sigma level of the model, while Figs. 8a-b exhibit the rms increments of the wind field. It is seen that both schemes converge quite rapidly with rms increments of P and wind at the end of the third iteration less than 10 $m^2 s^{-2}$ and 0.2 ms^{-1} respectively.

We now examine the controlling aspect of the VINMI-2D scheme. Height field changes will be shown at $\sigma = 0.25$, which is typical. The changes at other levels are similar but with a reduced amplitude at lower levels due to the absence of vertical variation of the weights. Figure 9a shows the differences in height field between INMI and no initialization for an area including the uniform high-resolution grid. We observe relatively uniform height change magnitudes (~ 20 metres) over continental and ocean regions. Assuming the previously defined weights, i.e. a highly reliable analyzed height field over the continent and low reliability over oceans, the VINMI-2D scheme produces the changes shown in Fig. 9b. Essentially the analyzed extratropical height field has been maintained over the continents and changes forced preferentially over the adjacent oceans. It is also important to look at the wind changes. Figure 10 shows the height and wind field differences between VINMI-2D and INMI initialized fields. These fields are clearly in quasi-geostrophic balance and essentially limited to the north-American continental region. This indicates that the VINMI-2D scheme has also changed the slow modes of the analysis in order to retain the analyzed mass field over the continent. Due to the constraint imposed by the present specification of weights, the variational scheme maintained the analyzed mass field at the expense of the rotational part of the wind field over the extratropical continents. Together with our choice of ω_T in the functional (3.5), this implies that the initialized surface pressure field should be closer to the analysis when the VINMI-2D, rather than the INMI scheme is used. Of course, with the INMI scheme there is no such control of the surface pressure changes. TR91's inversion procedure is used in the VINMI algorithm to deduce ΔT and $\Delta \ln p_s$ from ΔP, and this procedure does not destroy the control operated implicitly on the former fields via ΔP. This is particularly evident for surface pressure (convincing results for temperature changes

are shown in the next sub-section) when Fig. 11a (INMI case) and Fig. 11b (VINMI-2D case) are compared. These figures show the surface pressure changes made by each initialization scheme. The latitudinal variation of the weights over the continent strongly forces the retention of the mass field north of 40° latitude. This effect is observed in Fig. 11b. At lower latitudes, over the continents, the constraint on the mass field is gradually relaxed in favor of the rotational part of the wind field.

We now show the impact of the VINMI-2D scheme on the vertical velocity field. Figure 12a shows the field of vertical velocity ω at 700 mb before initialization. This field is noisy over the Rocky mountains and Greenland. Elsewhere the structure of the field of ω is synoptically unacceptable. For example consider the low pressure system over the Great Lakes, with a trough extending northward to Hudson bay (see Fig. 4). The value of ω is positive (i.e. downward motion) in this region and negative (i.e. upward motion) over the ridge associated with the high system over the Northwest Territories. After initialization (VINMI-2D), the ω field is synoptically acceptable as judged from Fig. 12b. Note also the narrow band of upward vertical motion west of Alaska, a feature noticeable before initialization but refined by the initialization scheme. The ω field for INMI is not shown since it is essentially identical to the VINMI-2D case (see also FT89 and Fillion 1991 concerning the initialized divergent circulation). In Figs. 13a-c, we show the evolution of the surface pressure at points 1, 2, and 3 respectively (see Fig. 1 for the location of these points). The variational scheme has eliminated the high-frequency oscillations just as well as the unconstrained scheme. Point 3 lies within the region of control of the mass field, and this is apparent in Fig. 13c where the initial surface pressure field for the uninitialized and variationally initialized cases are the same. The evolution of the surface pressure at this point exhibits the familiar behaviour already noted for the height field traces with shallow-water models (see Fig. 8 of FT89 and Fig. 16 of Fillion 1991). This characteristic signature reflects fast oscillations around a slowly-evolving meteorologically-significant state. The

variationally initialized case however shows a slow evolution which differs from the unconstrained case.

d. Three dimensional control

Puri (1983) studied the impact of variational initialization in the context of data assimilation using a multilevel spectral model. His study was severely restricted to latitudinally varying weights thus ignoring the very sharp variation in analysis accuracy between oceans and continents seen in Figs. 5a-f. The VINMI-2D scheme described in the preceding sub-section does not have this limitation. Also, an important result reported by Puri (1983) is the improvement of the assimilation cycle by allowing the weights to vary in the vertical. This was achieved in his study by assigning lower weights to the higher vertical modes initialized. As mentioned by Puri, this kind of specification of the weights is inappropriate in order to reflect the availability of data and analysis reliability in the vertical. The VINMI scheme developed in Section 3 resolves this problem and allows a full variation of the weights in the vertical.

We now examine the behaviour of the VINMI-3D scheme when vertical variation of the weights is introduced. For this purpose, we define

$$\omega_T = H_T^{\lambda} H_T^{\sigma} \left[1 - \cos^8(\theta)\right]$$

where H^{λ}_T is defined as before. We perform two experiments as follows

$$\text{Experiment 1} \quad : \quad H_T^{\sigma} = 1 \quad \text{for } 0.5 < \sigma < 1$$
$$= 0.1 \quad \text{for } \sigma \leq 0.5$$

$$\text{Experiment 2} \quad : \quad H_T^{\sigma} = 1 \quad \text{for } \sigma \leq 0.5$$
$$= 0.1 \quad \text{for } 0.5 < \sigma < 1 \quad .$$

and ω_ψ is as before. Note that at $\sigma = 1$, for both experiments, ΔT and $\Delta \ln p_s$ are jointly constrained in the same manner as for the INMI case, i.e. (3.6) is applied, and we set $\omega_T = 1$ at $\sigma = 1$ for a maximum retention of the surface pressure information. This latter choice together with the formulation retained for VINMI as discussed in Section 3 (see (3.5)) forces the retention of the analyzed temperature field at $\sigma = 1$, which is not as accurate as the surface

pressure field. Results concerning this point are discussed at the end of this sub-section. The present choice of H^{σ}_{T} creates (as in the horizontal case) a clear distinction between the region of control for the VINMI-3D scheme as compared to the standard unconstrained INMI scheme.

Figures 14a-b show the temperature changes after initialization using the INMI and VINMI-3D schemes respectively for Experiment 1, for the level $\sigma = 0.786$ i.e. within the region of control of the temperature changes. It is clear that the analyzed temperature field over the continent has been maintained by the variational scheme. We also note that outside the horizontal region of control, the structure of temperature field changes is comparable for both schemes. Figure 15 gives the rms temperature changes over the North-American region of control as a function of the vertical coordinate σ, for both INMI and VINMI-3D schemes. Clearly the variational scheme has forced temperature changes in the desired way, i.e., retention of the analyzed temperature field in the lower half of the vertical domain and relaxation of the constraint in the upper domain (i.e. $\sigma < 0.5$). The transition zone near $\sigma = 0.5$ is also evident, indicating the robustness of the scheme. The same remark applies to the results of Experiment 2 shown in Fig.16. The structure of the temperature changes in the lower half of the domain ($\sigma > 0.5$) for the VINMI-3D scheme are now closer to the changes forced by the unconstrained scheme. However, in the upper domain, the variational scheme clearly forced the retention of the analyzed temperature field whereas the unconstrained scheme produced significant temperature changes. As in the case of the VINMI-2D scheme, this retention of the temperature field over the continents is done at the expense of the wind field. This is shown in Fig. 16b for Experiment 2 for the rms changes of the wind field over the horizontal region of control as a function of the vertical coordinate σ. As mentioned at the beginning of this sub-section, in order to limit the changes to the surface pressure field, ω_r at $\sigma = 1$ was set to unity for both experiments. This choice appears to be more consistent for Experiment 1 where the temperature changes are controlled in the lower half of the model

domain. The temperature changes at the top and bottom of the model are very small for both constrained and unconstrained experiments. The effect of constraining the initialization scheme to produce small changes in the temperature and surface pressure fields at $\sigma = 1$ (at the expense of the rotational part of the wind field) has forced changes to the wind field (Experiment 2) which produce an increased rms value near $\sigma = 1$ (see Fig. 16b). We note however that even for these extreme situations where strong retention of the temperature field (i.e. Experiment 1 and 2) is forced by the variational scheme, the associated changes to the wind (not shown) were still smaller than the expected analysis errors. Finally, Figures 17a-b show the surface pressure traces for Experiment 2 for point 1 (i.e. within the tropics) and point 4, and indicate the ability of the VINMI scheme to cope with such extreme weights specification (in the horizontal as well as in the vertical) and produce balanced initial conditions. The resulting surface pressure traces for Experiment 1 (not shown) are similar.

5. Summary and discussion

The technique of variational implicit nonlinear NMI (VINMI) is here extended to a multilevel operational model where (a) the weights vary horizontally and vertically and (b) the required computer time is maintained below 10 % of the CPU time needed for a 48 h forecast. By using VINMI, it is now possible to envisage controlling the initialization scheme to respect the degree of accuracy of the analyzed fields produced by operational data assimilation procedures. Results of VINMI in the context of horizontally varying weights (VINMI-2D) indicate that it is possible to maintain the same degree of balance in the initial fields as that produced by the INMI scheme and to take into account the very sharp variation in mass and wind field analysis accuracy between continents and oceans. Also, by allowing the weights to vary abruptly in the vertical, it was possible to demonstrate the ability of VINMI-3D to operate efficiently in cases of strong variations of the weights in the vertical.

Variational initialization, however, does not resolve the inherent deficiency of the analysis or the initialization schemes. Rather, it fills the need for abrupt variations in physical space in the changes caused by initialization to mass and wind fields, permitting horizontal and vertical variation that is consistent with expected analysis errors and is basically different than the one used by INMI in order to reach a balanced initial state. In this study, an efficient and robust scheme to force this constrained adjustment was developed. For practical purposes, the real test of any modifications to the initialization scheme is the improvement of the quality of the short-range forecast (typically 6 hours) which serves as a trial or first guess for the analysis. The next goal will be to examine whether VINMI can lead to such an improvement in the context of data assimilation. The optimal specification of the variational weights will have to be considered as a prerequisite to that study.

Finally, we note that a more coherent approach to the analysis-initialization problem for data assimilation may be to use either a Kalman filtering (Ghil et al., 1982) or optimal control (Talagrand 1988, Courtier and Talagrand 1988) technique. By combining the model

and approximate balancing constraints, these techniques can produce balanced initial conditions which make "optimal" use of the available data during the assimilation period. However, at present, both of these techniques require prohibitive computer resources which limits their use in an operational context. Variational initialization in the context of an operational data assimilation scheme should help (if properly used) to lessen some of the negative impact of unconstrained nonlinear NMI schemes on the analyzed fields. This should allow for a better evaluation of the sequential analysis-initialization techniques currently used for numerical weather prediction.

Acknowledgments. The authors wish to thank Clive Temperton and Andrew Staniforth for useful suggestions and Herschel Mitchell and Jean Côté for their useful reviews of the paper. We also thank Sylvie Gravel for help with some figures.

APPENDIX

Calculus of variations in the vertical

a. Boundary conditions for the vertical normal-modes

During the initialization process, it is required to specify the appropriate vertical boundary conditions on P in a manner which is consistent with the derivation of the vertical normal modes of the model. To avoid repetition, we refer the reader to Béland and Beaudoin (1985, Section 2), for the basic derivation. Note however the misprint in their equation (2.10). The correct form which applies at $\sigma = \sigma_1, 1$ is

$$\frac{\partial}{\partial t}\left[\left(\frac{\sigma}{R\gamma^*}\right)\frac{\partial P}{\partial \sigma} + \frac{\sigma}{RT_s^*}P_s\right]\Bigg|_{\sigma=\sigma_1,1} = \left[-\sigma\,\hat{\mathbf{V}}\cdot\nabla q - \frac{I}{\gamma^*}\right]\Bigg|_{\sigma=\sigma_1,1} \quad . \qquad \text{(A1)}$$

Setting the right-hand side of (A1) to zero (i.e. neglecting nonlinear terms), the appropriate vertical boundary conditions for the linearized system about a resting and isothermal basic state is

$$\frac{\partial P}{\partial \sigma} + \frac{\gamma^*(\sigma)}{T^*}P_s = 0 \quad \text{at} \quad \sigma = \sigma_1, 1 \quad , \qquad \text{(A2)}$$

where

$$\gamma^* = \frac{\kappa T^*}{\sigma} \quad ; \quad \kappa \equiv \frac{R}{c_p} \quad ,$$

and we take $T^* = 300$ K. The constants R and c_p are respectively the specific gas constant and heat capacity for dry air. Successive use of the continuous form of (2.2c), the hydrostatic equation, and the definition of P, leads to the following equation

$$\nu\frac{\partial P}{\partial t} - D = Q_P \quad , \qquad \text{(A3)}$$

where $D = \nabla\cdot\mathbf{V}$, Q_P represents a combination of nonlinear terms, and

$$\nu \equiv \frac{\partial}{\partial \sigma}\left(\frac{\sigma}{R\gamma^*(\sigma)}\frac{\partial}{\partial \sigma}\right) \qquad \text{(A4)}$$

is referred to as the *vertical structure operator*. We note that (2.5) is the analogue of (A3) obtained from the same set of equations but in a discretized form. From (2.5) and (A3) (nonlinear terms set to zero) it follows that the eigenvectors of the consistent discretization of $\mathcal{V}$ and the eigenvectors of matrix C are the same but the eigenvalues of $\mathcal{V}$ satisfy the relation

$$\mathcal{V}\,\xi_n(\sigma) = \frac{1}{\Phi_n}\,\xi_n(\sigma)$$

where Φ_n is an eigenvalue and ξ_n an eigenvector of matrix C.

b. Euler-Lagrange equations

We consider the minimization of the functional

$$J = \int_D \left[\frac{\omega_T}{m^2} \left\{ \frac{\sigma}{R\gamma^*} \left(\frac{\partial(\Delta P)}{\partial \sigma} \right)^2 \right\} + \omega_\psi \left(\nabla(\Delta\psi)\right)^2 + \frac{\omega_T}{RT^* m^2} \left(\Delta P\right)_s^2 \right] dx\, dy\, d\sigma \equiv$$

$$\equiv \int_D L_1 \; dx\, dy\, d\sigma \qquad \text{(A5)}$$

under the constraint M defined by (3.3a) and the boundary conditions (3.3b,c). The barotropic case has been discussed by FT89, therefore we need only consider the new form of potential energy. For this purpose, we set to zero the first variation of the Lagrangian $\mathcal{L}$ defined by

$$\mathcal{L} \equiv \int_D \left(L_1 + 2\lambda M\right) dx\, dy\, d\sigma \equiv \int_D L \; dx\, dy\, d\sigma$$

with respect to ΔP, (ref. Smirnov 1964) where λ is a variable Lagrange multiplier. It follows that

$$0 = \int_D \left[L_{(\Delta P)} - \left(\frac{\partial}{\partial \sigma} L_{(\Delta P)_\sigma} \right) + \frac{\partial^2}{\partial x^2} L_{(\Delta P)_{xx}} + \frac{\partial^2}{\partial y^2} L_{(\Delta P)_{yy}} \right] \delta(\Delta P)\ dx\ dy\ d\sigma\ +$$

$$\int_{\sigma_1}^{1} \left\{ \oint_\Gamma \left[\delta(\Delta P)_x L_{(\Delta P)_{xx}} - \delta(\Delta P) \frac{\partial}{\partial x} L_{(\Delta P)_{xx}} \right] dy - \right.$$

$$\left. \left[\delta(\Delta P)_y L_{(\Delta P)_{yy}} - \delta(\Delta P) \frac{\partial}{\partial y} L_{(\Delta P)_{yy}} \right] dx \right\} d\sigma$$

$$+ \int_{D_h} \left[F_{(\Delta P)_\sigma} \delta(\Delta P) \right] \Big|_{\sigma=\sigma_1}^{\sigma=1} dx\ dy + \int_{D_h} \left[G_{(\Delta P)} \delta(\Delta P) \right]_{\sigma=1} dx\ dy$$

$$+ \int_{D_h} \left[H_{(\Delta P)} \delta(\Delta P) \right] \Big|_{\sigma=\sigma_1} dx\ dy \quad .$$

where

$$F(x, y, \sigma) \equiv \frac{\omega_\Gamma}{m^2} \frac{\sigma}{R\gamma^*} \left(\frac{\partial(\Delta P)}{\partial \sigma} \right)^2 \quad ; \quad G(x, y, \sigma) \equiv \frac{\omega_\Gamma(x,y,1)}{m^2 R T^*} (\Delta P)^2 \quad ;$$

and

$$H(x, y, \sigma_1) \equiv - \frac{\sigma_1\, \omega_\Gamma(x,y,\sigma_1)}{m^2 R T^*} (\Delta P)^2 \quad .$$

We introduced the notation

$$Z_{(\zeta)} \equiv \frac{\partial Z}{\partial \zeta} \quad .$$

Since the normal derivative of (ΔP) on Γ is fixed when $(\Delta \psi)$ is fixed (see (3.3b-c)), the extremum condition is satisfied provided that on the boundary of Γ we have the same type of *natural boundary conditions* as specified in the barotropic version, i.e. (3.7b). The Euler-Lagrange equation for (ΔP) must be satisfied for an extremum, i.e.

$$L_{(\Delta P)} - \left(\frac{\partial}{\partial \sigma} L_{(\Delta P)_\sigma} \right) + \frac{\partial^2}{\partial x^2} L_{(\Delta P)_{xx}} + \frac{\partial^2}{\partial y^2} L_{(\Delta P)_{yy}} = 0 \quad .$$

Therefore it follows that

$$\nabla^2\lambda = \frac{-1}{m^2}\frac{\partial}{\partial\sigma}\left[\frac{\omega_T\,\sigma}{R\,\gamma^*}\frac{\partial(\Delta P)}{\partial\sigma}\right] \qquad \text{(A6a)}$$

$$\left(\frac{\partial\lambda}{\partial n}\right)_\Gamma = 0 \quad . \qquad \text{(A6b)}$$

We have not yet introduced the vertical boundary conditions that must be imposed on ΔP. The natural vertical boundary conditions are

$$F_{(\Delta P)_\sigma} + G_{(\Delta P)} = 0 \qquad \text{at} \quad \sigma = 1 \qquad \text{(A7)}$$

$$F_{(\Delta P)_\sigma} - H_{(\Delta P)} = 0 \qquad \text{at} \quad \sigma = \sigma_1 \quad . \qquad \text{(A8)}$$

Assuming $\omega_T > 0$, (A7) becomes

$$\frac{\partial}{\partial\sigma}(\Delta P) + \frac{\gamma^*}{T^*}(\Delta P) = 0 \qquad \text{at} \quad \sigma = 1 \quad ,$$

which has the same form as the lower boundary condition of (A2). Also, condition (A8) gives the same form as the upper boundary condition of (A2), i.e.

$$\frac{\partial}{\partial\sigma}(\Delta P) + \frac{\gamma^*(\sigma_1)}{T^*}(\Delta P)_s = 0 \qquad \text{at} \quad \sigma = \sigma_1 \quad .$$

Thus, the introduction of the term involving H in the functional J generates the same vertical boundary conditions as for the unconstrained INMI scheme. We show in subsection 3d that the present constrained minimization problem reduces to the INMI scheme when the weights ω_T and ω_v are constant and equal. Finally, we note that for models with the top at $\sigma = 0$, the usual boundary condition used to define the vertical normal modes, i.e

$$\frac{\partial}{\partial\sigma}(\Delta P) = 0 \qquad \text{at} \quad \sigma = 0$$

is the natural boundary condition that appears in order to minimize the energy functional, and for that case, the term involving H vanishes as $\sigma_1 \to 0$.

References

Béland, M., and C. Beaudoin, 1985: A global spectral model with a finite-element formulation for the vertical discretization: Adiabatic formulation. *Mon. Wea. Rev.*, **113**, 1910-1919.

Benoit, R., J. Côté and J. Mailhot, 1989: Inclusion of a TKE boundary layer parameterization in the Canadian regional finite-element model. *Mon. Wea. Rev.*, **117**, 1726-1750.

Bourke, W., and J.L. McGregor, 1983: A nonlinear vertical mode initialization scheme for a limited area prediction model. *Mon. Wea. Rev.*, **111**, 2285-2297.

Concus, P., G.H. Golub, and D.P. O'leary, 1976: A generalized conjugate gradient method for the numerical solution of elliptic partial differential equations, *Sparse Matrix Computations*, J. R. Bunch and D.J. Rose, Eds. Academic Press, 309-332.

Courtier, P., and O. Talagrand, 1988: Variational assimilation of meteorological observations with the direct and adjoint shallow-water equations. *Data assimilation and the use of satellite data. ECMWF Seminar 1988*, 31-69. [Available from the European Centre for Medium Range Weather Forecasts, Shinfield Park, Reading RG2 9AX, England].

Daley, R., 1978: Variational nonlinear normal mode initialization. *Tellus*, **30**, 201-218.

_____, 1979: The application of nonlinear normal mode initialization to an operational forecast model. *Atmos. Ocean*, **17**, 97-124.

Errico, R.M., and D.L.Williamson, 1988: The behaviour of gravitational modes in numerical forecasts. *Mon. Wea. Rev.*, **116**, 1737-1756.

Errico, R.M., 1989: The degree of Machenhauer balance in a climate model. *Mon. Wea. Rev.*, **117**, 2723-2733.

Fillion, L., and C. Temperton, 1989: Variational implicit normal mode initialization. *Mon. Wea. Rev.*, **117**, 2219-2229.

Fillion, L., 1991: Variational implicit normal mode initialization on the sphere. *Mon. Wea. Rev.*, **119**, 631-652.

Ghil, M., S. Cohn, J. Tavantzis, K. Bube, and E. Isaacson, 1982: Applications of estimation theory to numerical weather prediction. In *Dynamic Meteorology, data assimilation methods* (L. Bengtsson, M. Ghil and E. Källén, editors), 139-224. Springer-Verlag, New York.

Juvanon du Vachat, R., 1988: Non-normal mode initialization: formulation and application to the inclusion of the beta terms in the linearization. *Mon. Wea. Rev.*, **116**, 2013-2024.

Ko, S.D., J. Tribbia and J. Boyd, 1989: Energetics analysis of a multilevel global spectral model. Part I: Balanced energy and transient energy. *Mon. Wea. Rev.*, **117**, 1941-1954.

Lorenc, A., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **99**, 303-319.

Machenhauer, B., 1977: On the dynamics of gravity oscillations in a shallow-water model, with application to normal mode initialization. *Contrib. Atmos. Phys.*, **50**, 253-271.

Mitchell, H., C. Charette, C. Chouinard, B. Brasnett, 1990: Revised interpolation statistics for the Canadian data assimilation procedure: Their derivation and application. *Mon. Wea. Rev.*, **118**, 1591-1614.

Puri, K., 1983: Some experiments in variational normal mode initialization in data assimilation. *Mon. Wea. Rev.*, **111**, 1208-1218.

Smirnov, V. I., 1964: *A course of Higher Mathematics*, Vol. IV, 4th ed. Pergamon Press, 811 pp.

Staniforth, A., and R. Daley, 1979: A baroclinic finite-element model for regional forecasting with the primitive equations. *Mon. Wea. Rev.*, **107**, 107-121.

_____, M. Béland, and J. Côté, 1985: An analysis of the vertical structure equation in sigma coordinates. *Atmos. Ocean*, **23**, 323-358.

Talagrand, O., 1988: Four-Dimensional Variational Assimilation. *Data assimilation and the use of satellite data. ECMWF Seminar 1988*, 1-30. [Available from the European Centre for Medium Range Weather Forecasts, Shinfield Park, Reading RG2 9AX, England].

Tanguay, M., A. Simard and A. Staniforth, 1989 : A three dimensional semi-Lagrangian scheme for the Canadian regional finite-element forecast model. *Mon. wea. Rev.*, **117**, 1861-1871.

Temperton, C., 1984: Variational normal mode initialization for a multilevel model. *Mon. Wea. Rev.*, **112**, 2303-2316.

_____, and D.L. Williamson, 1979: Normal mode initialization for a multilevel grid-point model. ECMWF Tech. Rep. No. 11, Shinfield Park, Reading, Berkshire RG2 9AX, U.K., 91 pp.

_____, 1988: Implicit normal mode initialization. *Mon. Wea. Rev.*, **116**, 1013-1031.

_____, and M. Roch, 1991: Implicit normal mode initialization for an operational regional model. *Mon. Wea. Rev.*, **119**, in press.

Tribbia, J. J., 1982: On variational normal mode initialization. *Mon. Wea. Rev.*, **110**, 455-470.

Williamson, D. L., and R. Daley, 1983: An iterative analysis-initialization technique. *Mon. Wea. Rev.*, **111**, 1517-1536.

Figure captions

Fig. 1. The horizontal domain and non-uniform grid. Points 1-4 are locations where surface pressure time-traces are stored during a model integration.

Fig. 2. The vertical distribution of sigma levels of the RFE model.

Fig. 3. Vertical modes 7-10 of the model. (a) solid line: mode number 7, dashed line: mode number 8; (b) solid line: mode number 9, dashed line: mode number 10.

Fig. 4. Mean sea-level pressure at 00 UTC, 22 January 1986, before initialization. Contour interval 8 mb.

Fig. 5. Root-mean-square expected analysis error. (a) Height-field (m) at 850 mb, contour interval 3 m; (b) Height-field (m) at 250 mb, contour interval 5 m; (c-d) Zonal component of the wind (ms^{-1}) at 850 and 250 mb (resp), contour interval 1 ms^{-1}; (e-f) Temperature (Celsius) at 850 and 250 mb (resp), contour interval 0.4° C.

Fig.6. The value of BAL as a function of iteration number for the first three vertical modes of the model for (a) unconstrained (INMI); (b) variational (VINMI-2D) initialization.

Fig. 7. Root-mean-square increments of the generalized-geopotential field ($m^2 s^{-2}$) as a function of the vertical coordinate σ for the first three iteration of the (a) unconstrained (INMI) scheme ; (b) variational (VINMI-2D) scheme.

Fig. 8. Root-mean-square increments of the modulus of the wind field (ms^{-1}) as a function of the vertical coordinate σ for the first three iterations of the (a) unconstrained (INMI) scheme; (b) variational (VINMI-2D) scheme.

Fig.9. Height-field differences (m) at $\sigma = 0.250$ between initialized and original (analyzed) fields. (a) INMI; (b) VINMI-2D. Contour interval 5m.

Fig. 10. Height (m) and wind field differences (ms^{-1}) at $\sigma = 0.250$ between variationally initialized (VINMI-2D) and unconstrained initialized (INMI) fields. Maximum wind modulus 2 ms^{-1}. Contour interval 10 m.

Fig. 11. Surface pressure changes (mb) of the original analyzed field done by the initialization schemes. (a) unconstrained INMI; (b) variational VINMI-2D Contour interval 0.5 mb.

Fig. 12. Vertical velocity ω (μbs^{-1}) at 700 mb. (a) before initialization; (b) after VINMI-2D. Contour interval 2 μbs^{-1} .

Fig. 13. Time trace of surface pressure at selected grid points (see Fig. 1). (a) point 1; (b) point 2; (c) point 3. Solid line: no initialization. Dashed line: after three iterations of INMI. Dotted line: after three iterations of VINMI-2D.

Fig. 14. Temperature changes at $\sigma = 0.786$ due to (a) unconstrained initialization (INMI); (b) variational initialization (VINMI-3D). Contour interval 0.05 deg.

Fig. 15. Root-mean-square changes of the original (analyzed) temperature field over the North-American region of control as a function of levels σ. Solid line: unconstrained initialization (INMI). Dashed line: variational initialization (VINMI-3D, Experiment 1).

Fig. 16. Root-mean-square changes of the original (analyzed) (a) temperature (deg); (b) modulus of the wind (ms^{-1}), over the North-American region of control as a function of levels σ. Solid line: unconstrained initialization (INMI). Dashed line: variational initialization (VINMI-3D, Experiment 2).

Fig. 17. Time-trace of surface pressure at selected grid points. (a) point 1; (b) point 4. Solid line: no initialization. Dashed line: unconstrained initialization (INMI, 3 iterations). Dotted line: variational initialization (VINMI-3D, 3 iterations).

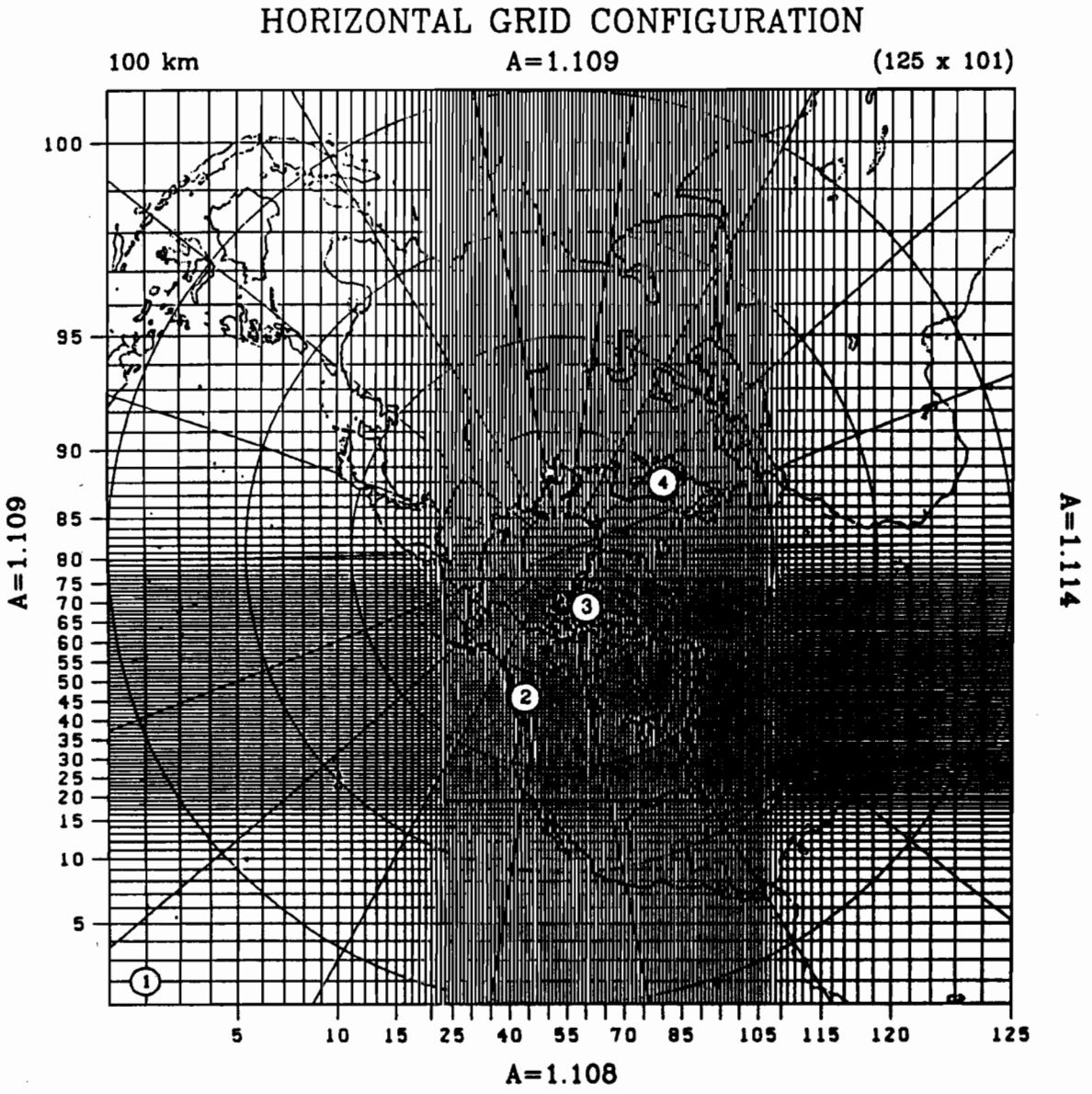


Fig. 1. The horizontal domain and non-uniform grid. Points 1-4 are locations where surface pressure time-traces are stored during a model integration.

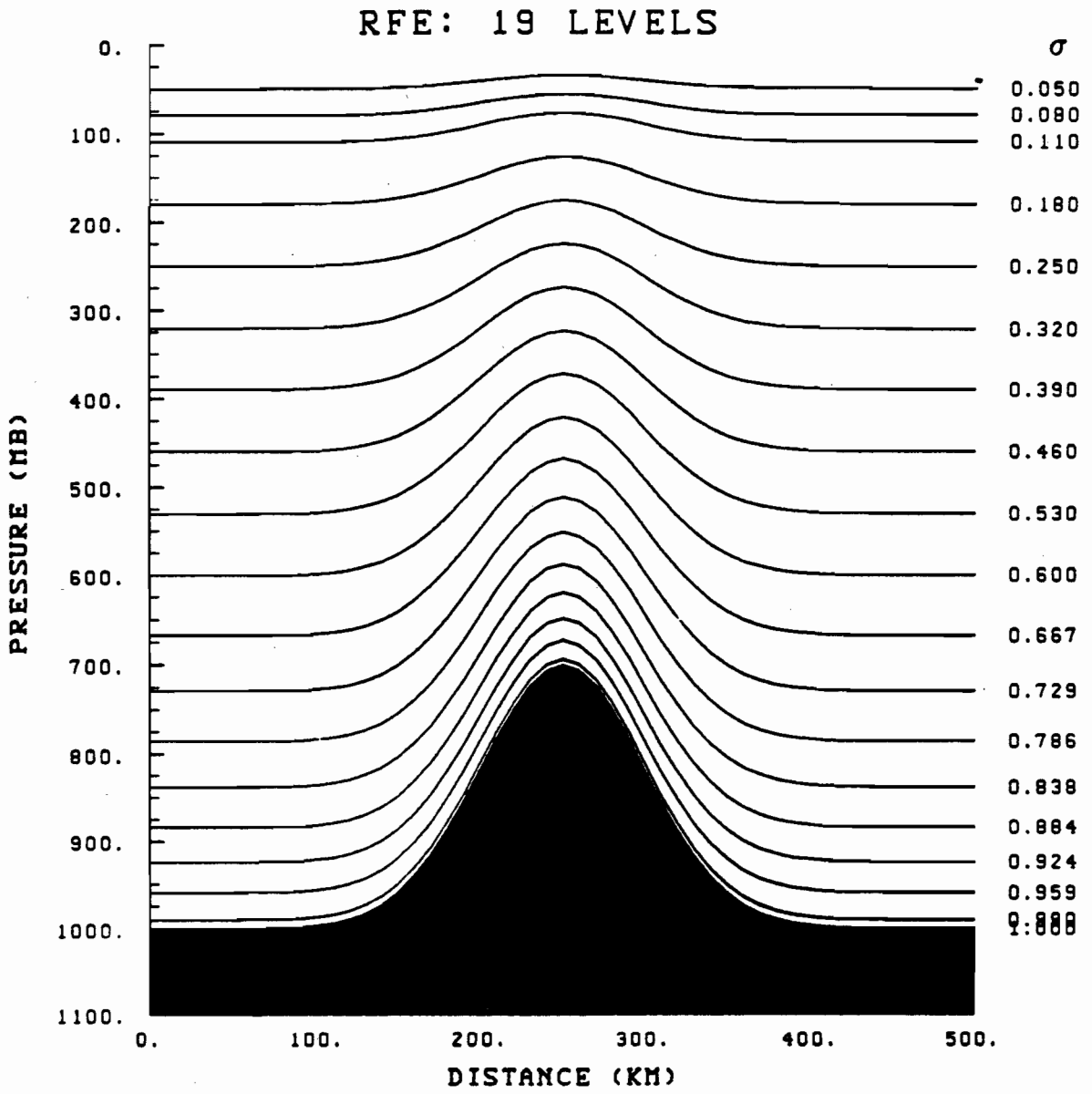


Fig. 2. The vertical distribution of sigma levels of the RFE model.

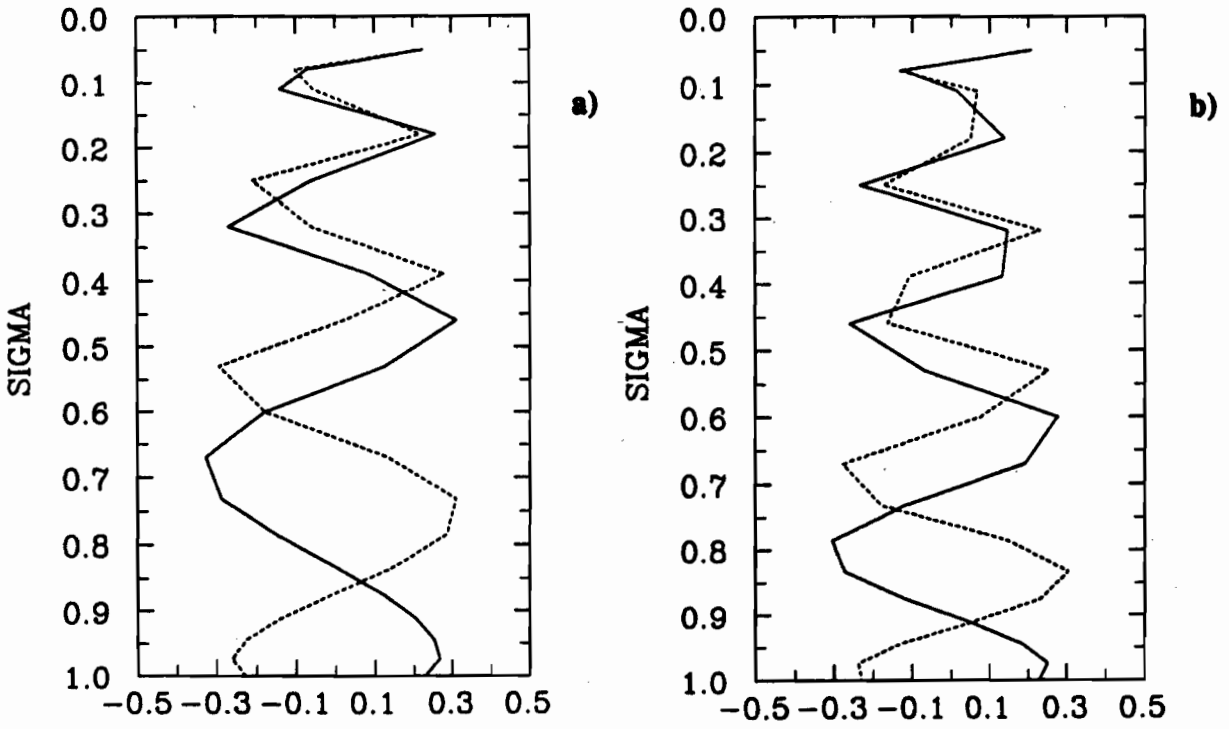


Fig. 3. Vertical modes 7-10 of the model. (a) solid line: mode number 7, dashed line: mode number 8; (b) solid line: mode number 9, dashed line: mode number 10.

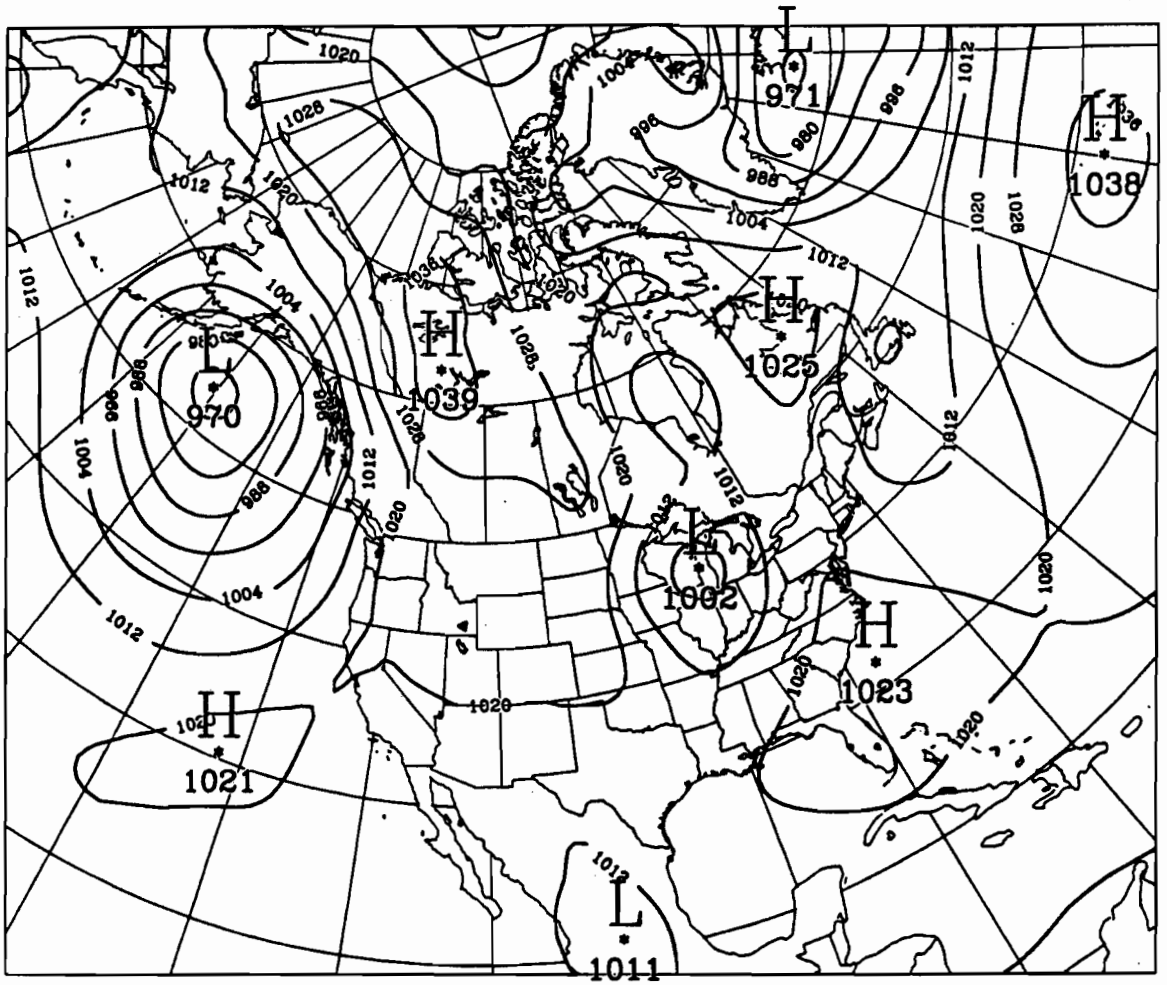


Fig. 4. Mean sea-level pressure at 00 UTC, 22 January 1986, before initialization. Contour interval 8 mb.

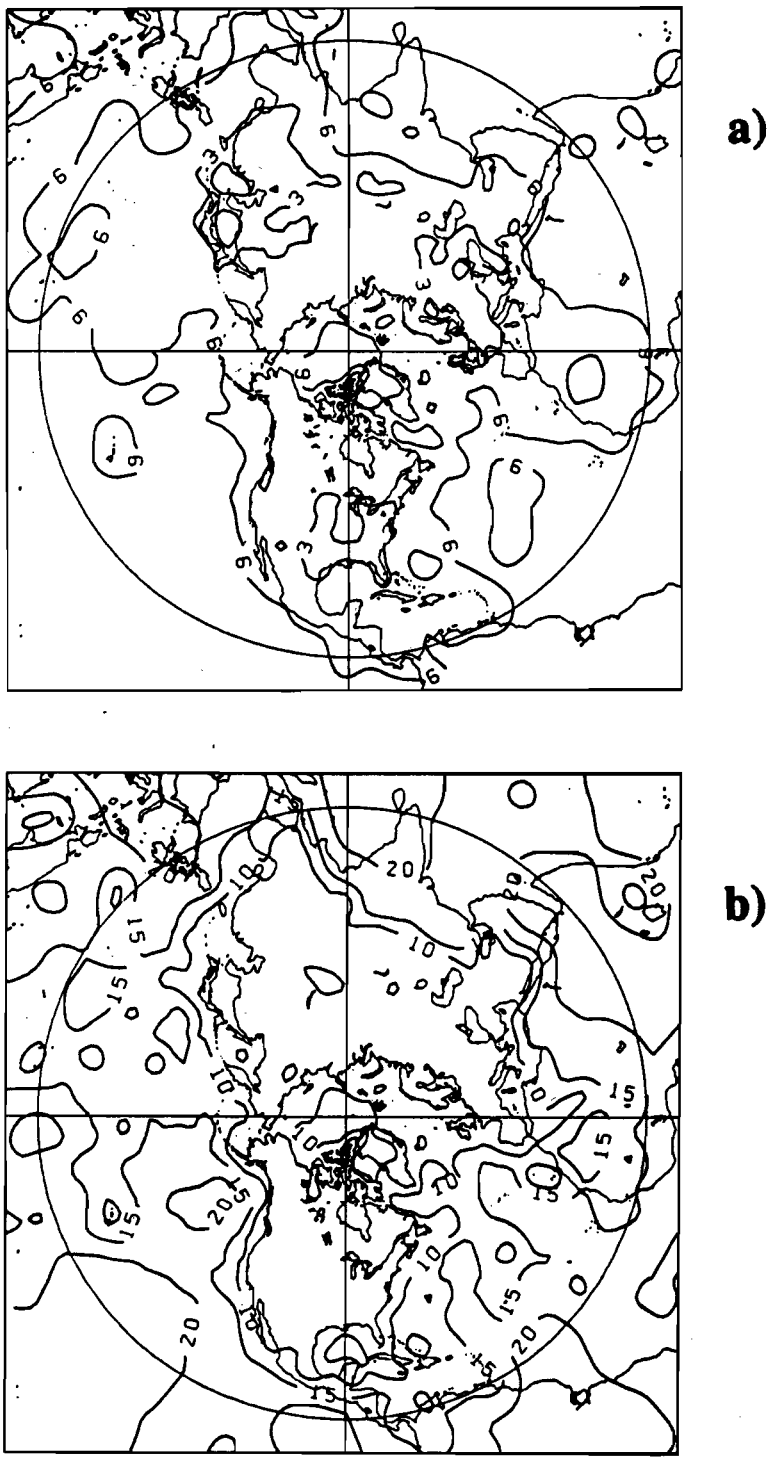
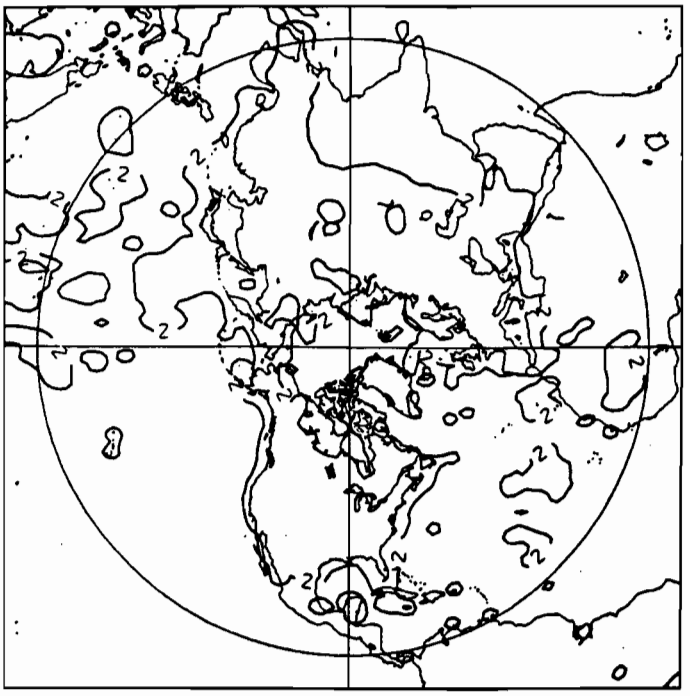
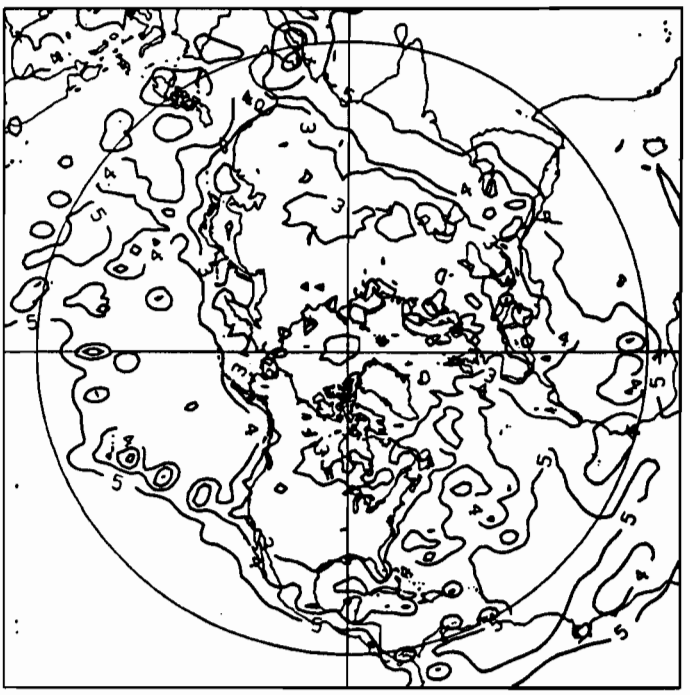


Fig. 5. Root-mean-square expected analysis error. (a) Height-field (m) at 850 mb, contour interval 3 m; (b) Height-field (m) at 250 mb, contour interval 5 m; (c-d) Zonal component of the wind (ms^{-1}) at 850 and 250 mb (resp), contour interval 1 ms^{-1}; (e-f) Temperature (Celsius) at 850 and 250 mb (resp), contour interval 0.4° C.

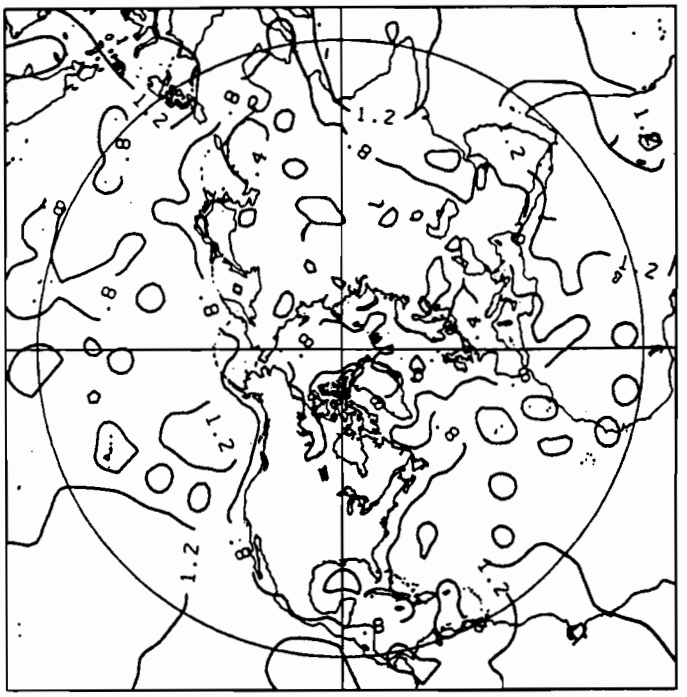


c)



d)

Fig. 5. Root-mean-square expected analysis error. (a) Height-field (m) at 850 mb, contour interval 3 m; (b) Height-field (m) at 250 mb, contour interval 5 m; (c-d) Zonal component of the wind (ms^{-1}) at 850 and 250 mb (resp), contour interval 1 ms^{-1}; (e-f) Temperature (Celsius) at 850 and 250 mb (resp), contour interval 0.4° C.



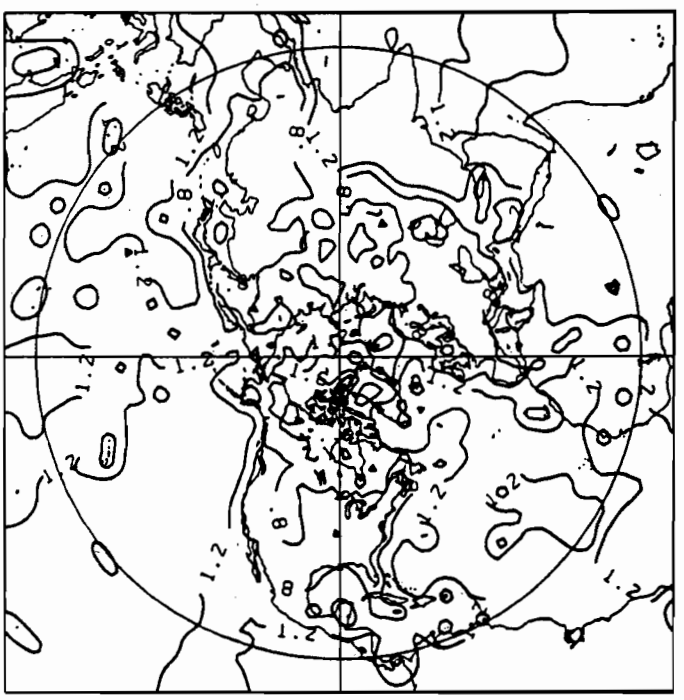


Fig. 5. Root-mean-square expected analysis error. (a) Height-field (m) at 850 mb, contour interval 3 m; (b) Height-field (m) at 250 mb, contour interval 5 m; (c-d) Zonal component of the wind (ms^{-1}) at 850 and 250 mb (resp), contour interval 1 ms^{-1}; (e-f) Temperature (Celsius) at 850 and 250 mb (resp), contour interval 0.4° C.

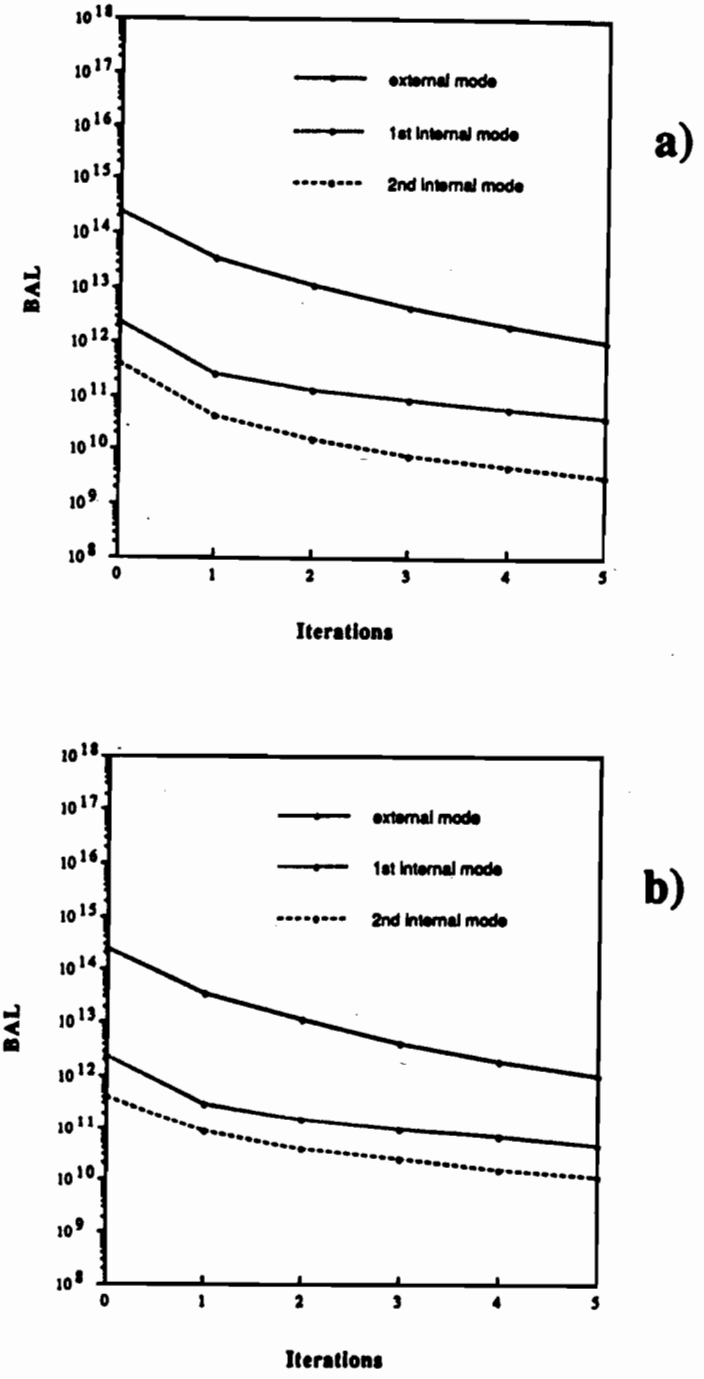


Fig.6. The value of BAL as a function of iteration number for the first three vertical modes of the model for (a) unconstrained (INMI); (b) variational (VINMI-2D) initialization.

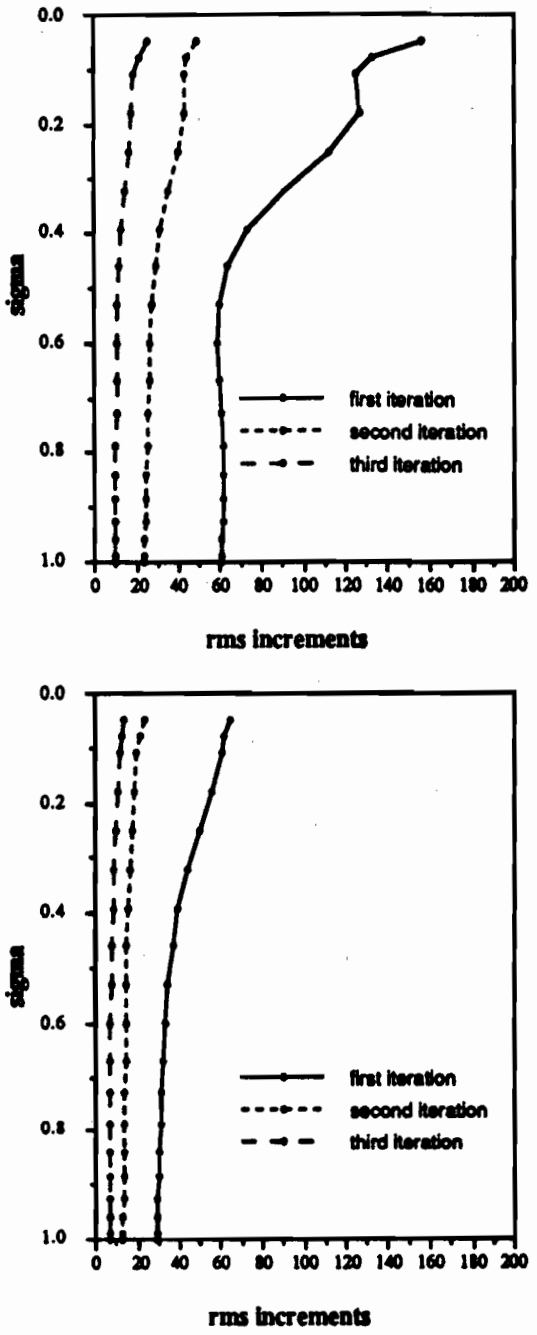


Fig. 7. Root-mean-square increments of the generalized-geopotential field ($m^2 s^{-2}$) as a function of the vertical coordinate σ for the first three iteration of the (a) unconstrained (INMI) scheme ; (b) variational (VINMI-2D) scheme.

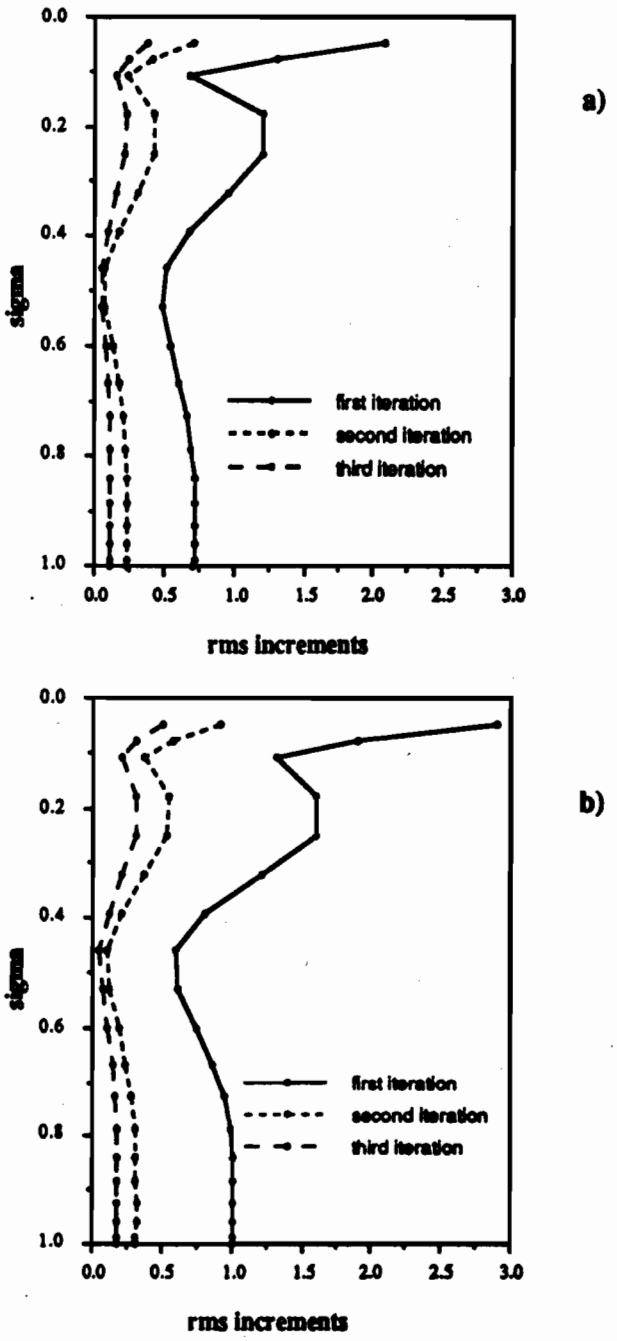


Fig. 8. Root-mean-square increments of the modulus of the wind field (ms^{-1}) as a function of the vertical coordinate σ for the first three iterations of the (a) unconstrained (INMI) scheme; (b) variational (VINMI-2D) scheme.

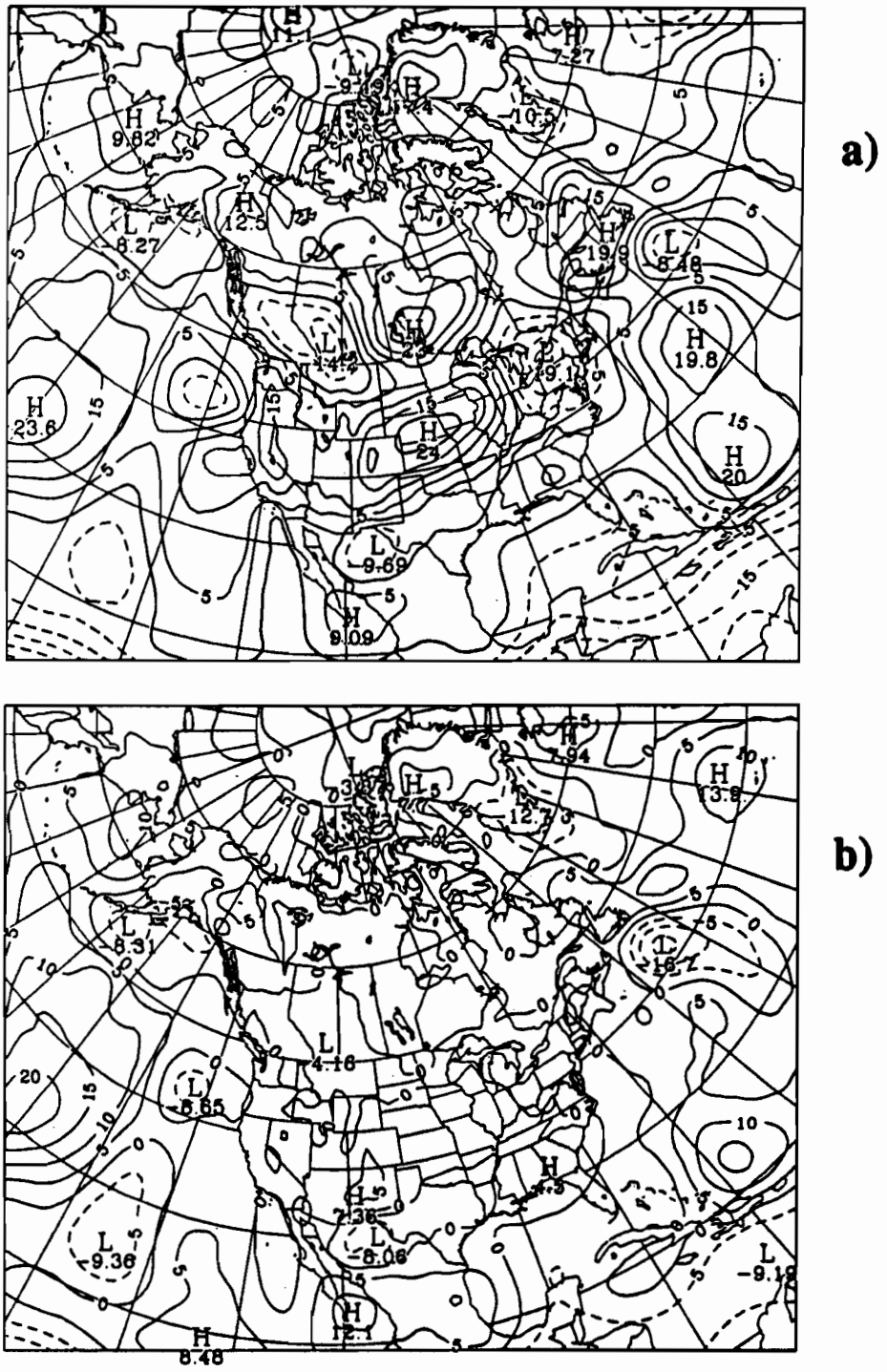


Fig.9. Height-field differences (m) at σ = 0.250 between initialized and original (analyzed) fields. (a) INMI; (b) VINMI-2D. Contour interval 5m.

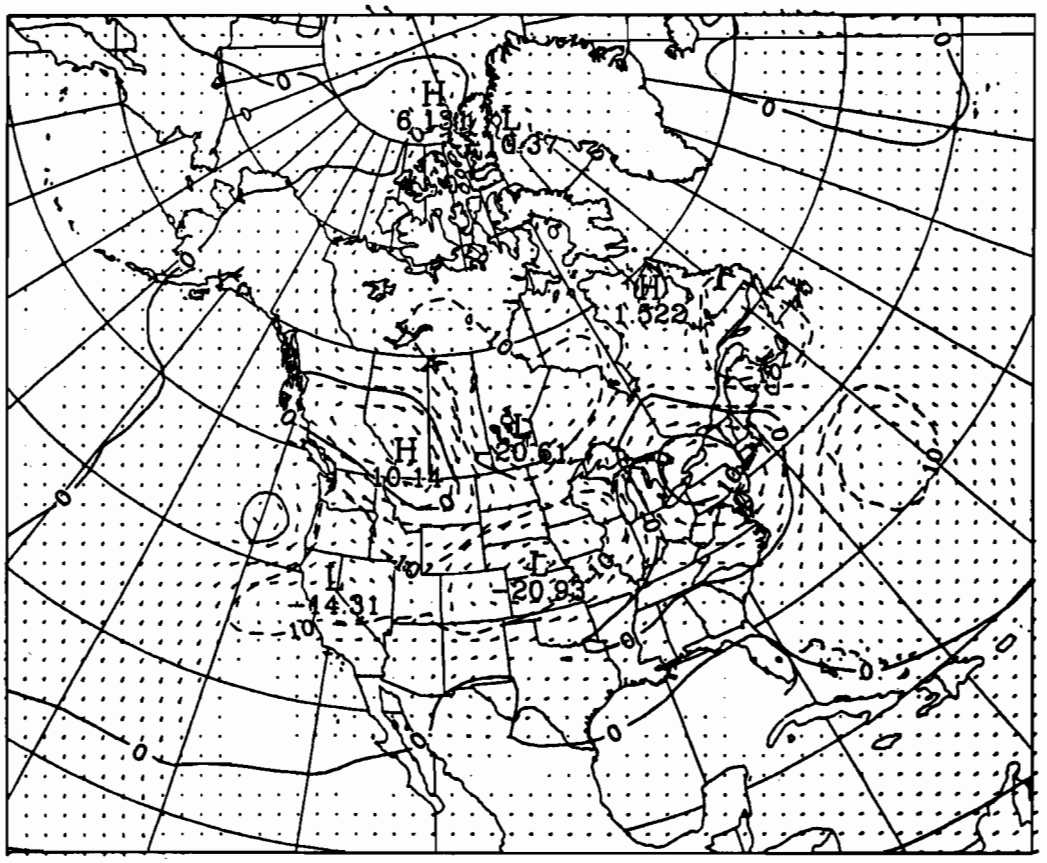


Fig. 10. Height (m) and wind field differences (ms^{-1}) at $\sigma = 0.250$ between variationally initialized (VINMI-2D) and unconstrained initialized (INMI) fields. Maximum wind modulus 2 ms^{-1}. Contour interval 10 m.

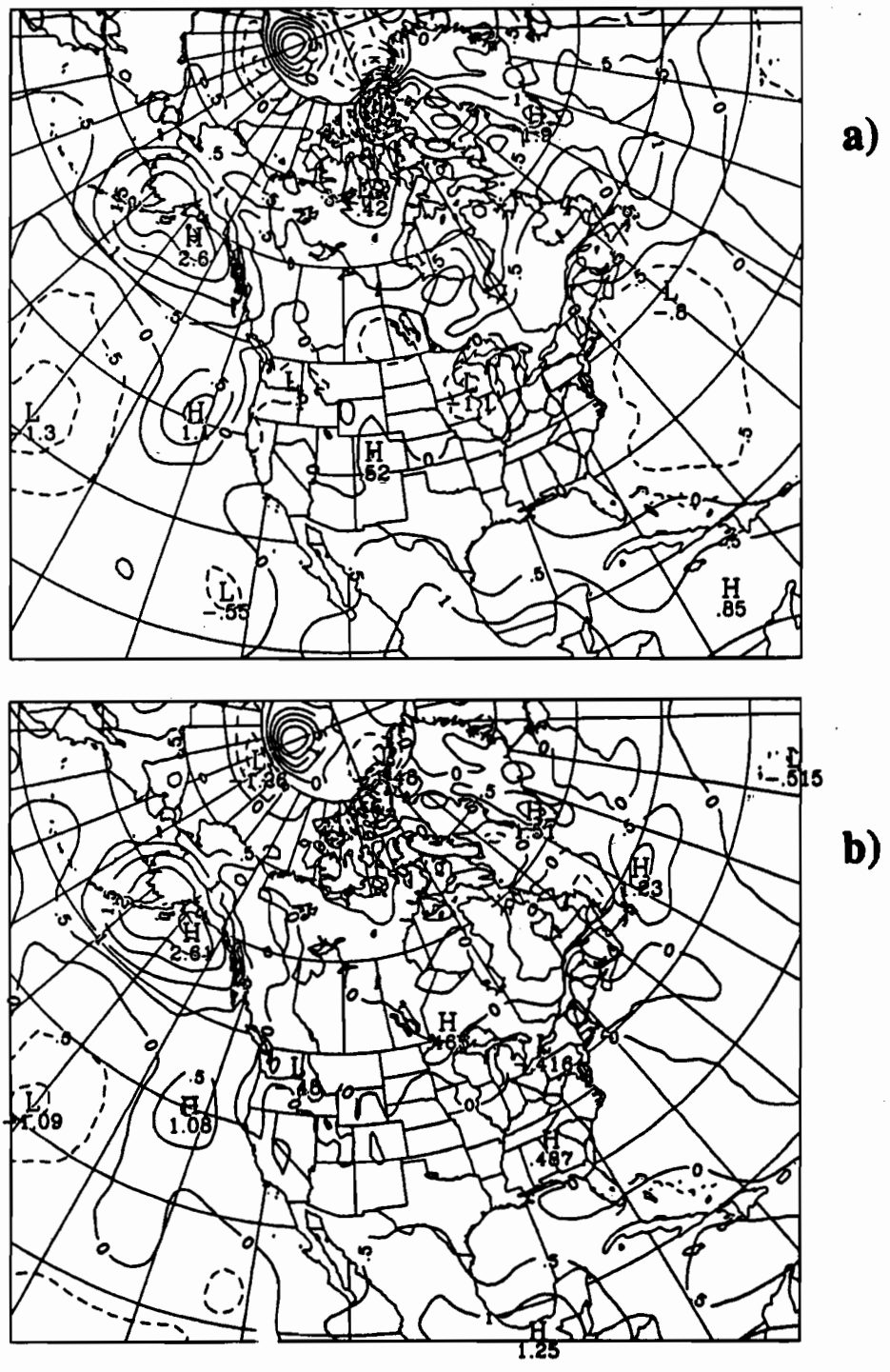


Fig. 11. Surface pressure changes (mb) of the original analyzed field done by the initialization schemes. (a) unconstrained INMI; (b) variational VINMI-2D Contour interval 0.5 mb.

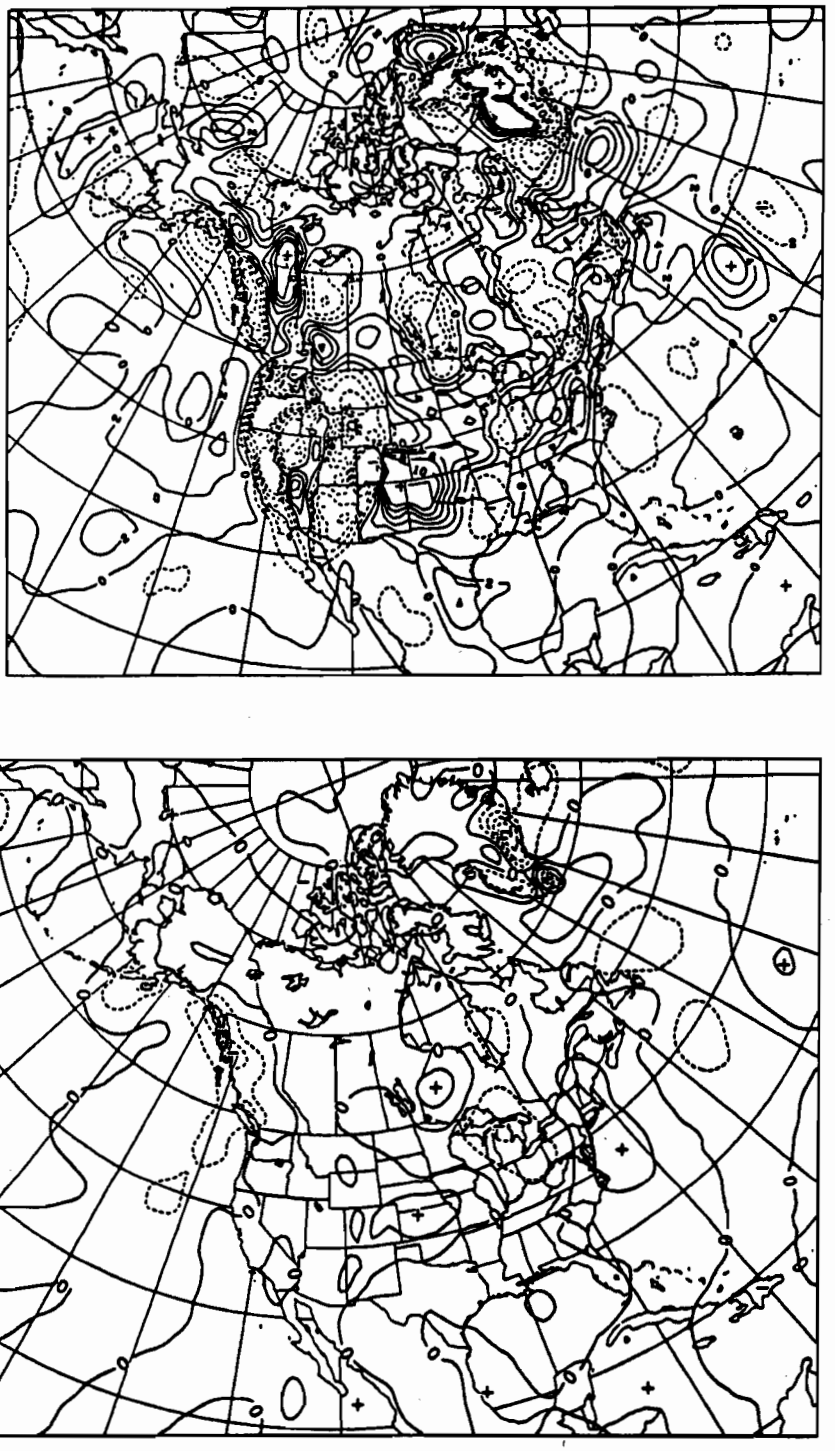


Fig. 12. Vertical velocity ω (μbs^{-1}) at 700 mb. (a) before initialization; (b) after VINMI-2D. Contour interval 2 μbs^{-1}.

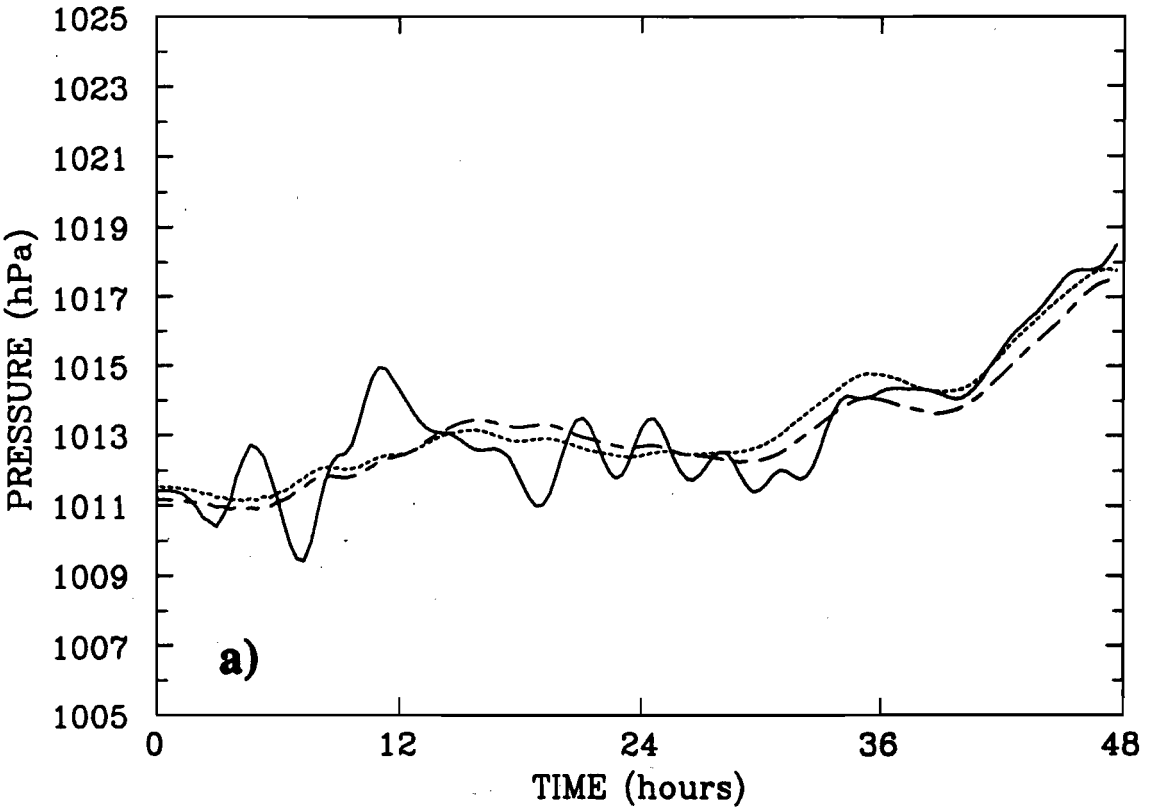


Fig. 13. Time trace of surface pressure at selected grid points (see Fig. 1). (a) point 1; (b) point 2; (c) point 3. Solid line: no initialization. Dashed line: after three iterations of INMI. Dotted line: after three iterations of VINMI-2D.

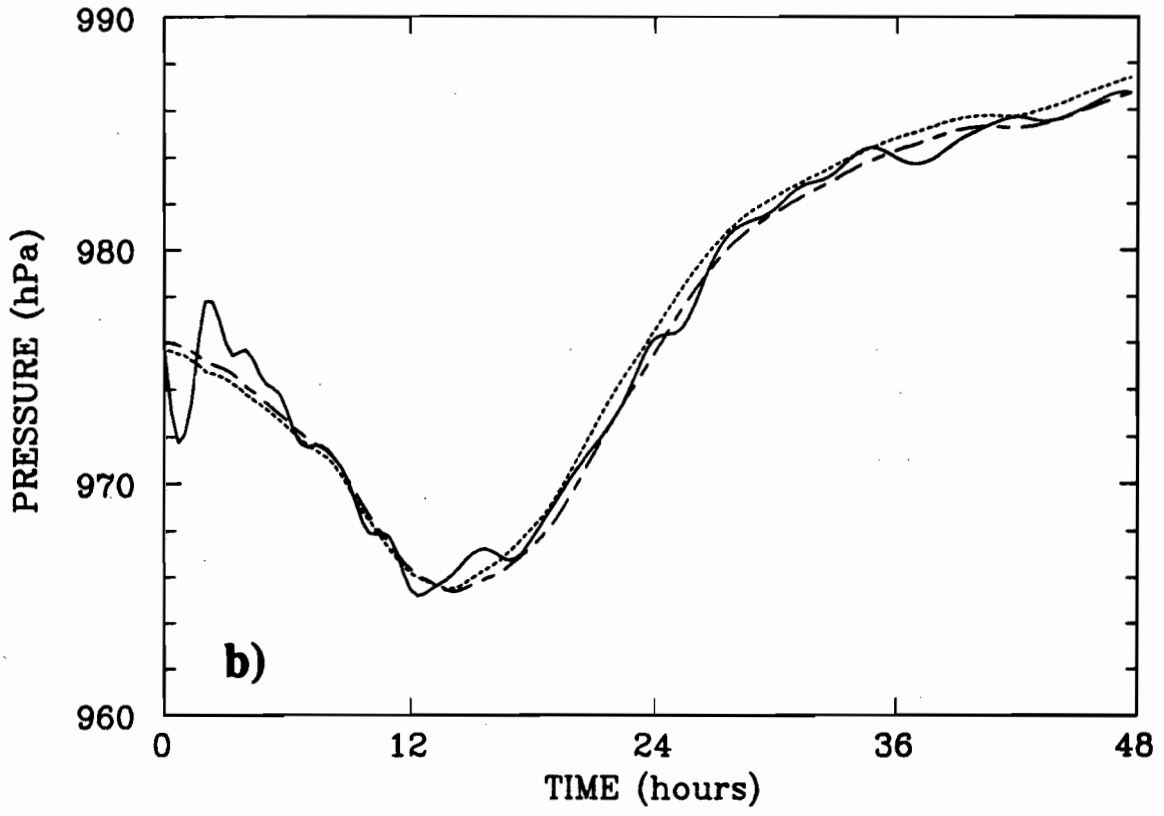


Fig. 13. Time trace of surface pressure at selected grid points (see Fig. 1). (a) point 1; (b) point 2; (c) point 3. Solid line: no initialization. Dashed line: after three iterations of INMI. Dotted line: after three iterations of VINMI-2D.

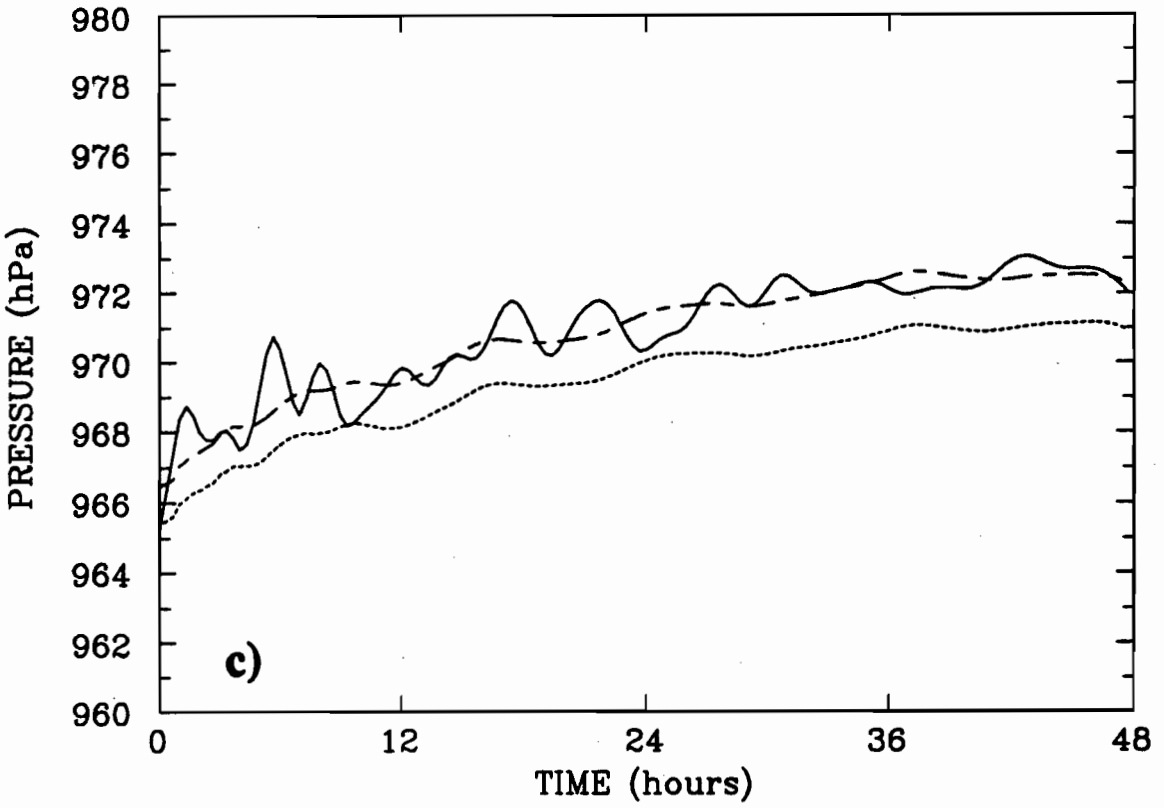


Fig. 13. Time trace of surface pressure at selected grid points (see Fig. 1). (a) point 1; (b) point 2; (c) point 3. Solid line: no initialization. Dashed line: after three iterations of INMI. Dotted line: after three iterations of VINMI-2D.

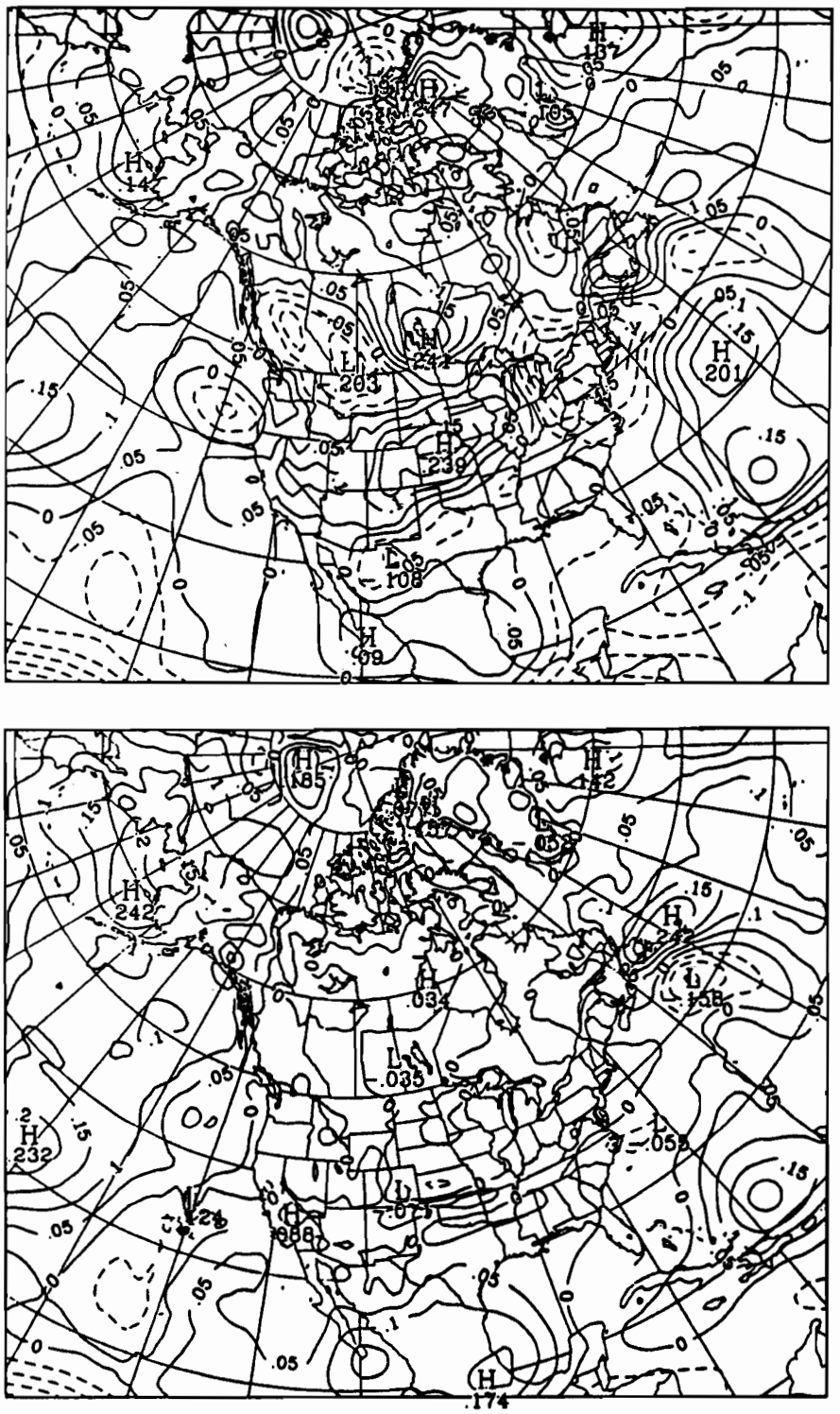


Fig. 14. Temperature changes at σ = 0.786 due to (a) unconstrained initialization (INMI); (b) variational initialization (VINMI-3D). Contour interval 0.05 deg.

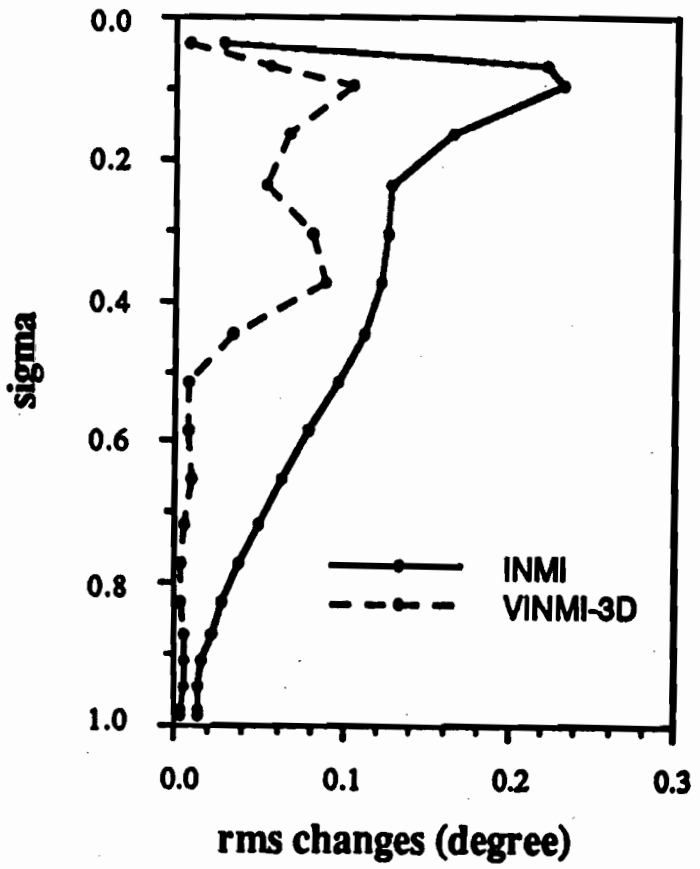


Fig. 15. Root-mean-square changes of the original (analyzed) temperature field over the North-American region of control as a function of levels σ. Solid line: unconstrained initialization (INMI). Dashed line: variational initialization (VINMI-3D, Experiment 1).

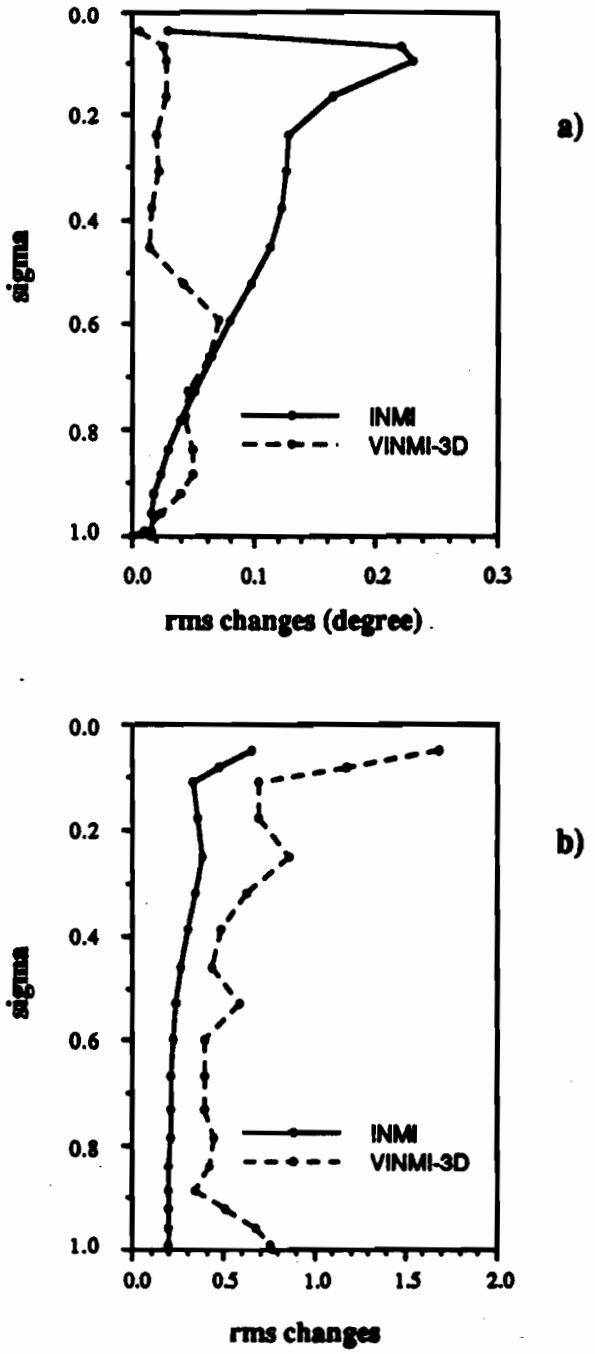


Fig. 16. Root-mean-square changes of the original (analyzed) (a) temperature (deg); (b) modulus of the wind (ms^{-1}), over the North-American region of control as a function of levels σ. Solid line: unconstrained initialization (INMI). Dashed line: variational initialization (VINMI-3D, Experiment 2).

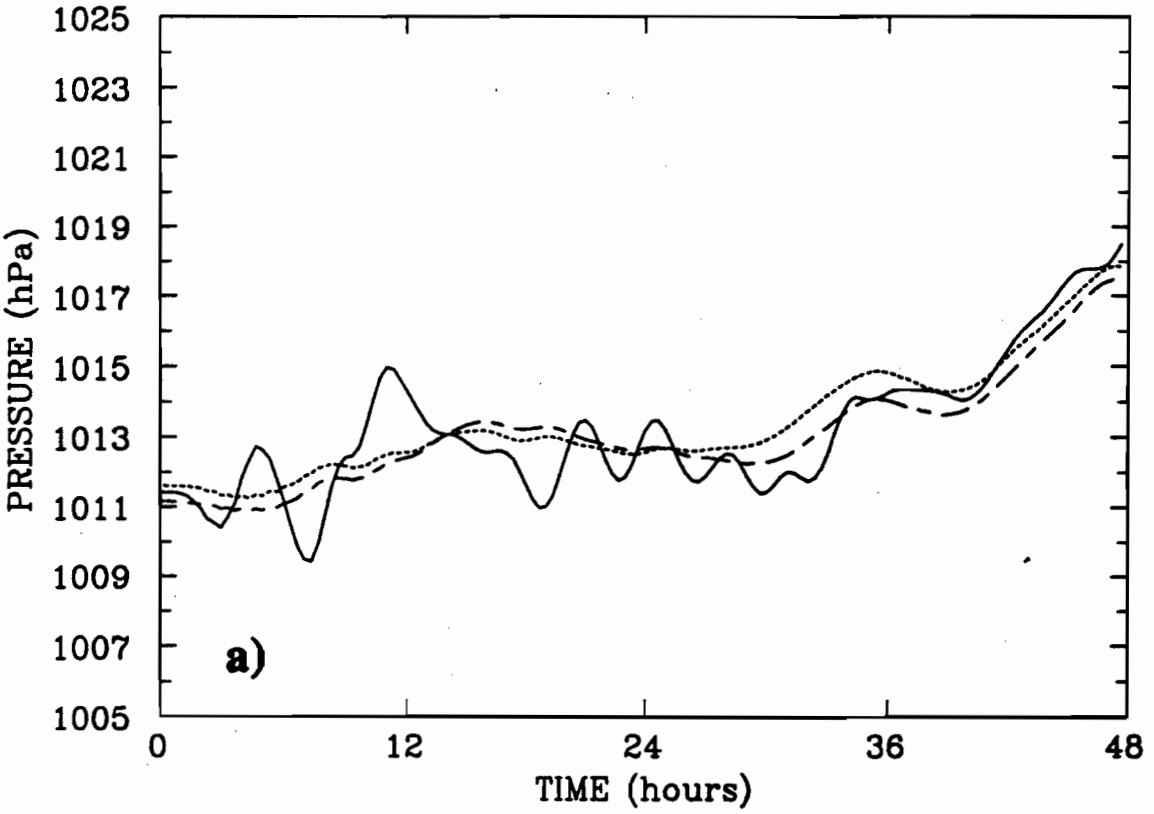


Fig. 17. Time-trace of surface pressure at selected grid points. (a) point 1; (b) point 4. Solid line: no initialization. Dashed line: unconstrained initialization (INMI, 3 iterations). Dotted line: variational initialization (VINMI-3D, 3 iterations).

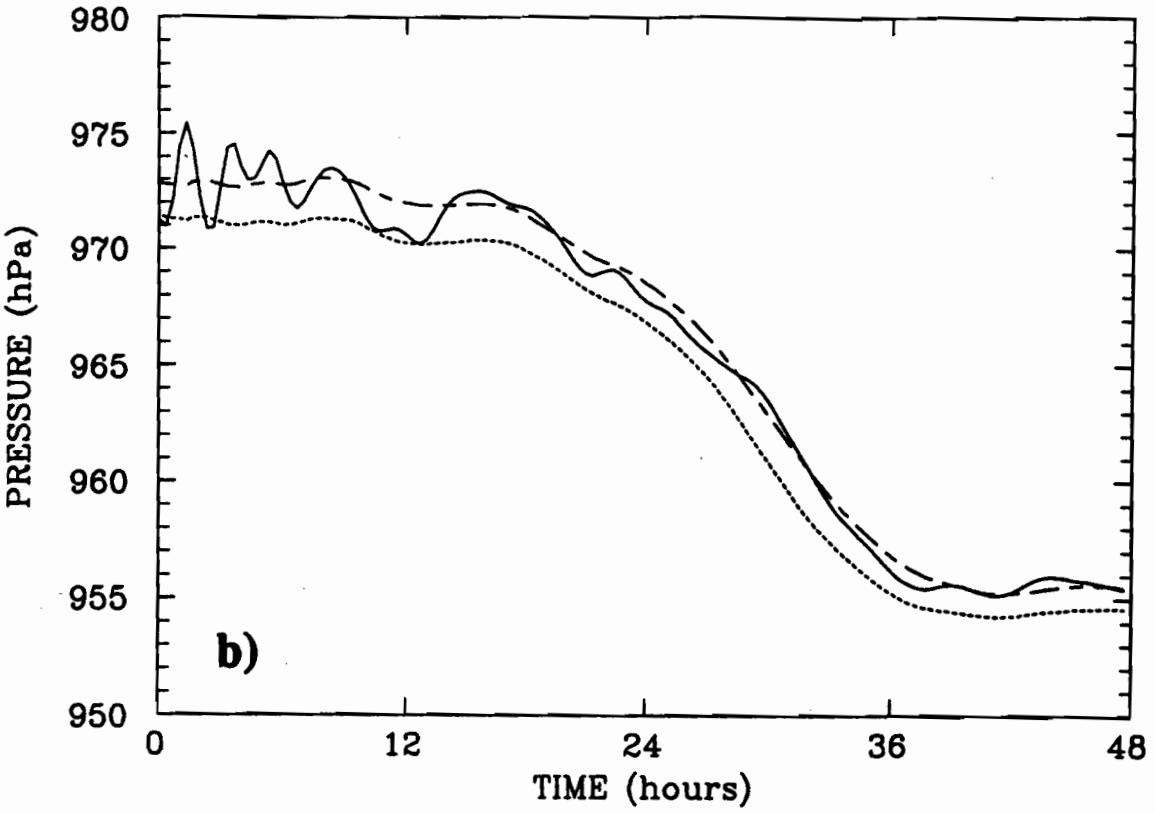


Fig. 17. Time-trace of surface pressure at selected grid points. (a) point 1; (b) point 4. Solid line: no initialization. Dashed line: unconstrained initialization (INMI, 3 iterations). Dotted line: variational initialization (VINMI-3D, 3 iterations).

Chapter VI

Conclusion

The idea of including the level of accuracy of analyzed fields (produced by optimal interpolation) into the nonlinear normal mode initialization procedure for numerical weather prediction was considered. We first investigated the applicability of a variational framework for INMI within simple shallow-water models. We were interested by the necessity to take into account the sharp distinction in reliability of analyzed fields over land and ocean within the nonlinear adjustment process between mass and wind fields. The extension to the multilevel case also appeared as a desirable aspect of the scheme since there is a known variation of reliability of analyzed fields with respect to the vertical.

Daley (1978) was the first to suggest a variational extension of nonlinear NMI. The essential difficulty associated with Daley's scheme is due to the dimensionality of the resulting variational problem due to the inherent "normal mode" space formulation. The emergence of an approximation of the "explicit " nonlinear NMI technique called "Implicit nonlinear NMI (INMI)" (Temperton 1988) which is formulated in physical space, was the basis for reconsidering Daley's proposal. By combining INMI with a given minimization criteria, the resulting problem could be studied in the context of horizontally and vertically varying weights to fully reflect the analysis accuracy.

The feasibility of variational explicit nonlinear NMI has already been demonstrated in the past for the spectral form of the shallow-water equations (Daley 1978, Tribbia 1982). To be considered as an original development, the present study needed to adress the feasibility and efficiency of VINMI for shallow-water models first. The efficiency problem was considered as an important practical aspect since it is essentially that point which limited the usefulness of variational "explicit" nonlinear NMI.

In chapter III, Machenhauer's scheme expressed in physical space was imposed as a strong constraint for the minimization of an energy type functional. Conservation of mass was further imposed to close the variational problem. Also, consistent boundary conditions were formulated for the model's fast and slow modes. The formulation allowed a strong variation of the weights in the horizontal in order to reflect the ocean-continents contrast in the analysis reliability. The INMI scheme developped by Temperton (1988) was shown in the present study to be a special case of VINMI when the weights on the rotational part of the wind and mass fields are set equal and constant. One important aspect distinguishes INMI and VINMI, that is VINMI has the ability to alter the evolution of the slow modes (to remain closer to the analysis) thus VINMI does generally lead to a *different* forecast of synoptic scale systems as compared to an uninitialized forecast. This aspect is not present in unconstrained explicit or implicit nonlinear NMI schemes since the slow modes of the analysis are left unchanged. One clearly sees here the importance of the specification of the variational weights in order to improve the performance of a data assimilation scheme. This aspect was not considered in the present study but will have to be examined carefully in the future. VINMI as formulated here does not control explicitly the changes on the divergent part of the analyzed flow. This restriction was *not* present in Daley's or Tribbia's variational scheme mentioned previously. As for INMI, VINMI is also unable to distinguish (in terms of frequency) which fast modes should be balanced. This latter aspect is the most serious one when this scheme is considered in a baroclinic context where (as shown by Errico 1989) the accuracy of Machenhauer's scheme rapidly diminishes when applied to gravity modes with resonant periods greater than 10 hours say. On a practical point of view, the VINMI scheme considered in chapter III deals with a variational adjustment of 30,000 variables in a very efficient way using an iterative procedure accelerated with a conjugate-gradient technique.

The VINMI scheme may also be used to improve the rather "brute" force adjustment imposed on the initial wind field near the wall boundary in order to satisfy the "no cross-boundary" mass flux necessary to integrate the RFE model. When directly applied, i.e. by

zeroing the normal component of the wind field on the wall boundaries, and without an appropriate adjustment of the mass field, this results in a local dynamical imbalance between the mass and wind fields which excites essentially the equatorially trapped gravity-inertia modes of the model. Applying the initialization scheme (INMI) on these modified fields will result in a redistribution of this original local imbalance throughout the horizontal domain. To limit the impact on midlatitude fields, VINMI could be used to explore the degree of flexibility of the method to control this problem. This aspect of the "initialization problem" for limited area models was not considered in the present work.

In the second part of this study, we extended in chapter IV the variational approach of chapter III to global spectral models. Using a triangularly truncated spherical harmonics representation up to zonal wavenumbers 63, similar results as those for the barotropic RFE model were obtained. Also, using an f-plane approximation, Leith's (1980) result was recovered due to the exact equivalence between explicit and implicit nonlinear NMI for that case. The familiar problem of divergence of the height-constrained initialization (a special case of VINMI where the weight on the rotational part of the wind field is much less than the weight on the mass field) was then put into evidence and related to the ellipticity condition for the nonlinear balance equation (see Charney 1955, Tribbia 1981).

The fact that the estimated errors of the analyzed fields have in addition to a strong horizontal contrast (see chapter IV) a vertical structure, asked for a consistent generalization of the VINMI method for multilevel models. This problem was faced in chapter V and it was shown that VINMI could naturally be extended to multilevel models using a known property of Machenhauer's scheme vis-à vis the linearized form of energy. The numerical procedure developed in chapters III and IV could also be naturally extended into three dimensional space without requiring excessive computer resources in terms of additional memory and computer time as compared to the operational version of the RFE model. The VINMI-3D scheme requires ~ 8% of the total computer time needed for a 48 hour forecast and compares favourably to the 2% required by the unconstrained INMI scheme. It is still possible to

improve this computer time but this has not been further considered in this study. Using two simple weight specification cases, the VINMI-3D scheme was shown to respond adequately to known structures of the variational weights. The horizontal and vertical structure of the changes on the analyzed fields were diagnosed for each cases. The adjustment process in 3D space requires an estimation of the error distribution of error variance of the analyzed fields in order to adequately set the initial state. The ensemble of weights used principally in this study were specified in order to demonstrate the potential and robustness of the VINMI scheme in extreme situations (altough still quite realistic, see chapter IV).

Future research based on the present work could be to extend the technique to

- Global variable resolution models of the atmosphere
- Mesoscale models.

Finally, the precise structure of the optimal weights to be used by the VINMI scheme has not been adressed in this study. This part is crucial in order that VINMI be beneficial for data assimilation.

BIBLIOGRAPHY

Andersen, J.H., 1977: A routine for normal mode initilization with nonlinear correction for a multilevel spectral model with triangular truncation. ECMWF Int. Rep. No. 15, 41 pp [ECMWF, Shinfield Park, Reading, Berkshire RG2 9AX, England.]

Baer, F., 1977: Adjustments of initial conditions required to suppress gravity oscillations in nonlinear flows. *Contrib. Atmos. Phys.*, **50**, 350-366.

Baer, F., and J. Tribbia, 1977: On complete filtering of gravity modes through non-linear initialization. *Mon. Wea. Rev.*, **105**, 1536-1539.

Ballish, B., 1980: Initialization, theory and application to the NMC spectral model. Ph.D, thesis, University of Maryland, 151 pp. [Dept. of Meteorology, University of Maryland, College Park, MD 20742.]

Benoit, R., J. Côté and J. Mailhot, 1989: Inclusion of a TKE boundary layer parameterization in the Canadian regional finite-element model. *Mon. Wea. Rev.*, **117**, 1726-1750.

Bourke, W., and J.L. McGregor, 1983: A nonlinear vertical mode initialization scheme for a limited area prediction model. *Mon. Wea. Rev.*, **111**, 2285-2297.

Brière, S., 1982: Nonlinear normal mode initialization of a limited area model. *Mon. Wea. Rev.*, **110**, 1166-1186.

Browning, G., A. Kasahara and H.O. Kreiss, 1980: Initialization of the primitive equations by the bounded derivative method. *J. Atmos. Sci.*, **37**, 1424-1436.

_____, and H.O. Kreiss, 1982: Initialization of the shallow-water equations with open boundaries by the bounded derivative method. *Tellus*, **34**, 334-351.

Burger, A. P., 1958: Scale considerations of planetary motions of the atmosphere. *Tellus*, **10**, 195-205.

Charney, J., 1955: The use of primitive equations of motion in numerical prediction. *Tellus*, **7**, 22-26.

Courant, R., and D. Hilbert 1962: *Partial differential equations. Methods of mathematical physics*. Vol. 2., Interscience, 830 pp.

Concus, P., and G.H. Golub, 1973: Use of fast direct methods for the efficient numerical solution of nonseparable elliptic equations, SIAM *J. Numer. Anal.*, **10**, 1103-1120.

_____, and D.P. O'leary, 1976: A generalized conjugate gradient method for the numerical solution of elliptic partial differential equations, *Sparse Matrix Computations*, J. R. Bunch and D.J. Rose, Eds. Academic Press, 309-332.

Courtier, P., and J.F. Geleyn, 1988: A global model with variable resolution—application to the shallow-water equations. *Quart. J. Roy. Meteor. Soc.*, **114**, 1321-1346.

Daley, R., 1978: Variational nonlinear normal mode initialization. *Tellus*, **30**, 201-218.

_____, 1979: The application of nonlinear normal mode initialization to an operational forecast model. *Atmos. Ocean*, **17**, 97-124.

_____, 1981: Normal mode initialization. *Rev. Geophys. Space Phys.*, **19**, 450-468.

_____, 1986 : The application of variational methods to initialization on the sphere. *Variational methods in geosciences*, Elsevier, New York, 309 pp.

Dickinson, R., and D.L. Williamson, 1972: Free oscillations of a discrete stratified fluid with application to numerical weather prediction. *J. Atmos. Sci.*, **29**, 623-640.

Errico, R.M., 1984: Dynamical balance produced by a general circulation model. *Mon. Wea. Rev.*, **112**, 2439-2454.

_____, and G.T. Bates, 1988 a: Implicit normal-mode initialization of the PSU/NCAR mesoscale model. *NCAR Tech. Note*, NCAR/TN-312+IA.112pp.

_____, and D.L.Williamson, 1988 b: The behaviour of gravitational modes in numerical forecasts. *Mon. Wea. Rev.*, **116**, 1737-1756.

_____, E.H. Barker, and R. Gelaro, 1988 c: A determination of balanced normal modes for two models. *Mon. Wea. Rev.*, **116**, 2717-2724.

_____, 1989 a: The forcing of gravitational modes by convective heating. *Mon. Wea. Rev.*, **117**, 2734-2752.

_____, 1989 b: The degree of Machenhauer balance in a climate model. *Mon. Wea. Rev.*, **117**, 2723-2733.

Fillion, L., and C. Temperton, 1989: Variational implicit normal mode initialization. *Mon. Wea. Rev.*, **117**, 2219-2229.

_____, 1991: Variational implicit normal mode initialization on the sphere. *Mon. Wea. Rev.*, **119**, 631-652.

Gandin, L. S., 1963: Objective analysis of Meteorological fields. Gidrometeorol. Izd., Leningrad (in Russian); English transl. Israel Program for Scientific Translations, Jerusalem, 1965, 242 pp. (available from NTIS, as N66-18047).

Ghil, M., 1989: Meteorological data assimilation for oceanographers. Part I: description and theoretical framework. *Dynamics of Atmospheres and Oceans.*, **13**, 171-218.

Haltiner, G.J., and R.T. Williams, 1980: *Numerical Weather Prediction and Dynamic Meteorology,* John Wiley, New York, 477 pp.

Hinkelmann, K., 1951: Der Mechanismus des meteorologischen Lärmes, *Tellus*, **3**, 285-296.

Houghton, D.D., 1968: Derivation of the elliptic condition for the balance equation in spherical coordinates. *J. Atmos. Sci.*, **25**, 927-928.

Juvanon du Vachat, R., 1986: A general formulation of normal modes for limited-area models: applications to initialization. *Mon. Wea. Rev.*, **114**, 2478-2487.

_____, 1988: Non-normal mode initialization: formulation and application to the inclusion of the beta terms in the linearization. *Mon. Wea. Rev.*, **116**, 2013-2024.

_____, 1989: Initialisation intrinsèque par modes normaux: Théorie et application à un modèle à domaine limité. Thèse de Doctorat de l'Université de Paris VI, septembre 1989.

Kasahara, A., 1976: Normal modes of ultralong waves in the atmosphere. *Mon. Wea Rev.*, **104**, 669-690.

Kasahara, A., 1982a: Nonlinear normal mode initialization and the bounded derivative method. *Rev. Geophys. Space Phys.*, **20**, 385-397.

_____, 1982b: Significance of nonelliptic regions in balanced flows of the tropical atmosphere. *Mon. Wea. Rev.*, **110**, 1956-1967.

Ko, S.D., J. Tribbia and J. Boyd, 1989: Energetics analysis of a multilevel global spectral model. Part I: Balanced energy and transient energy. *Mon. Wea. Rev.*, **117**, 1941-1954.

Kreiss, H.-O., 1979: Problems with different time scales for ordinary differential equations. *SIAM J. Num. Anal.*, **16**, 980-998.

_____, 1980: Problems with different time scales for partial differential equations. *Comm. Pure Appl. Math.*, **33**, 399-439.

Leith, C.E., 1980: Nonlinear normal mode initialization and quasigeostrophic theory. *J. Atmos. Sci.*, **37**, 954-964.

Longuet-Higgins, M. S., 1968: The eigenfunctions of Laplace's tidal equations over a sphere. *Phil. Trans. Roy. Soc. London*, A262, 511-607.

Lorenz, E. N., 1980: Attractor sets and quasi-geostrophic equillibrium. *J. Atmos. Sci.*, **37**, 1685-1699.

Lorenz, E. N., 1986: On the existence of a slow manifold. *J. Atmos. Sci.*, **43**, 1547-1557

Machenhauer, B., 1977: On the dynamics of gravity oscillations in a shallow-water model, with application to normal mode initialization. *Contrib. Atmos. Phys.*, **50**, 253-271.

_____, 1982: Fundamentals of nonlinear normal mode initialization. *Proceedings of the Fourteenth Stansted seminar*, publication in *Meteor.*, No. **127**, McGill University.

Mitchell, H., C. Charette, C. Chouinard, B. Brasnett, 1990: Revised interpolation statistics for the Canadian data assimilation procedure: Their derivation and application. **118**, in press.

Miyakoda, K., and R. W. Moyer, 1968: A method of initialization for dynamic weather forecasting. *Tellus*, **20**, 115-128.

Moura, A.D., 1976: The eigensolutions of the linearized balance equations over a sphere. *J. Atmos. Sci.*, **33**, 877-907.

Müller, R., 1989: On the structure of the global linearized primitive equations Part II: Laplace's tidal equations. *Beitr. Phys. Atmosph.*, **62**, 112-125.

Nitta, T. and J.B. Hovermale, 1969: A technique of objective analysis and initialization for the primitive forecast equations. *Mon. Wea. Rev.*, **97**, 652-658.

Phillips, N.A., 1960: On the problem of initial data for the primitive equations. *Tellus*, **12**, 121-126.

_____, 1981 : Variational analysis and the slow manifold. *Mon. Wea. Rev.*, ***109*** , 2415-2426.

Rutherford, I., 1972 : Data assimilation by statistical interpolation of forecast error fields. *J. Atnos. Sci.*, **29**, 809-815.

Rellich, F., 1934: Uber die Reduktion gewisser ausgearteter Systeme von partiellen Differentialgleichungen. Math. Ann., Vol. 109, pp. 714-745.

Sasaki, Y., 1958: An objective analysis based on the variational method. *J. Meteor. Soc. Japan*, **36**, 77-78.

Smirnov, V.I., 1964: *A course of higher mathematics*, Vol. I-V, 4th ed. Pergamon Press.

Staniforth, A.N., and H. Mitchell, 1978: A variable-resolution finite-element technique for regional forecasting with the primitive equations. *Mon. Wea. Rev.*, **106**, 439-447.

_____, A, M. Béland, and J. Côté, 1985: An analysis of the vertical structure equation. *Atmos. and Ocean.*, **23**, 323-358.

Tanguay, M., A. Simard and A. Staniforth, 1989 : A three dimensional semi-Lagrangian scheme for the Canadian regional finite element forecast model. *Mon. wea. Rev.*, **117**, 1861-1871.

_____, A. Robert and R. Laprise, 1990 : A semi-implicit semi-Lagrangian fully compressible regional forecast model. *Mon. Wea. Rev.*, **118**, in press.

Temperton, C., 1973: Some experiments in dynamic initialization for a simple primitive equation model. *Quart. J. Roy. Meteor. Soc.*, **99**, 303-319.

_____, 1976: Dynamic initialization for barotropic and multilevel models. *Quart. J. Roy. Meteor. Soc.*, **102**, 297-311.

_____, and D. Williamson, 1979: Normal mode initialization for a multilevel grid point model. *ECMWF te ical note*, Shinfield Park, Reading, Berkshire, RG9 2AX, England.

_____, 1984: Variational normal mode initialization for a multilevel model. *Mon. Wea. Rev.*, **112**, 2303-2316.

_____, 1985: Applications of a new principle for normal mode initialization. *Proceedings of the Seventh Conference on Numerical Weather Prediction of the American Meteorological Society*, Montréal, 105-107.

_____, 1988: Implicit normal mode initialization. *Mon. Wea. Rev.*, **116**, 1013-1031.

_____, 1989: Implicit normal mode initialization for spectral models. *Mon. Wea. Rev.*, **117**, 432-447.

_____, 1990 : Implicit normal mode initialization for the ECMWF model. Report of the WMO International symposium on assimilation of observations in meteorology and oceanography. Clermont-Ferrand (France) 9-13 july 1990.

_____, and M. Roch, 1991: Implicit normal mode initialization for an operational regional model. **119**, 667-677.

Thaning, L., 1984: On the existence of solutions to Machenhauer's non-linear normal mode initialization. *Tellus*, **36A**, 30-41.

Tribbia, J.J., 1979: Nonlinear initialization on an equatorial beta-plane.. *Mon. Wea. Rev.*, **107**, 704-713.

_____, 1981: Nonlinear normal-mode balancing and the ellipticity condition. *Mon. Wea. Rev.*, **109**, 1751-1761.

_____, 1982: On variational normal mode initialization. *Mon. Wea. Rev.*, **110**, 455-470.

_____, 1984: A simple scheme for higher order nonlinear normal mode initialization. *Mon. Wea. Rev.*, **112**, 278-284.

Wiin-Nielsen, A., 1979: On normal mode initialization on the sphere. *J. Atmos. Sci.*, **36**, 2040-2048.

Williamson, D.L., 1976: Normal mode initialization procedure applied to forecasts with the global shallow-water equations. *Mon. Wea. Rev.*, **104**, 195-206.

_____, and C. Temperton, 1981: Normal mode initialization for a multi-level grid point model. Part II : Nonlinear aspects. *Mon. Wea. Rev.*, **119**, 744 - 757.

_____, R. Daley, and T.W. Schlatter, 1981: The balance between mass and wind fields resulting from multivariate optimal interpolation. *Mon. Wea. Rev.*, **109**, 2357 - 2376.

_____, and R. Daley, 1983: A unified analysis - initialization technique. *Mon. Wea. Rev.*, **111**, 1517 - 1536.

Corrections to chapter III and IV

1) p.37 Dr. Luc Fillion should be Luc Fillion.

2) p.50 Dr. Luc Fillion should be Luc Fillion.

3) equation (2.5)

$$\hat{D}_n^m = i\,a\,[n(n+1)]\,\frac{1}{2}\,D_n^m \qquad (2.5)$$

4) p. 53 ... algebraic system (2.13) should be algebraic system (2.14).

5) p. 58 ϕ_D should be ω_D .