Tellus (1980), *32*, 198–206

The compatible balancing approach to initialization, and four-dimensional data assimilation

By MICHAEL GHIL, *Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, N.Y. 10012, U.S.A. and Laboratory for Atmospheric Sciences, NASA Goddard Space Flight Center, Greenbelt, MD. 20771, U.S.A.*

(Manuscript received July 30, 1979; in final form January 15, 1980)

ABSTRACT

We present an attempt at a unified theoretical approach to both the four-dimensional assimilation of asynoptic data and the initialization problem. This approach relies on the derivation of certain relationships between geopotential tendencies and tendencies of the horizontal velocity field in primitive-equation models of atmospheric flow. The approach is worked out and analyzed in detail for some simple barotropic models. Certain independent results of numerical experiments for the time-continuous assimilation of real asynoptic meterological data into a complex, baroclinic weather prediction model are discussed in the context of the present approach. Tentative inferences are drawn for practical assimilation procedures.

1. Introduction

The initialization problem, which has been with dynamic meteorology since the beginning of the use of primitive-equation systems for numerical weather prediction (NWP), has met with renewed interest in the last few years. The problem can be briefly described as follows: the initial data required to solve a baroclinic system of primitive equations in cartesian coordinates (x, y, z), say, are the horizontal velocity components u, v, the temperature T and the density ρ. These data are not available at the same time, either with sufficient accuracy, or with a density and spatial distribution comparable to that of the grid points in numerical prediction models currently used, over all regions of interest in numerical forecasting.

The initialization problem consists in finding ways to compensate for the lack of (i) *accuracy* and (ii) *completeness* in the data. The lack of completeness in the initial data will manifest itself in an erroneous specification of the large-scale structure of the atmosphere. It thus leads to an incorrect forecast of the slow, synoptic and planetary-scale waves which carry most of the meterologically significant information; their forecasting will deteriorate with simulated time due to predictability decay in the atmosphere itself, as well as due to physical and numerical inaccuracies in the model.

The lack of *accuracy* in the initial data manifests itself in the generation of fast inertia-gravity waves (Hinkelmann, 1951; Charney, 1955). The latter will not affect the forecast of the slow, meterologically significant waves, but do affect the vertical motion field and hence precipitation forecasts.

In the atmosphere inertia-gravity waves are present only with small amplitudes: in the process of *geostrophic adjustment* such waves are dispersed and eventually dissipated. Geostrophic adjustment results in a so-called balanced state, in which slow, meterologically significant motions are predominant.

The question of *accuracy*, tied with the concept of balanced, slow evolution, is at the basis of a number of recent approaches to the initialization problem, which bear considerable similarity to each other. Results were published by Baer (1977), Baer and Tribbia (1977), Browning et al. (1979), Daley (1978), Leith (1978), Machenhauer (1977), and Tribbia (1979). Some of these approaches require

0040-2826/80/030198-09$02.50/0

the computation of the normal modes of a linearized problem (Baer, 1977; Daley, 1978; Machenhauer, 1977), others do not (Kreiss, 1979, 1980). For lack of a better term, we shall call these approaches collectively "initial adjustment".

Initial adjustment consists in assuming the instantaneous availability of initial data corresponding only to the slow modes of the prognostic system, and in determining the remaining initial data necessary for solving the full system by a method which guarantees that the solution will only contain slow modes, at least for a certain time. The diagnostic system obtained thus requires as data only a subset of the complete initial data apparently required by the prognostic system; this seems to solve in a certain way also the question of completeness of data.

The procedure most widely used in current practice to address the *completeness* question is *updating*. This procedure is very well suited to the presently available asynoptic observational systems, in particular to satellite-borne temperature sounding instrumentation. In it, the NWP model is furnished additional data, specifically temperature data, as they become available, in order to compensate for the incompleteness of the initial data, especially for the lack of velocity field data (Charney et al., 1969). The theoretical justification for updating is the conjecture (Charney et al., 1969) that the time history of the mass field determines the velocity field for large-scale atmospheric flows. Many versions of the basic procedure above are known collectively as four-dimensional (4-D) data assimilation; they are reviewed in Bengtsson (1975).

The difficulties with updating are well known (cf. Ghil et al., 1977, for a summary and references). It suggests, however, a different approach to the initialization problem. In this approach, we assume the instantaneous availability of data on the mass field and on its time history; the time history in our analysis will be expressed mathematically as data on the first and second time derivatives of the mass field at the initial time. Such an assumption is a reasonable idealization of the data provided by presently available observational systems and data analysis procedures (e.g., Ghil et al., 1979a). Time derivatives could be interpreted in an operational context as time differences between suitably analyzed fields at successive synoptic or subsynoptic times (e.g., Bolin and Charney, 1951). An operationally feasible assimilation procedure capable of providing the required mass field history in a time-continuous form from asynoptic temperature data appears in Ghil et al. (1979a,b). Time derivatives, although not necessary to the practical procedure itself, can be computed from the time history *a posteriori*, by suitable difference formulae and for suitable time intervals.

We call the theoretical approach outlined in the previous paragraph *compatible balancing* (Ghil, 1975; Ghil et al., 1977). It leads to diagnostic systems which differ from those produced by the initial adjustment approach. Further mathematical results related to this approach have also been obtained by Bube and Oliger (1977) and by Bube (1978).

The purpose of the present article is to show that the velocity field obtained by solving such a diagnostic system, together with the given mass field, will also lead to a solution to the prognostic system which is free of fast modes; the necessary assumptions are similar to those required in the initial adjustment approach. This result will enable us to establish a connection between the practice of 4-D assimilation of asynoptic data and the theory of initialization.

In Section 2 we develop the theoretical approach for a linear shallow-fluid model. The diagnostic system compatible with the model is derived, and error estimates for the solution in terms of the data are outlined. In Section 3 we examine the diagnostic system compatible with a nonlinear shallow-fluid prognostic system (cf. Ghil et al., 1977). Again it is shown that the solution of the prognostic system will contain only slow modes under suitable assumptions on the mass field data.[1] Practical considerations are brought in and a comparison with the initial adjustment approach is made in Section 4.

2. Diagnostic equations for a linear shallow-fluid model

We develop the compatible balancing approach first for a model governed by the shallow-fluid equations linearized around a state of rest. The present section serves mostly to introduce in as

[1] Some results on baroclinic primitive-equation models appear in Ghil (1975) and will be omitted here.

simple a context as possible the concepts which will be used in Section 3; the latter contains our main results.

In a rotating Cartesian x,y-coordinate system the equations we study in this section are

$$u_t + \phi_x - fv = 0 \tag{2.1a}$$

$$v_t + \phi_y + fu = 0 \tag{2.1b}$$

$$\phi_t + \Phi(u_x + v_y) = 0 \tag{2.1c}$$

Here u,v are the velocity components in the x,y directions respectively and f is the Coriolis parameter. For convenience the geopotential ϕ is introduced instead of the height of the free surface h by $\phi = gh$, with g the acceleration of gravity; the equilibrium value of the geopotential is $\Phi =$ constant.

In this model the assumption of the mass field history being known is equivalent simply to assuming that we know $\phi = \phi\,(x,y,t)$, and hence in particular ϕ_t and ϕ_{tt} at any given instant.

2.1. *Derivation of the diagnostic system*

Our purpose with respect to system (2.1) is to obtain two equations in which u_t, v_t are absent, although ϕ_t, ϕ_{tt} may appear. Clearly (2.1c) itself is such an equation, and only one additional equation satisfying the requirements has to be derived. We proceed to do so by differentiating (2.1c) with respect to t, (2.1a) and (2.1b) with respect to x and y. This yields

$$u_{tx} + \phi_{xx} - fv_x = 0 \tag{2.2a}$$

$$v_{ty} + \phi_{yy} + fu_y = 0 \tag{2.2b}$$

$$\phi_{tt} + \Phi(u_{xt} + v_{yt}) = 0 \tag{2.2c}$$

We assume for simplicity in the sequel that $f =$ constant. Substituting u_{xt} and v_{yt} from (2.2a) and (2.2b) into (2.2c) yields

$$\Phi f(v_x - u_y) - \Phi(\phi_{xx} + \phi_{yy}) + \phi_{tt} = 0 \tag{2.3}$$

Hence (2.3) together with (2.1c) form a system of two first-order partial differential equations for u, v,

$$u_x + v_y = -\phi_t/\Phi \tag{2.4a}$$

$$u_y - v_x = -f^{-1}(\nabla^2\phi - \phi_{tt}/\Phi) \tag{2.4b}$$

here ∇^2 is the two-dimensional Laplacian operator,

$$\nabla^2\phi = \phi_{xx} + \phi_{yy}$$

Eqs. (2.4) are the required set of diagnostic equations for the model whose time evolution is given by the prognostic equations (2.1). We immediately notice that the instantaneous (synoptic) determination of u, v at $t = t_0$ from (2.4) actually requires only knowledge of ϕ_t, ϕ_{tt} at $t = t_0$, rather than the whole time history of ϕ from some $t = t_{-1} < t_0$ until $t = t_0$: knowledge of the time history over $t_{-1} < t < t_0$ would imply a knowledge of *all* the time derivatives of ϕ at $t = t_0$, not just the first two.

The linear system (2.4) is *elliptic*, i.e., it has no real characteristics (e.g., Ghil et al., 1977, p. 1225). In fact it is just a set of inhomogeneous Cauchy–Riemann equations for the functions v, u; by cross-differentiation these equations lead to a Poisson equation for either u or v. A boundary-value problem (BVP) having a unique solution for (2.4) would be the Dirichlet problem. This means prescribing u say on a closed contour ∂D. These boundary conditions determine u completely and v up to an additive constant in the domain D bounded by the contour ∂D; the value of this constant can be given by prescribing v at some point in D (e.g., Miranda, 1970, p. 265 ff.). Alternatively, one can prescribe v on ∂D and u at one point. Also u can be given on one part of ∂D and v on the rest. Hence, within the framework of model (2.1), solving the Dirichlet problem for system (2.4) solves the *completeness-of-data* problem. An efficient algorithm for the numerical solution of (2.4) appears in Ghil and Balgovind (1979).

We turn now to discussing the *accuracy* problem for the model at hand. The accuracy problem in our formulation can be handled in two ways. In one approach, one shows *first* that the solution (u, v) of system (2.4) in D depends continuously on the data of (2.4), i.e., on the mass field data ϕ, ϕ_t, ϕ_{tt} in D, and on the boundary data u, v on ∂D. *Second*, one shows that the solution of (2.1) depends smoothly on the initial data (ϕ, u, v) with ϕ given and u, v computed as the solution of (2.4). This approach has been carried out by Ghil (1975) and by Bube (1978). It has the disadvantage that it does not seem to generalize easily to the nonlinear problem of Section 3 or to baroclinic primitive-equation models. Hence we shall proceed here with another approach.

First, we shall show that for (2.1) the smallness of the geopotential tendencies ϕ_t and ϕ_{tt} at an instant $t = t_0$ implies the smallness of the wind tendencies u_t, v_t at the same instant as well. *Second*, we shall show that the smallness of initial tendencies ϕ_t, u_t, v_t at $t = t_0$ guarantees their smallness over a finite time interval $t_0 < t < t_1$.

We shall consider here the systems (2.1) and (2.4) in the entire plane, avoiding thus for the moment questions related to boundary conditions. The latter have been considered in Ghil (1975) and will be omitted here for the sake of simplicity.

It is well known (Hinkelmann, 1951; Williamson and Dickinson, 1972) that the linearized shallow-fluid equations have three independent plane-wave solutions, one corresponding to slow, geostrophic Rossby waves, the other two to rapid inertia-gravity waves propagating in opposite directions. In the case we treat, in which the unperturbed velocity is zero, the Rossby mode is actually stationary and the inertia-gravity waves have phase velocity

$$c = \pm(k^2 \Phi + f^2)^{1/2}/\mathbf{k} \tag{2.5}$$

where $\mathbf{k} = (k_1, k_2)$ is the wave vector and $k^2 = k_1^2 + k_2^2$. Any solution of (2.1) in the entire x,y-plane can be represented by a series expansion with respect to $\mathbf{k}$ in these plane waves. Thus small ϕ_t, u_t, v_t are equivalent to quasi-geostrophicity of the solution.

We start with a mathematical fact which is needed in the sequel:

Proposition P1. The solution (u, v) of the inhomogeneous Cauchy–Riemann equations

$$u_x + v_y = F(x,y) \tag{2.6a}$$

$$u_y - v_x = K(x,y) \tag{2.6b}$$

in the entire x,y-plane depends continuously on the right-hand sides F, K, provided F, K, u and v satisfy suitable conditions at infinity.[2]

In particular, if F and K satisfy certain smallness conditions in the sense of asymptotic theory, $F = O(\varepsilon) = K$, then u and v have the same order of magnitude of smallness, $u = O(\varepsilon) = v$.

This means that errors in the solution can be bounded by errors in the data, when using suitable error measures, or norms. For a more precise formulation and further details, we refer to the vast modern literature on elliptic partial differential equations (e.g., Miranda, 1970).

[2] This proposition holds and our whole discussion could be conducted also for F, K, u and v periodic, in x as well as in y (compare Bube, 1978).

2.2. *Dimensional analysis and smoothness proof*

It is convenient to carry out our argument on the nondimensional form of the equations. The underlying concept is that of two time scales (Blumen, 1972; Baer and Tribbia, 1977).

Let L be the characteristic length scale of the flow, U its characteristic velocity. The slow time scale is then

$$T = L/U \tag{2.7a}$$

Let f_0 be the characteristic value of the Coriolis parameter f. Then (2.5) shows that either $\theta_1 = 1/f_0$ or $\theta_2 = L/\sqrt{\Phi}$ could be used as a fast time scale. For simplicity we shall assume that these two scales are equal, $\theta_1 = \theta_2 = \theta$, or

$$L^2 f_0^2/\Phi = O(1) \tag{2.8}$$

This assumption is equivalent to

$$L/\lambda = O(1) \tag{2.8$'$}$$

where λ is the Rossby radius of deformation (Blumen, 1972),

$$\lambda = \sqrt{\Phi}/f_0 \tag{2.9}$$

Then we can take

$$\theta = 1/f_0 \tag{2.7b}$$

and the ratio of the two time scales is

$$\theta/T = U/Lf_0 = R_0 \tag{2.10}$$

i.e. the Rossby number R_0. We will be concentrating on mid-latitude synoptic motions, for which the Rossby number is small, $R_0 = O(10^{-1})$; we will denote it therefore in the sequel by ε. This is a reasonable approach since areas of sparse coverage for winds are in middle and high latitudes; tropical wind coverage is provided by geostationary satellites (Bengtsson, 1975; Ghil et al., 1977).

We introduce a non-dimensional *fast time* t^* and a non-dimensional *slow time* τ by

$$t^* = f_0 t \tag{2.11a}$$

$$\tau = (U/L)t = \varepsilon t^* \tag{2.11b}$$

The other variables are non-dimensionalized by

$$x = Lx', \quad y = Ly', \quad u = Uu', \quad v = Uv' \tag{2.12}$$

It is also useful to introduce (cf. Phillips, 1963)

$$\phi_0 = Lf_0 U \tag{2.13a}$$

as a characteristic deviation of the geopotential ϕ from its equilibrium value Φ, with

$$\phi = \Phi + \phi_0 \phi' \tag{2.13b}$$

After substituting the non-dimensional variables into (2.1) and dropping primes, we obtain

$$u_{t^*} + \varepsilon u_\tau + \phi_x - v = 0 \tag{2.14a}$$

$$v_{t^*} + \varepsilon v_\tau + \phi_y + u = 0 \tag{2.14b}$$

$$\phi_{t^*} + \varepsilon\phi_\tau + (\lambda^2/L^2)(u_x + v_y) = 0 \tag{2.14c}$$

Let

$$G \equiv \phi_x - v \tag{2.15a}$$

$$H \equiv \phi_y + u \tag{2.15b}$$

Then

$$G_x + H_y = \nabla^2 \phi + u_y - v_x \tag{2.16a}$$

$$G_y - H_x = -(u_x + v_y) \tag{2.16b}$$

The procedure of cross-differentiation which applied to (2.1) led to (2.4) yields in non-dimensional form, starting with (2.14), the equations

$$u_x + v_y = -(L^2/\lambda^2)(\phi_{t^*} + \varepsilon\phi_\tau) \tag{2.17a}$$

$$u_y - v_x + \nabla^2 \phi = (L^2/\lambda^2)(\phi_{t^*t^*} + 2\varepsilon\phi_{t^*\tau} + \varepsilon^2 \phi_{\tau\tau}) \tag{2.17b}$$

Combining (2.17) and (2.16) we obtain

$$G_x + H_y = (L/\lambda)^2(\partial/\partial t^* + \varepsilon\partial/\partial\tau)^2 \phi \tag{2.18a}$$

$$G_y - H_x = (L/\lambda)^2(\phi_{t^*} + \varepsilon\phi_\tau) \tag{2.18b}$$

If $\phi_{t^*} = O(\varepsilon)$, $\phi_{t^*t^*} = O(\varepsilon)$, and $\phi_\tau = O(1)$, $\phi_{\tau\tau} = O(1)$, it follows from (2.8′), (2.18) and Proposition P1 that $G = O(\varepsilon)$, $H = O(\varepsilon)$. Hence, by the definition (2.15) of G, H, and by (2.14a,b), we have $u_{t^*} = O(\varepsilon)$, $v_{t^*} = O(\varepsilon)$, $u_\tau = O(1)$, $v_\tau = O(1)$.

If the first time derivatives of ϕ, u, v are $O(\varepsilon)$ at $t = t_0$, they will stay so for any finite $O(1)$ time interval $t_0 < t < t_1$ (Kreiss, 1980). This completes our proof that small geopotential tendencies ϕ_t, ϕ_{tt} at $t = t_0$ will guarantee a slowly evolving solution to (2.1) with initial data at $t = t_0$ computed from (2.4). Furthermore it shows that small errors in the data ϕ_t, ϕ_{tt} for (2.4) will only lead to small changes in the solution to (2.1).

The observation that initial smallness of tendencies in system (2.1) is preserved over finite time intervals is at the core of the initial adjustment approach as well. It is called sometimes the *bounded derivative principle* (Browning et al., 1979; Kreiss, 1980). We shall use another form of it in the next section and comment further on differences in its application and interpretation in Section 4.

At this point we observe that differentiating (2.18) with respect to t will yield $(\partial/\partial t)^2(u,v) = O(\varepsilon)$, provided also $\phi_{ttt} = O(\varepsilon)$. By induction, $(\partial/\partial t)^{r+1}\phi = O(\varepsilon)$ implies $(\partial/\partial t)^r(u,v) = O(\varepsilon)$ at $t = t_0$, and hence for $t_0 < t < t_1$. Thus increased smoothness of the solution (ϕ, u, v) to (2.1) over $t_0 < t < t_1$ can be obtained, given greater smoothness of the initial geopotential data for (2.4) at $t = t_0$. In other words, if the past history of the geopotential for $t < t_0$ is such that the one-sided derivatives of ϕ at $t = t_0$ satisfy $(\partial/\partial t)^{r+1}\phi = O(\varepsilon)$, this will imply certain conditions on the data of systems (2.4) and (2.18). These conditions in turn will imply that the solution to these systems will satisfy certain relations which lead to the additional smoothness of the solution to (2.1) for $t > t_0$. The connection between spatial smoothness in (2.4) and smoothness in time for solutions of (2.1) results from the hyperbolic character of system (2.1) (John, 1975, pp. 48–86) in general, and from its special properties analyzed here in particular.

3. Diagnostic equations for a nonlinear shallow-fluid model

The nonlinear shallow-fluid equations are given by

$$u_t + uu_x + vu_y + \phi_x - fv = 0 \tag{3.1a}$$

$$v_t + uv_x + vv_y + \phi_y + fu = 0 \tag{3.1b}$$

$$\phi_t + u\phi_x + v\phi_y + \phi(u_x + v_y) = 0 \tag{3.1c}$$

The notation is the same as in Section 2. The corresponding diagnostic system was derived in Ghil et al. (1977). The derivation is similar to that of (2.4) from (2.1) and will not be repeated here. The diagnostic system corresponding to (3.1) is

$$u_x + v_y = -\Psi_x u - \Psi_y - \Psi_t \tag{3.2a}$$

$$u_x^2 + 2u_y v_x + v_y^2 + f(u_y - v_x) = f(\Psi_x v - \Psi_y u) + \frac{d\Psi_t}{dt} + u\frac{d\Psi_x}{dt} + v\frac{d\Psi_y}{dt} - \frac{\phi_x^2 + \phi_y^2}{\phi} - \nabla^2 \phi \tag{3.2b}$$

Here we introduced $\Psi = \log \phi$ (since $\phi > 0$), and $d/dt = \partial/\partial t + u\partial/\partial x + v\partial/\partial y$. System (3.2) reduces to the classical balance equation when $\delta = 0 = d\delta/dt$, with $\delta = u_x + v_y$. Furthermore, the diagnostic system corresponding to a *baroclinic* primitive-equation model in isentropic coordinates (Ghil et al., 1977) has exactly the same form as (3.2).[3]

In Ghil et al. (1977) it was shown that system (3.2) can be solved numerically in most situations of meterological interest. The solution is computationally much more efficient than initial data determination by dynamic initialization methods (e.g., Morel and Talagrand, 1974).

We shall show here that for initial data ϕ given and u, v computed from (3.2), the solution to (3.1) will be quasi-geostrophic, provided the initial geopotential tendencies ϕ_t, ϕ_{tt} are small. It is again convenient to work with the equations in non-dimensional form.

We introduce non-dimensional variables by (2.11–2.13), dropping primes again. In the sequel all variables will be understood to be non-dimensional, without carrying primes. For brevity, we shall also use the symbol ∂_t, to denote

$$\partial_t \equiv \partial/\partial t^* + \varepsilon\partial/\partial \tau \tag{3.3}$$

Assume now that: (A1) the non-dimensional geopotential data satisfy, for $\varepsilon \ll 1$,

$$\phi_x, \phi_y = O(1) \tag{3.4a}$$

$$\phi_t, \phi_{tt} = O(\varepsilon) \tag{3.4b}$$

and (A2) that the solution (u,v) of (3.2) with such data has a characteristic length $L_{(u,v)}$ which equals that of the data $L_{(\phi)}$,

$$L_{(u,v)} = L_{(\phi)} = L \tag{3.5a}$$

and a characteristic magnitude U such that the corresponding Rossby number is small,

$$R_0 \equiv U/f_0 L = \varepsilon \ll 1 \tag{3.5b}$$

Given (A1), it follows from the extensive numerical experiments of Ghil et al. (1977) that (A2) is a reasonable assumption.

Let the non-dimensional G, H be defined as before by

$$G = \phi_x - v \tag{3.6a}$$

$$H = \phi_y + u \tag{3.6b}$$

[3] A baroclinic primitive-equation model in *pressure* coordinates has also been treated in Ghil (1975).

By cross-differentiation of (3.6) one obtains again

$$G_x + H_y = \nabla^2 \phi + u_y - v_x \tag{3.7a}$$

$$G_y - H_x = -(u_x + v_y) \tag{3.7b}$$

Rewriting now (3.2) in non-dimensional form yields

$$-\frac{U}{L}(u_x + v_y) = \frac{f_0 \phi_0}{\Phi}\{\partial_t \phi + \varepsilon(u\phi_x + v\phi_y) + \ldots\} \tag{3.8a}$$

$$\begin{aligned}\frac{Uf_0}{L}(u_y - v_x) + \frac{\phi_0}{L^2}\nabla^2 \phi = -\frac{U^2}{L^2}\mathscr{H}(u,v) \\ + \frac{f_0 U\phi_0}{\Phi L}(\Psi_x v - \Psi_y u) + \frac{f_0^2 \phi_0}{\Phi}\{\partial_t^2 \Psi + \ldots\} \\ - \frac{\phi_0^2}{\Phi L^2}(\phi_x^2 + \phi_y^2)\end{aligned} \tag{3.8b}$$

where higher-order terms appear as dots in the curly brackets. In writing (3.8) we used the fact that a characteristic magnitude for ϕ is Φ, while for derivatives of ϕ, e.g., ϕ_x, ϕ_0/L is characteristic; this further implies that $\phi_0/\Phi L$ is characteristic for Ψ_x, and so on. In writing (3.8a) we also used assumptions (A1) and (A2), in particular (3.4a) and (3.5b), to non-dimensionalize the material derivative $d\phi/dt$. They are further used in (3.8b), both to write $d/dt = \partial_t + \varepsilon(u\partial/\partial_x + v\partial/\partial_y)$, and to compare dimensionally Ψ_t with $u\Psi_x$, $v\Psi_y$; as a result, the term of lowest order in the curly brackets is $\partial_t^2 \Psi$, and higher-order terms have been replaced by dots. The quantity

$$\mathscr{H}(u,v) = u_x^2 + 2u_y v_x + v_y^2 \tag{3.8c}$$

has been non-dimensionalized using (3.5a).

Use of (2.9) and of (2.13a) to express Φ and ϕ_0 in (3.8) gives

$$-(u_x + v_y) = (L^2/\lambda^2)(\partial_t \phi + \ldots) \tag{3.9a}$$

$$\begin{aligned}u_y - v_x + \nabla^2 \phi = -\varepsilon\mathscr{H}(u,v) + (L/\lambda)^2 \\ \times \{(\partial_t^2 \Psi + \ldots) + \varepsilon(\Psi_x v - \Psi_y u + \phi_x^2 + \phi_y^2)\}\end{aligned} \tag{3.9b}$$

with (3.4a) and (3.5b) again being used as needed and dots replacing $O(\varepsilon)$ terms in (3.9a). Finally, combining (3.9) with (3.7) and using (A1), i.e., (3.4b), we obtain

$$G_x + H_y = \{1 + O(L^2/\lambda^2)\}\, O(\varepsilon) \tag{3.10a}$$

$$G_y - H_x = (L/\lambda)^2\, O(\varepsilon) \tag{3.10b}$$

From (2.8′) and Proposition P1 it follows then that $G, H = O(\varepsilon)$, and hence (3.1) and (3.6) yield

$$du/dt = \partial_t u + \varepsilon(u\partial u/\partial x + v\partial u/\partial y) = O(\varepsilon) \quad (3.11a)$$

$$du/dt = \partial_t v + \varepsilon(u\partial v/\partial x + v\partial v/\partial y) = O(\varepsilon) \quad (3.11b)$$

This in turn implies $\partial_t u = O(\varepsilon) = \partial_t v$ at $t = t_0$.

The bounded derivative principle for (3.1) (discussed in detail in Browning *et al.*, 1979) implies that ϕ_t, u_t, v_t, if small at $t = t_0$, will stay so for a finite $O(1)$ time interval, $t_0 < t < t_1$. This completes the proof that small ϕ_t, ϕ_{tt} at $t = t_0$ will guarantee a solution to (3.1) which contains no high-frequency components for $t_0 < t < t_1$, given initial data computed from (3.2). Also if the errors in the geopotential data to (3.2) are small, the resulting errors in the solution to (3.1) will be small.

It is worthwhile commenting at this point on the requirement (2.8′), necessary in order to derive (3.11) from (3.10). The difficulties with geostrophic adjustment of ultra-long planetary waves, for which $L/\lambda \gg 1$, are well known (e.g., Washington, 1964); they have been addressed from the point of view of the initialization problem by Tribbia (1979). It appears that in practice ultra-long waves can still be handled by reasonable initialization procedures, even though less effectively than the shorter waves.

Differentiating (3.9) with respect to t will yield u_{tt}, $v_{tt} = O(\varepsilon)$ at $t = t_0$, provided $\phi_{ttt} = O(\varepsilon)$ initially. The only difference from the reasoning at the end of Section 2 is that u_t, v_t will appear on the right-hand side of (3.9) after differentiation. But these were already shown to be $O(\varepsilon)$. Successive differentiations of (3.9) will yield recursively $\partial_t^r(u,v) = O(\varepsilon)$, provided $\partial_t^{r+1}\phi = O(\varepsilon)$. If these relations hold at $t = t_0$, they will hold for $t_0 < t < t_1$. We thus conclude that solutions of (3.1) with initial data ϕ given and (u,v) computed from (3.2) will have high smoothness, provided sufficient smoothness of geopotential data is given initially.

4. Discussion and conclusions

The assimilation of asynoptic mass field data is an important ingredient in the operational determination of forecast initial states (Bengtsson, 1975). It is usually based on some form of *updating*, either time-continuous or intermittent. The updating approach to four-dimensional (4-D) assimilation of mass-field data is based on the conjecture that the history of the mass field determines the wind field for large-scale motions of the atmosphere (Charney et al., 1969). This conjecture was given a precise formulation for some simple atmospheric models (Ghil, 1975; Ghil et al., 1977) by prescribing the required mass field data to consist of the mass field itself, together with its first two time derivatives at the initial time. Within this formulation, the conjecture was proven analytically to be correct, namely the required mass field data do determine the wind field at the initial instant, and hence the complete initial state.

In the present article, based on the preceding sections' theoretical conclusions, we attempt to make a step in the direction of a unified approach to both initialization and 4-D data assimilation. In current practice, 4-D assimilation is used to construct an initial state from available synoptic and asynoptic information; this corresponds to solving what we called the *completeness-of-data* problem. Then some initialization procedure is used to change the previously computed initial state in such a way as to minimize "noise", or inertia-gravity wave activity, in the forecast carried out from the "initialized" state. The latter corresponds to solving what we called the *accuracy* problem.

The two problems are clearly related and it seems wasteful and inefficient, both theoretically and computationally, to solve them separately. We have shown in Sections 2 and 3 theoretically that for two simple models it is not necessary to do so. If the observed time history of the mass field is smooth enough, i.e., has small time derivatives, then the forecast from an initial state obtained using this time history will be noise-free. This statement was proven about the models themselves, independently of any practical procedure for computing and observing model solutions. Within the framework of the models analyzed here, the initial state would have to be obtained in practice by solving our diagnostic equations for the wind field. The theoretical implication, however, seems to be that an initial field obtained by any other reasonable procedure for the 4-D assimilation of asynoptic mass field data will also have the same property of leading to a relatively noise-free forecast. This tentative conclusion is supported by the preliminary evidence in Ghil et al. (1979a), where it was shown that time-continuous assimilation by linear-regression methods of remote-sounding temperature data in the Goddard Laboratory for Atmospheric Sciences (GLAS) model produced satisfactory initial states; high-frequency motions present to a certain extent at the beginning of forecasts were quickly and efficiently damped out by the model, without any special initialization procedure being applied. Clearly our conclusion is model-dependent and might not be supported by numerical experiments with other models, whose dispersive and dissipative properties are different. In this case, however, a well-matched forecasting model and assimilation scheme eliminated the need for a special initialization procedure.

Observed mass field history can be non-smooth for two reasons:

(a) rapid synoptic and sub-synoptic scale developments during the observational period preceding initial forecast time, or

(b) observational error.

Initial adjustment approaches consist in projecting the available data onto the slow manifold of the prognostic system (Leith, 1978). This projection can distort the data, unless additional constraints are imposed and the approach modified (Leith, 1978). Such distortion would appear to be particularly undesirable in situation (a) above, leading to large forecast errors within a short simulated time.

In situation (b), assimilation methods which take into account the statistical error structure of the observations should be able to produce the required amount of smoothing in time, as well as in space. The detailed synoptic study of the real-data experiments presented in Ghil et al. (1979a) seems to provide examples of both situations, and of their satisfactory handling by the statistical, time-continuous assimilation methods recommended there (Atlas et al., 1979): observational errors appear to be successfully reduced on the average. Furthermore, the assimilation-prediction system developed at GLAS led to significant improvements in the forecasting of severe weather over the United States in some synoptic situations occurring over the North Pacific which had led to major forecast errors over North America within 48–72 h when other assimilation methods were used. In any case, the theoretical analysis in this article, as well as the practical experience with the GLAS assimilation-prediction system, suggests that the careful combination of assimilation method and forecast model is particularly important in situations of rapid development.

Asynoptic information from satellites and other non-conventional observing systems is not limited to mass field data. Many other practical complexities cannot be handled by the theoretical approach presented here. A broader framework which can encompass completeness and accuracy problems, 4-D assimilation and initialization, appears to be provided by filtering and estimation theory. We are in the process of applying such an approach to the simple models treated here, and hope to extend it to real data and complex models in the near future (Ghil and Tavantzis, 1978; Ghil et al., 1979c).

5. Acknowledgements

It is a pleasure to acknowledge useful discussions at different times with Ferdinand Baer, Bert Bolin, Kenneth Bube, Milton Halem, Eugene Isaacson, Akira Kasahara, Heinz-Otto Kreiss, Peter D. Lax, Cecil E. Leith, Joseph Oliger, Ian D. Rutherford, Olivier Talagrand and Joseph J. Tribbia. A careful reading by Eugenia Kalnay-Rivas and the referees' comments were also very helpful in improving the presentation. This work was performed in part while the author was at the International Meterological Institute in Stockholm; it was supported in part by NASA Grant NSG-5130 at the Courant Institute.

REFERENCES

Atlas, R., Halem, M. and Ghil, M. 1979. Subjective evaluation of the combined influence of satellite temperature sounding data and increased model resolution on numerical weather forecasting. *Proc. Fourth Conf. NWP*, Amer. Meteorol. Soc., Boston, Mass., 319–328.

Baer, F. 1977. Adjustment of initial conditions required to suppress gravity oscillations in nonlinear flows. *Beitr. Phys. Atmos. 50*, 350–366.

Baer, F. and Tribbia, J. J. 1977. On complete filtering of gravity modes through nonlinear initialization. *Mon. Wea. Rev. 12*, 1536–1539.

Bengtsson, L. 1975. *4-Dimensional assimilation of meteorological observations.* GARP Publications Series, No. 15, WMO/ICSU, Geneva. [NTIS #76N10651, available from the WMO Secretariat, P.O. Box 5, CH-1211 Geneva, Switzerland]

Blumen, W. 1972. Geostrophic adjustment. *Rev. Geophys. Space Phys. 10*, 485–528.

Bolin, B. and Charney, J. 1951. Numerical tendency computations from the barotropic vorticity equation. *Tellus 3*, 248–255.

Browning, G., Kasahara, A. and Kreiss, H.-O. 1979. Initialization of the primitive equations by the bounded derivative method, NCAR Ms. 0501-79-4, submitted to *J. Atmos. Sci.*

Bube, K. P. 1978. *The construction of initial data for hyperbolic systems from nonstandard data.* Ph.D. Thesis, Stanford University. [Available from University Microfilms International, Ann Arbor, MI 48106, and also as Report STAN-CS-79-691, from Computer Science Dept., Stanford University, Stanford, CA 94305, U.S.A..]

Bube, K. P. and Oliger, K. 1977. Hyperbolic partial differential equations with nonstandard data. In *Advances in computer methods for partial differential equations II* (ed. R. Vichnevetsky). IMACS, Rutgers University, New Brunswick, NJ., U.S.A. 256–263.

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

Charney, J., Halem, M. and Jastrow, R. 1969. Use of incomplete historical data to infer the present state of the atmosphere. *J. Atmos. Sci. 26*, 1160–1163.

Daley, R. 1978. Variational non-linear normal mode initialization. *Tellus 30*, 201–218.

Ghil, M. 1975. *Initialization by compatible balancing,* Report No. 75-16, Institute for Computer Applications in Science and Engineering, Hampton, Virginia. [Available from ICASE, Mail Stop 132C, NASA Langley Research Center, Hampton, VA 23665, U.S.A., or from the author.]

Ghil, M. and Balgovind, R. 1979. A fast Cauchy–Riemann solver. *Math. Comput. 33*, 585–635.

Ghil, M. and Tavantzis, J. 1978. Recursive estimation for numerical weather prediction, in NASA Tech. Memo 80253, NASA Goddard Space Flight Center, Greenbelt, Md. 20771, U.S.A., 34–41.

Ghil, M., Shkoller, B. and Yangarber, V. 1977. A balanced diagnostic system compatible with a barotropic prognostic model. *Mon. Wea. Rev. 105*, 1223–1238.

Ghil, M., Halem, M. and Atlas, R. 1979a. Time-continuous assimilation of remote-sounding data and its effect on weather forecasting. *Mon. Wea. Rev. 107*, 140–171.

Ghil, M., Halem, M. and Atlas, R. 1979b. Effects of sounding temperature assimilation on weather forecasting: model dependence studies. In *Remote sounding of the atmosphere from space* (ed. H.-J. Bolle). Pergamon Press, 21–25.

Ghil, M., Tavantzis, J., Bube, K. and Isaacson, E. 1979c. Applications of estimation theory to numerical weather prediction. *Atmospheric and Oceanic Research Review—1979.* Laboratory for Atmospheric Sciences, NASA Goddard Space Flight Center, Greenbelt, MD 20771, U.S.A.

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

John, F. 1975. *Partial differential equations*, 2nd ed. Springer, New York.

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

Kreiss, H.-O. 1980. *Problems with different time scales for partial differential equations. Comm. Pure Appl. Math.*, in press.

Leith, C. E. 1978. Objective methods for weather prediction. *Ann. Rev. Fluid Mech. 10*, 107–128.

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

Miranda, C. 1970. *Partial differential equations of elliptic type*, 2nd ed. Springer, New York.

Morel, P. and Talagrand, O. 1974. Dynamic approach to meteorological data assimilation. *Tellus 25*, 334–344.

Phillips, N. 1963. Geostrophic motion. *Rev. Geophys. 1*, 123–164.

Tribbia, J. J. 1979. Nonlinear initialization on an equatorial beta-plane. *Mon. Wea. Rev. 107*, 704–713.

Washington, W. 1964. A note on the adjustment towards geostrophic equilibrium in a simple fluid system. *Tellus 16*, 530–534.

Williamson, D. L. and Dickinson, R. E. 1972. Periodic updating of meteorological variables. *J. Atmos. Sci. 29*, 190–193.

СОВМЕСТНЫЙ СБАЛАНСИРОВАННЫЙ ПОДХОД К ИНИЦИАЛИЗАЦИИ И ЧЕТЫРЕХМЕРНОЕ УСВОЕНИЕ ДАННЫХ

Мы представляем попытку единого подхода как к четырехмерному усвоению асиноптических данных, так и к проблеме инициализации. Подход основан на выводе определенных соотношений между тенденциями полей геопотенциала и горизонтальной скорости в моделях атмосферных течений с примитивными уравнениями. Подход разрабатывается и анализируется в деталях для некоторых простых баротропных моделей. Проводятся сравнения с результатами численных экспериментов при использовании данных и комплексной бароклинной модели прогноза погоды. Делаются предварительные выводы для практических процедур усвоения.