Tellus (1992), *44A*, 297–313

The effect of model errors in variational assimilation

By WERNER WERGEN, *Deutscher Wetterdienst, Frankfurter Strasse 135, D-6050 Offenbach, Germany*

(Manuscript received 17 April 1991; in final form 4 March 1992)

ABSTRACT

A linearized, one-dimensional shallow water model is used to investigate the effect of model errors in four-dimensional variational assimilation. A suitable initialization scheme for variational assimilation is proposed. Introducing deliberate phase speed errors in the model, the results from variational assimilation are compared to standard analysis/forecast cycle experiments. While the latter draws to the data and reflects the model errors only in the data-void areas, variational assimilation with the model used as strong constraint is shown to distribute the model errors over the entire analysis domain. The implications for verification and diagnostics are discussed. Temporal weighting of the observations can reduce the errors towards the end of the assimilation period, but may deteriorate the subsequent forecasts. An extension to variational assimilation is proposed, which seeks not only to determine the initial state from the observations but also some of the tunable parameters of the model. The potentional usefulness of this approach for parameterization studies and for a separation of forecast errors into model- and analysis errors is discussed. Finally, variational assimilations with the model used as weak constraint are presented. While showing a good performance in the assimilation, forecasts can suffer severely if the extra term in the equations up to which the model is enforced are unable to compensate for the real model error. In the discussion, an overall appraisal of both assimilation methods is given.

1. Introduction

Variational assimilation has in the last few years become an area of intensive research and experimentation (Courtier and Talagrand, 1990, and references therein). It offers a number of potential advantages over the current operational assimilation schemes. In particular, it allows in principle to translate temporal resolution into spatial resolution and might thus be ideally suited for the exploitation of future observing systems. Other advantages include the possibility to use observed quantities nonlinearly related to the model variables, e.g., satellite observed radiances. Also, all observations can be used at their correct location in time and space and no data selection processes are required. Thus, it might contribute to the solution of a number of problems currently of interest in meteorological data assimilation (Lorenc, 1988).

On the other hand, there are still a number of problems which need to be solved before an operational application of variational assimilation. They are both theoretical and practical. On the practical side, the computational costs of variational assimilation are at present at least an order of magnitude too high to allow an operational implementation. Even with further significant increases in computer power, the method will not be easily applicable in short range forecast centers, where typically less than one hour remains between the data cut-off and the dissemination of the one-day-forecast. On the theoretical side it is not yet clear how to deal with threshold processes, such as convection, in the assimilating model. Multiple minima might cause further problems. Also, the question of quality control of the observations is still open. Furthermore, the fact that models will always have errors needs to be taken into account. However, in most of the recent applications (e.g., Courtier and Talagrand, 1990) through the definition of the cost-function and by reducing the control variable to the initial state only, the model is implicitly assumed to be correct. It is not clear, how serious this assumption is. The present paper is meant to shed some light into this problem.

A one-dimensional, linearized shallow water model will be used as forecast vehicle. The linearization allows a largely analytical solution of the model equations and reduces the variational problem to a straightforward inversion of a linear system of equations. It thus allows a large number of experiments at low computational cost. At the same time the model is still meteorologically relevant, as it predicts both Rossby- and gravity modes. It therefore requires an explicit control of unwanted gravity wave activity during the assimilation.

The variational assimilation will be compared to the standard intermittent scheme (Bengtsson, 1975), consisting of an Optimum Interpolation (OI) analysis, a normal mode initialization (NMI), and a 6-hour-forecast serving as a first guess for the next analysis. An identical twin approach will be employed in which the true solution will be known. Both methods will be verified against that truth. Finally, an extended variational approach will be presented, in which not only the initial state, but also certain model parameters will be used as control variables. This extension can be interpreted as variational assimilation with weak constraint (Sasaki, 1970) or as variational tuning algorithm for the free parameters of the model.

2. The model

A one-dimensional, linearized shallow water model is chosen for the experiments. It still incorporates most of the meteorologically relevant processes while at the same time drastically reducing the computational costs. As a stratified linear model can be reduced to a series of shallow water models by projecting into vertical normal mode space, the subsequent results can readily be extended to baroclinic models. The only sacrifice made are the nonlinear terms.

2.1. The model equations

The model is formulated on a periodic mid-latitude β-plane. Thus, the Rossby-modes are non-stationary and dispersive. The equations are:

$$\frac{\partial \zeta}{\partial t} + \beta v + fD = K \frac{1}{r^2 \cos^2 \varphi} \frac{\partial^2 \zeta}{\partial \lambda^2} + Q_\zeta, \tag{1}$$

$$\frac{\partial D}{\partial t} + \beta u - f\zeta + \frac{1}{r^2 \cos^2 \varphi} \frac{\partial^2 \Phi}{\partial \lambda^2} = K \frac{1}{r^2 \cos^2 \varphi} \frac{\partial^2 D}{\partial \lambda^2} + Q_\mathrm{D}, \tag{2}$$

$$\frac{\partial \Phi}{\partial t} + \bar{\Phi} D = K \frac{1}{r^2 \cos^2 \varphi} \frac{\partial^2 \Phi}{\partial \lambda^2} + Q_\Phi. \tag{3}$$

Standard meteorological notation is used: ζ denotes the vorticity, D the divergence, Φ is the deviation of the geopotential from its mean value $\bar{\Phi}$. The Coriolis parameter is denoted by f, β is its meridional derivative, r is the radius of the earth, φ the latitude and K is a diffusion coefficient. The external forcing terms Q are assumed to be constant in time. They will be specified later.

2.2. Numerical solution

A normal mode approach (Machenhauer, 1977) is chosen for the solution of equations (1)–(3), as it allows a largely analytical treatment. Moreover, a normal mode decomposition will be required anyhow later on for the initialization process. First, the model, re-written in terms of streamfunction ψ and velocity potential χ, will be subjected to a Fourier transform.

$$\begin{pmatrix} \psi \\ \chi \\ \Phi \\ \mathrm{Q} \end{pmatrix} = \sum_{m=-M}^{M} \begin{pmatrix} \psi_m \\ i\chi_m \\ (m\sqrt{\bar{\Phi}}/r\cos\varphi)\Phi_m \\ Q_m \end{pmatrix} \mathrm{e}^{im\lambda}. \tag{4}$$

The factors in front of χ_m and Φ_m will facilitate the subsequent analysis, as they will lead to a Hermetian coefficient matrix. Inserting (4) into (1)–(3) and defining a vector z_m holding the predictive variables (ψ_m, χ_m, Φ_m) yields:

$$\frac{\mathrm{d}z_m}{\mathrm{d}t} = i\underline{\underline{A}}_m z_m + Q_m \tag{5}$$

The matrix $\underline{\underline{A}}_m$ is given by

$$\underline{\underline{A}}_m = \begin{bmatrix} \frac{\beta r \cos\varphi}{m} + iK\frac{m^2}{r^2\cos^2\varphi} & -f & 0 \\ -f & \frac{\beta r\cos\varphi}{m} + iK\frac{m^2}{r^2\cos^2\varphi} & \frac{m}{r\cos\varphi}\sqrt{\bar{\Phi}} \\ 0 & \frac{m}{r\cos\varphi}\sqrt{\bar{\Phi}} & iK\frac{m^2}{r^2\cos^2\varphi} \end{bmatrix}. \tag{6}$$

$\boldsymbol{Q}_m$ holds the amplitudes of the forcing terms. $\underline{\underline{A}}_m$ can easily decomposed into a matrix polynomial.

$$\underline{\underline{A}}_m = \frac{iKm^2}{r^2 \cos^2 \varphi} \underline{\underline{I}} + \underline{\underline{A}}'_m \tag{7}$$

$\underline{\underline{I}}$ is the identity matrix.

As $\underline{\underline{A}}'_m$ is symmetric, its eigenvalues l_j are real and its eigenvectors ar orthogonal. The eigenvalues Λ_j of $\underline{\underline{A}}_m$ are given by:

$$\Lambda_j = \frac{iKm^2}{r^2 \cos^2 \varphi} + l_j. \tag{8}$$

The eigenvectors of $\underline{\underline{A}}'_m$ are also the eigenvectors of $\underline{\underline{A}}_m$. They are the normal modes of the model. We proceed by projecting into normal mode space. With the 3 eigenvectors of $\underline{\underline{A}}$ making up the columns of the spectral matrix $\underline{\underline{S}}$ we write (dropping subscript m):

$$\boldsymbol{Z} = \underline{\underline{S}}\boldsymbol{Y}. \tag{9}$$

$\boldsymbol{Y}$ are the unknown amplitudes in normal mode space. Inserting (9) into (5) and multiplying from the left with $\underline{\underline{S}}^{-1}$ yields:

$$\frac{\mathrm{d}\boldsymbol{Y}}{\mathrm{d}t} = i\underline{\underline{S}}^{-1}\underline{\underline{A}}\,\underline{\underline{S}}\boldsymbol{Y} + \underline{\underline{S}}^{-1}\boldsymbol{Q}. \tag{10}$$

The similarity transform $\underline{\underline{S}}^{-1}\underline{\underline{A}}\,\underline{\underline{S}}$ reduces $\underline{\underline{A}}$ to a diagonal matrix with the eigenvalues Λ_j of $\underline{\underline{A}}$ as elements. Thus, we can write (10) in component form.

$$\frac{\mathrm{d}y_r}{\mathrm{d}t} = i\Lambda_r y_r + q_r. \tag{11}$$

The solution of (11) is:

$$y_r = \left(y_r(0) - \frac{iq_r}{\Lambda_r} \right) \mathrm{e}^{i\Lambda_r t} + \frac{iq_r}{\Lambda_r}. \tag{12}$$

Returning to matrix notation we can express $\boldsymbol{Y}$ as:

$$\boldsymbol{Y} = \underline{\underline{D}}(\boldsymbol{Y}(0) - \boldsymbol{P}) + \boldsymbol{P}, \tag{13}$$

where $\underline{\underline{D}}$ is a diagonal matrix with elements $\exp(i\Lambda_r t)$ and the entries of $\boldsymbol{P}$ are given by (iq_m/Λ_r).

Finally, inserting (9) into (13) yields:

$$\boldsymbol{Z}(t) = \underline{\underline{B}}(t)\, \boldsymbol{Z}(0) + \boldsymbol{E}^*, \tag{14}$$

where $\underline{\underline{B}}$ is given by $\underline{\underline{S}}\,\underline{\underline{D}}\,\underline{\underline{S}}^{-1}$ and $\boldsymbol{E}^*$ by $\underline{\underline{S}}(\boldsymbol{P} - \underline{\underline{D}}\boldsymbol{P})$. Eq. (14) allows to compute the amplitudes at any arbitrary time and, together with the Fourier transform, at any arbitrary point in space as a function of the initial amplitudes $z(0)$. This will be very convenient, as the required gradients with respect to the initial conditions can be worked out easily. Of course, all terms in (14) depend on the wavenumber m.

3. Experimental set-up

An identical twin approach was chosen for all experiments. The run generating the truth was started from the observed 500 hPa geopotential height field along 40°N for 12 UTC, 6 January 1986. Initial winds were derived geostrophically. Observation points were modelled after the northern hemisphere radiosonde network, with stations 2.5° apart over North America and 5° apart over Europe and Asia. This amounts to 52 stations in total. At these locations observations of u, v, and Φ were extracted from the truth every 6 hours and superimposed by normally distributed random noise. The RMS observation errors thus obtained amount to 88 $\mathrm{m^2\,s^{-2}}$ for the geopotential and to 2.9 $\mathrm{m\,s^{-1}}$ for the wind components. Using these observations, a variational and a standard analysis/forecast cycle assimilation were performed. Finally, the assimilated fields and the ensuing forecasts were compared to the truth.

4. Variational assimilation

4.1. Definition of the cost function

Initially the cost function J will measure only the distance of the model solution from the observations. In vector notation we write:

$$J_0 = \sum_{t=1}^{N} [\boldsymbol{H}(t) - \bar{\boldsymbol{H}}(t)]^{\mathrm{T}} [\boldsymbol{H}(t) - \bar{\boldsymbol{H}}(t)] \tag{15}$$

The overbar indicates observed quantities, superscript $^{\mathrm{T}}$ indicates the transpose of a vector or matrix. The vector $\boldsymbol{H}$ of the model predicted values at the observation points in (15) is given by:

$$\boldsymbol{H}=\left[v^1, v^2, \ldots, v^{N_v}, u^1, u^2, \ldots, u^{N_u}, \frac{\varepsilon_u}{\varepsilon_\Phi}(\Phi^1, \Phi^2, \ldots, \Phi^{N_\Phi})\right]^{\mathrm{T}}. \tag{16}$$

The superscript ranges over N_u, N_v, N_ϕ, the numbers of u, v, and ϕ observations, respectively; N is a number of time levels where observations are available, ε_u and ε_ϕ are the estimated RMS observations errors for wind and mass.

4.2. *Solution of the variational problem*

First, we express $\boldsymbol{H}(t)$ as function of the Fourier components of χ, ψ, and Φ at the initial time. We combine all the initial amplitudes into the vector of unknowns $\boldsymbol{X}$.

$$\boldsymbol{X}=\frac{1}{r\cos\varphi}\left[i\psi_1, -\chi_1, \frac{\varepsilon_u}{\varepsilon_\Phi}\Phi_1, i\psi_2, -\chi_2, \frac{\varepsilon_u}{\varepsilon_\Phi}\Phi_2, \ldots, i\psi_M, -\chi_M, \frac{\varepsilon_u}{\varepsilon_\Phi}\Phi_M\right]^{\mathrm{T}}. \tag{17}$$

By using eq. (14), $\boldsymbol{H}$ can now be expressed as:

$$\boldsymbol{H}=\underline{\underline{F}}\,\underline{\underline{G}}\boldsymbol{X}+\underline{\underline{F}}\boldsymbol{E}. \tag{18}$$

$\underline{\underline{F}}$ is the Fourier transform matrix with elements $f_{jm}=\exp(im\lambda_j)$ transforming from the Fourier components to the values at observation points. $\underline{\underline{G}}$ is a block diagonal matrix with the time dependent 3×3 matrices $\underline{\underline{B}}_m$ from eq. (14) as elements, $\boldsymbol{E}$ holds the forcing terms $\boldsymbol{E}_m$ for all wavenumbers. Inserting (18) into (15) reduces the problem of finding the minimum of J_0 to the solution of an overdetermined linear algebraic system in a least square sense. Thus, no recourse to the adjoint method as pioneered by Le Dimet and Talagrand (1986) for variational assimilation is required. Minimizing (15) thus yields:

$$\sum_{t=1}^{N}[\underline{\underline{F}}\,\underline{\underline{G}}(t)]^{\mathrm{T}}[\underline{\underline{F}}\,\underline{\underline{G}}(t)]\boldsymbol{X} = \sum_{t=1}^{N}[\underline{\underline{F}}\,\underline{\underline{G}}(t)]^{\mathrm{T}}[\bar{\boldsymbol{H}}(t)-\underline{\underline{F}}\boldsymbol{E}(t)]. \tag{19}$$

Of course, in order to be properly posed, the number of observations during the assimilation interval has to exceed the degrees of freedom of the model. For formal convenience we re-write (19) as

$$\underline{\underline{V}}\boldsymbol{X}=\boldsymbol{W}. \tag{20}$$

$\underline{\underline{V}}$ is a positive definite matrix of the order $3\times\mathrm{M}$. The standard Cholesky algorithm can be used to solve (20). Extending the assimilation interval simply means adding new summation terms to the elements of $\underline{\underline{V}}$ and $\boldsymbol{W}$.

4.3. *Initialization*

In principle, the variational scheme will use all degrees of freedom of the model to fit the data. Noisy observations and/or errors in the assimilating model might lead to unrealistic gravity-mode amplitudes. Therefore, an explicit control of unwanted gravity wave activity is required during the assimilation.

Courtier and Talagrand (1990) proposed a scheme consisting of a combination of an additional penalty term for the gravity mode amplitudes in the cost function and a standard NMI before each integration of the model in the descent process. This combined scheme was shown to converge and to yield the required balance.

Here, we will follow a different approach based on the very nature of normal mode initialization. In its standard form (Machenhauer, 1977), NMI first splits the analysed fields into their Rossby and gravity-mode parts. While leaving the Rossby modes untouched, it modifies the gravity mode amplitudes in such a way that the resulting gravity mode tendencies vanish. In essence, only the Rossby mode projection of the final analysis stems from the data, the gravity-mode projection is diagnosed from the analysed Rossby modes (and a suitably derived diabatic forcing) through the initialization condition.

It is straightforward to transfer this concept to variational assimilation: We include only the Rossby mode projection of the model variables in vector $\boldsymbol{H}$ of forecast values at observation points. With subscript R denoting the Rossby mode projection, we re-write the cost function (15) as:

$$J^I=\sum_{t=1}^{N}[\boldsymbol{H}_{\mathrm{R}}(t)-\bar{\boldsymbol{H}}(t)]^{\mathrm{T}} \times[\boldsymbol{H}_{\mathrm{R}}(t)-\bar{\boldsymbol{H}}(t)]. \tag{21}$$

The required modification can simply be implemented by setting the gravity mode eigenvectors in the spectral matrix $\underline{\underline{S}}$ to zero. Once the Rossby mode information is extracted from the observations by minimizing (21), the gravity mode amplitudes $y_{E,W}$ at the beginning of the assimilation interval are diagnosed from (12) as:

$$y^{I}_{E,W} = \frac{iq_{E,W}}{\Lambda_{E,W}}. \tag{22}$$

The indicies $_E$ and $_W$ denote east- and westward travelling gravity modes, superscript ^{I} indicates the initialized values.

5. Standard assimilation

The standard assimilation is of the intermittent form consisting of an Optimum Interpolation analysis, an (optional) normal mode initialization and a 6-h forecast serving as first guess for the next analysis. The OI analysis largely follows Lorenc (1981). No "super observation" formation or quality control is performed. The values of u, v, and Φ are analysed on $2 \times M$ equally spaced grid-points. Only the deviations of the first guess from the observations are analysed. First guess values are interpolated linearly to the observation points. The analysis is done globally, i.e., all observations are used to analyse all grid-points and variables. The analysis is multivariate in height and stream-function, based on the geostrophic constraint with a coupling coefficient of 0.95. Error covariances for the divergent wind are assumed to vanish. First guess height errors are modelled as the area mean of the RMS difference between the 6-h forecast and the truth. All assimilations start from a first guess with zero values at all gridpoints. The height-height correlation is assumed to be homogeneous and isotropic and is modelled as a Gaussian function with the correlation length scale of 600 km. With respect to the resolution of the model, this value might appear quite small. However, in the continental areas the data coverage is high enough not to require much interpolation. A higher value for the length scale would result in more averaging of the observations for the analysis of a grid point. On the positive side, it would lead to a better extrapolation into the data-void areas. The proposed value was chosen experimentally on the basis of the smallest analysis and forecast departures from the truth. The required wind errors and correlations follow from the height-stream function relation as given by Lorenc (1981). The observation errors ε_u and ε_ϕ are the same as the variational assimilation; the observation errors are assumed to be uncorrelated.

The (optional) initialization is based on the familiar Machenhauer (1977) scheme. Before the start of each 6-h forecast the analysed fields are initialized according to (22). The 6-h forecast is based on (14). The entire forecasting process reduces to a Fourier transform, a matrix multiplication, and an inverse Fourier transform.

6. Results

Before proceeding with the discussion of the results, it remains to specify the forcing terms Q in (1)–(3). We set Q_ζ and Q_D to zero while Q_ϕ is set to

$$Q_\Phi = b \cos \lambda. \tag{23}$$

This form is ment to model the large scale mass field forcing arising from land-sea contrasts. The parameter b is assumed to be 3.4 $m^2 s^{-3}$.

Table 1 summarizes the various parameters used in the experiments.

6.1. Correct model

First, assimilations with the correct model will be presented. This will serve as a control demonstrating the performance of both schemes under optimal conditions. It will also highlight the need for a further modification of the cost function. Finally, it will allow an assessment of the proposed initialization scheme for variational assimilation.

Fig. 1 shows the results after 24 h of assimilation. The top two panels show the meridional and zonal wind components in $m s^{-1}$, the bottom panel is for the geopotential in $m^2 s^{-2}$. The observations are marked by circles. Note the poor signal-to-noise ratio in the u-component which is purely divergent in this model. In well observed areas the two methods are close for the v-component and the geopotential. The divergent wind (center) is very noisy for the standard assimilation (dashed), while the variational scheme (dotted) is close to the truth (full). The

Table 1. *Parameters used in the experiments*

Parameter	Meaning	value	(Units)
φ	latitude	50	degrees
f	Coriolis parameter	$1.11 \cdot 10^{-4}$	s^{-1}
β	meridional derivative of f	$1.47 \cdot 10^{-11}$	$(ms)^{-1}$
Φ	mean geopotential	$11.55 \cdot 10^{4}$	$m^2 s^{-2}$
r	earth radius	$6.371 \cdot 10^{6}$	m
K	diffusion coefficient	$6.7 \cdot 10^{5}$	$m^2 s^{-1}$
M	maximum wave number	17	
ε_u	estimated RMS wind error	3	$m s^{-1}$
ε_ϕ	estimated RMS geopotential error	88	$m^2 s^{-2}$
b	external forcing parameter	3.4	$m^2 s^{-3}$

Forced primitive equation model
Assimilation after 24 hours

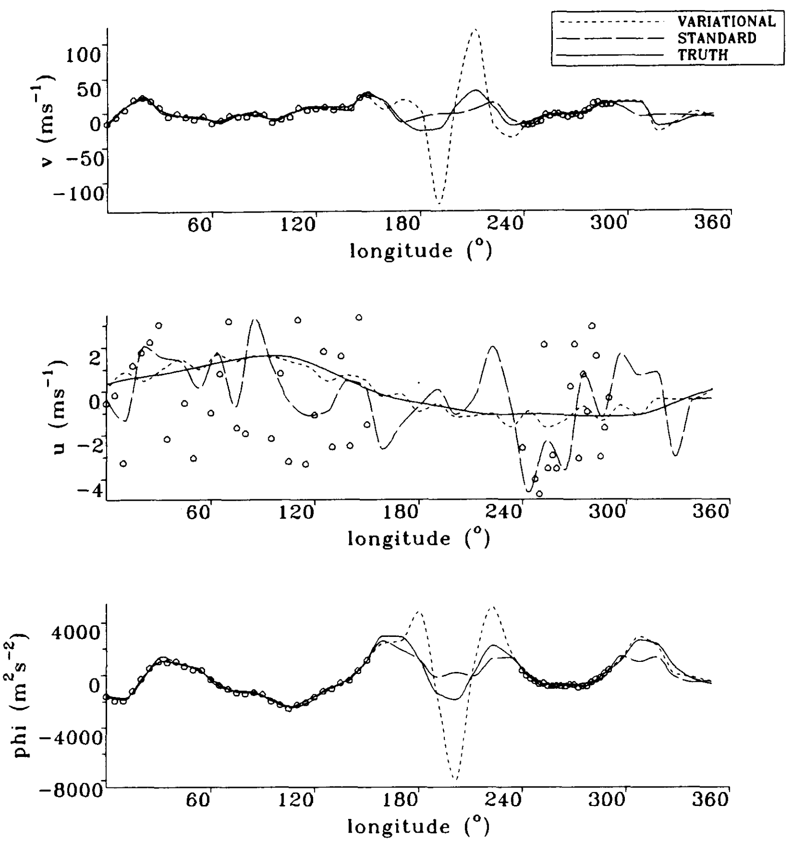

Fig. 1. Analysed v (top), u (middle) and ϕ (bottom) fields after 24 h of standard (dashed) and variational (dotted) assimilation as a function of longitude. The full line indicates the truth, the circles mark the observations.

poor result for the analysis of the divergent wind by the standard scheme is a consequence of the non-divergent error covariances used. In data void areas, particularly in the Pacific, the variational scheme shows very large, unrealistic oscillations. By contrast, in the standard scheme some information has propagated from the continents into the ocean areas so that the analysed fields are closer to the truth. Extending the assimilation in time would further reduce the errors.

The large oscillations of the variational scheme in data void areas are a well-known property of global least-square fitting methods as for instance reported by Sasaki (1970). They are a result of the noisy data. This was verified by a run with no noise added to the data. It yielded a perfect variational analysis. The fact that the divergent wind does not show the problem indicates that the gravity modes are not involved. Therefore, the problem can not be cured by initialization. Similar to Courtier and Talagrand (1987) we include an additional term in the cost function, which penalizes the "roughness" of the field. This can simply be achieved by re-defining J as

$$J_2 = J^I + \sum_m \alpha m^4 \left[u_m u_m^* + v_m v_m^* + \left(\frac{\varepsilon_u}{\varepsilon_\Phi}\right)^2 \Phi_m \Phi_m^* \right], \tag{24}$$

where the superscript * indicates the complex conjugate of a spectral coefficient and α is an arbitrary tuning parameter. As there is no obvious algorithm for estimating an optimal value for α, it was

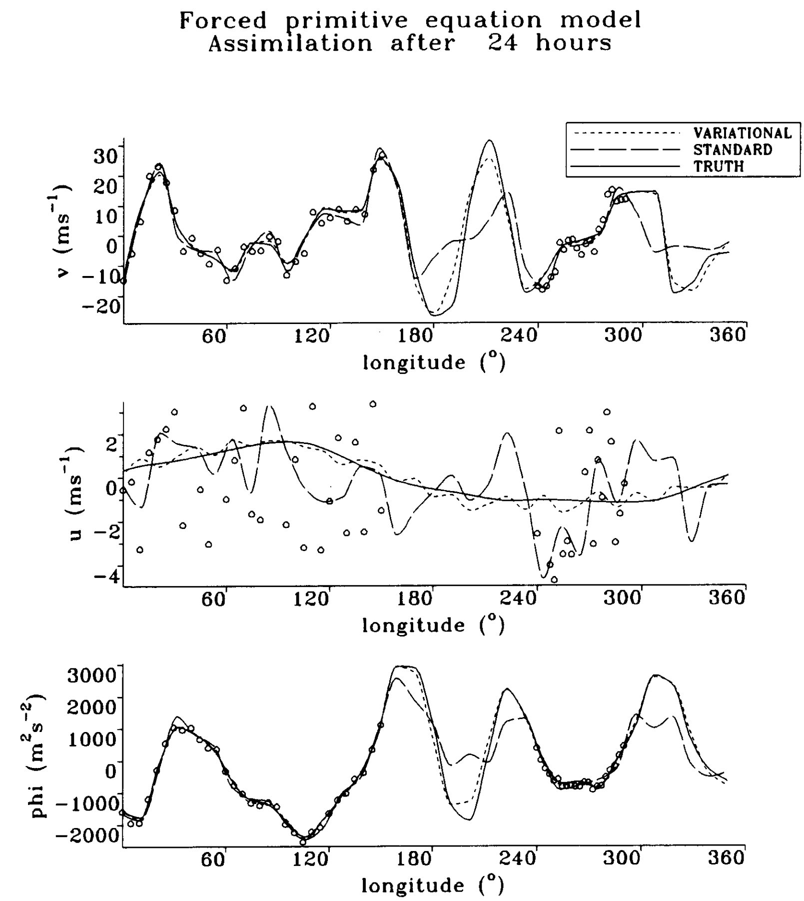

Fig. 2. As Fig. 1, but with additional smoothing term in cost function for variational assimilation. $\alpha = 10^{-5}$.

determined experimentally for each run. It turned out that it varied over several orders of magnitude, depending on the experimental setting. This large variability is a somewhat unsatisfactory aspect of variational assimilation.

Through the extra penalty term in (24), we have incorporated prior knowledge (Lorenc, 1986) into the assimilation process. We require, that the analysed fields are smooth. In the absence of any observations, the scheme would yield a flat analysis. This is consistent with the usage of a flat first guess for the initial analysis in the standard scheme. It too would result in a flat final analysis in the absence of observations. Experiments with different forms of the smoothness constraint showed little sensitivity of the results to the actual formulation.

Fig. 2 shows the result after 24 h of assimilation with α set to 10^{-5}. Note, that the scale has changed. The unrealistic oscillations have disappeared and the variational analysis is very close to the correct solution. The scheme has successfully extracted all information from the well observed areas and used the (perfect) model to transport the information into the data void areas. The standard solution can also be improved upon by extending the assimilation in time. This was verified in a separate experiment (not shown). As most operational centers have been running their assimilation systems for years without disruption, the discussed superiority of the variational scheme is perhaps rather more theoretical, at least in the context of a perfect model.

The effect of initialization is demonstrated in Fig. 3. As expected, it influences most the divergent wind component (middle panel). All the noise has

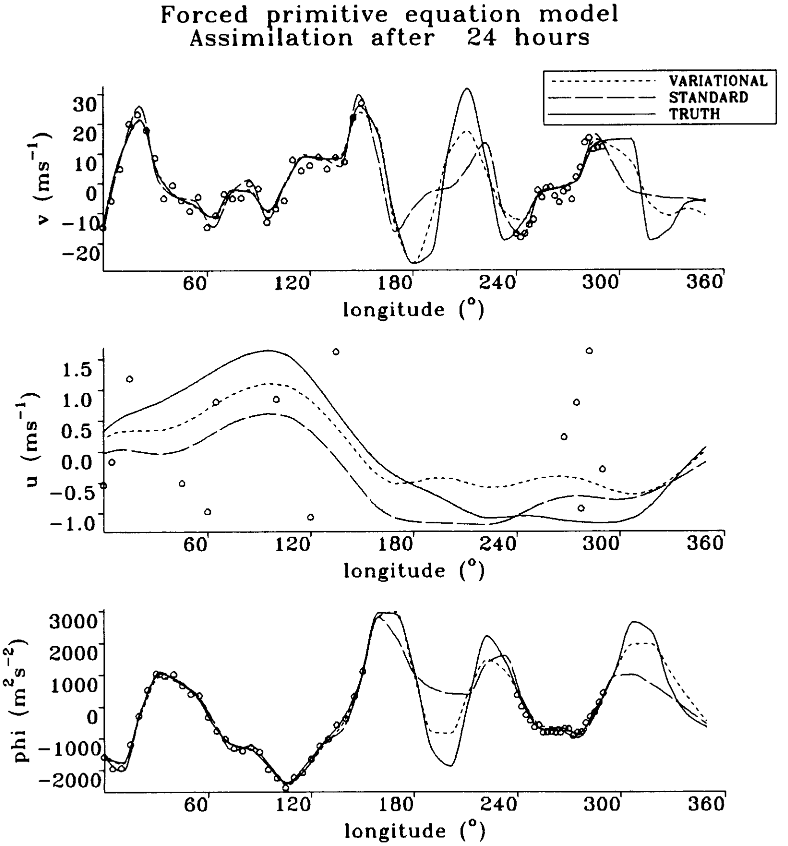

Fig. 3. Same as Fig. 2, but with normal mode initialization in variational and standard assimilation. $\alpha = 10^{-3}$.

disappeared in the analysed fields while most of the large scale features are correctly reproduced. The v-component and the mass-field are a little further away from the truth than in the case without initialization. As the truth (and the observations) were generated from geostrophic initial conditions, they contain gravitational energy because of the β-terms in the model. The initialization diagnoses the gravity mode amplitudes from the forcing terms. This is only an approximation which cannot be expected to exactly recover the original amplitudes. However, from the diagrams shown it is concluded that the analyses are acceptable regarding the high level of noise in the u-component. Again, the variational assimilation is superior after a 24-h cycle.

The fact that the proposed initialization scheme for the variational assimilation seems to work has important implications for the practical implementation of the variational approach. As it reduces the degrees of freedom of the problem, it requires less computer time. Also, as mentioned earlier, in order to be properly posed, the number of independent observations must exceed the number of degrees of freedom of the model. Fitting only the Rossby modes might therefore allow to shorten the assimilation interval. Whether the proposed initialization scheme will work in a different

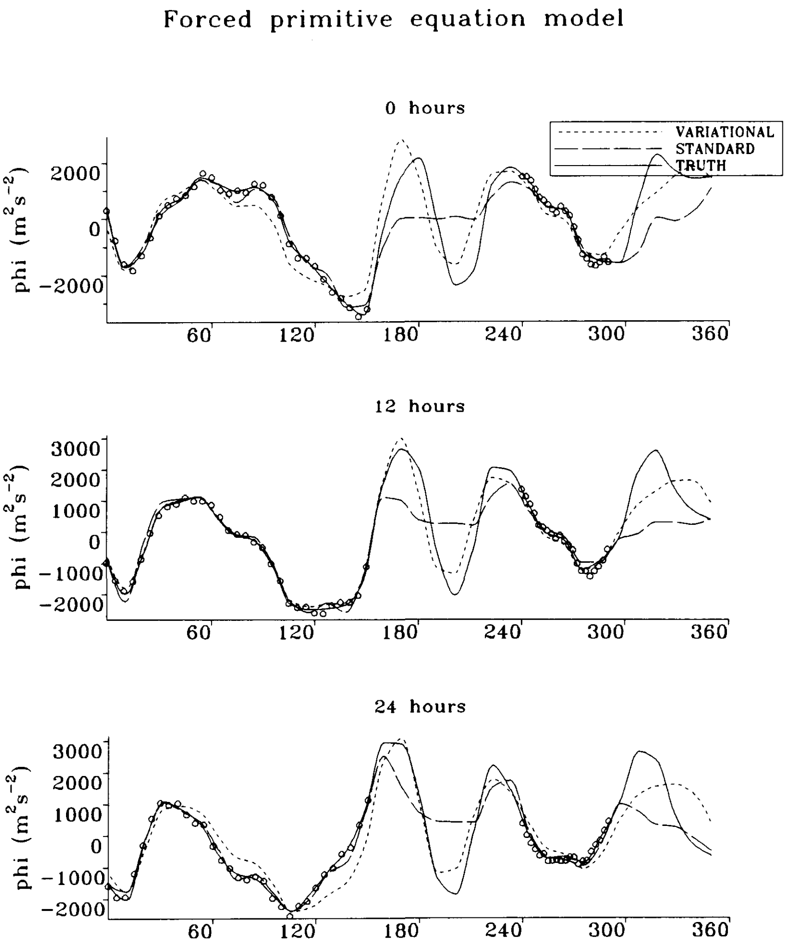

Fig. 4. Analysed geopotential at the beginning (top), middle (center) and end (bottom) of the standard (dashed) and variational (dotted) assimilation as a function of longitude. A 50% phase speed error for the Rossby modes was introduced in the model. The full line indicates the truth, observations are marked by circles. $\alpha = 10^{-3}$.

environment will be the subject of future work. In particular, its performance when only mass- or only wind-data are available and the importance of constraints needs to be investigated. Also, tropical aspects need to be addressed.

6.2. *Phase speed errors*

After having discussed the performance of both schemes under optimal conditions, we now consider the more realistic case with errors in the assimilating model. We introduce a 50% phase speed reduction for all Rossby waves. This error is not considered realistic; it merely serves to highlight qualitatively how both schemes react on model errors.

Fig. 4 shows the resulting initialized geopotential analyses for the beginning (top), the middle (center), and the end (bottom) of a 24-h assimilation. At the end of the assimilation (bottom) the characteristic difference between the 2 methods is obvious: The standard scheme draws to the data and reflects the model errors in the data void areas. By contrast, the variational scheme distributes the model errors over the entire analysis domain, thus leading to considerable analysis errors even in areas with good data coverage. Inspection of the time sequence reveals how the variational scheme tries to minimize the cost function. As no zonal mean wind has been used in the linearization, all Rossby waves are travelling westward, i.e., from right to left. At the beginning, the variational analysis is put in front of the truth, at the middle it is roughly in phase with the truth and after 24 h it is lagging behind. The best analysis is obtained for the center of the interval where information from both the past and the future are available to the scheme. Again, the standard assimilation can be further improved upon by extending the assimilation interval. By contrast, extending the interval would aggrevate the problems in varia-

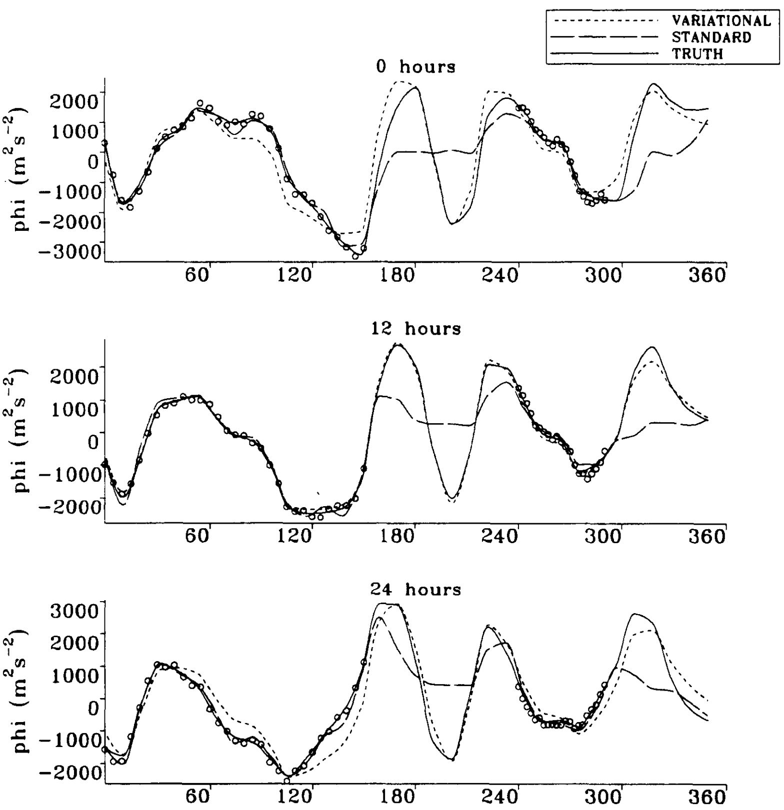

Fig. 5. Similar to Fig. 4, but smoothness term in cost function replaced by squared difference to the truth, $\varepsilon = 10$.

tional assimilation, if the model has errors. One would rather like to shorten the assimilation period. However, we need more observations than degrees of freedom and the signal must be given enough time to travel between the observations.

One might argue that the simple form of the additional smoothness term in (24) does not do justice to the variational scheme. One could try to incorporate more adequate prior knowledge, e.g., through the pragmatic use of first guess fields from a parallel OI assimilation. In an attempt to simulate such an approach, an experiment was run where the smoothness term in (24) was replaced by the squared deviation from the truth. The definition of the cost function is thus given by:

$$J_2^I = J^I + \varepsilon \sum_m (u_m - \tilde{u}_m)(u_m - \tilde{u}_m)^* + (v_m - \tilde{v}_m)(v_m - \tilde{v}_m)^* + \left(\frac{\varepsilon_u}{\varepsilon_\Phi}\right)^2 (\Phi_m - \tilde{\Phi}_m)(\Phi_m - \tilde{\Phi}_m)^*, \quad (25)$$

where the symbol $\tilde{}$ indicates the truth and the superscript * denotes the complex conjugate. The weighting ε of the truth was determined experimentally so as to control the unwanted oscillations but not to dominate the solution entirely. This experiment most favourable for the variational scheme would return the truth in the absence of any observations.

Fig. 5 shows the results for an experiment with $\varepsilon = 10$. It is obvious that the problems highlighted in Fig. 4 still exist. The variational solution lags behind at the end of the assimilation interval and fits the data less well than the standard scheme. Thus we conclude that it is not the form of the prior knowledge; it is the disregard of model errors in the variational scheme which causes the problems. As assimilations with the truth as first guess are highly unrealistic, all remaining experiments are done with the smoothness constraint (24).

The reflection of model errors in the entire domain has serious consequences if variational analyses are to be used for verification and diagnostics. They would probably increase the scores, as the verifying analyses reflect the model and its errors to a much larger extent than the standard scheme. Obviously, these improvements would be fictitious. Diagnostic studies would also be dealing more with the model than with the real atmosphere. Of course, in the present example the problem is obvious. However, with more subtle errors, such as phase speed errors associated with finite differencing, the problem might be less easy to detect. If variational analyses were to be used to diagnose the propagation of smaller scale waves,

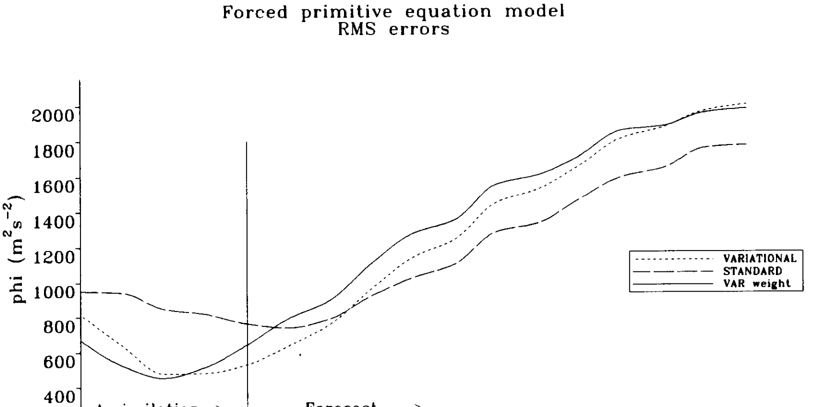

Fig. 6. RMS geopotential error in $m^2 s^{-2}$ as a function of time for standard (dashed), variational (dotted) and variational assimilation with temporal weighting of observations (full). A 50% phase speed error was introduced for all Rossby waves. A 24-h assimilation was followed by a 3-day forecast. $\alpha = 10^{-3}$.

the true phase speeds would be underestimated. The standard scheme would give better results, at least in areas with good data coverage.

The problem with model errors in variational assimilation is particularly severe in the tropics, where forecast errors are much larger than in mid-latitudes (Heckley, 1985). Systematic errors contribute significantly to the total error. Also, through the predominance of stationary components, the variational approch meets with principal limitations, as there is less tendency information to extract in the tropics.

6.3. *Temporal weighting of observations*

An obvious way to avoid the unsatisfactory fit at the end of the variational assimilation interval is a temporal weighting of the observations (Courtier and Talagrand, 1990). Usually, one wants to start a forecast from the last analysis. Therefore, one wants the best fit at the end of the assimilation period. An experiment was run with the observations at the beginning and the end of the interval weighted in the ratio 1 to 100. The sum of the weights was kept to 1 so not to change the relative importance of the smoothing term in the cost function. As in the previous experiment, the Rossby waves travel westward with only half their true phase speed. Fig. 6 shows the RMS difference of the initialized analysis to the truth for the geopotential as a function of time. A 24-h assimilation was followed by a 3-day forecast. Clearly, at the end of the assimilation the variational scheme with temporal weighting (full) has the best fit to the truth. At the beginning it is worse than the scheme without weighting (dotted). As pointed out before, this scheme shows the best fit in the middle of the interval. The standard scheme (dashed) has the worst fit of all. However, during the forecast things change. The standard scheme initially has a further decrease in error through the observations brought in at 24 h. The variable weighting forecast has a somewhat larger error growth rate than the no-weighting scheme. At the end of the forecast the advantage of temporal weighting is lost. Note that the problem becomes only apparent in the later stages of the forecast. This might explain why it did not show up in Courtier and Talagrand (1990), who limited their forecasts to 24 h. Also, it might be masked by other forecast errors in their experiments.

The present result can easily be explained by inspection of the Rossby mode spectrum at the beginning of the forecast (Fig. 7). Apart from wave-numbers 3, 5 and 6 all amplitudes of the variational assimilation without weighting are closer to the truth than the weighting scheme. As the Rossby modes are dispersive, the different spectra result in different error growth rates in

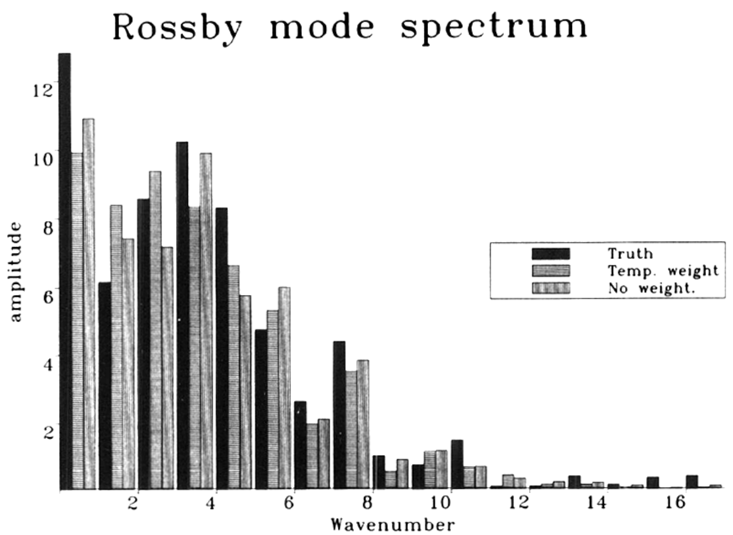

Fig. 7. Rossby mode spectrum at the end of the assimilation interval for truth (full), variational (vertical shading) and temporal weighting variational (horizontal shading) assimilation. Units are arbitrary. $\alpha = 10^{-3}$.

the forecast. The reported close RMS fit, which is just the sum of the squared differences of the various Fourier amplitudes, does not necessarily guarantee a good forecast.

The weighting scheme chosen here is extremely simple. More sophisticated algorithms would probably give better results. In any case, the proper definition of the cost function will require considerable further work in order to extract the maximum amount of information from the observations. This will bring about additional complications to the initially simple concept of variational assimilation. The success of the schemes will always have to be judged by the fit to the observations *and* by the quality of the ensuing forecasts.

7. Extended variational assimilations

While in the previous experiments the initial state was the only control variable, we will now extend the approach by including the forcing parameters into the optimization process. We will generate the observations from a run with a specified initial state and with specified forcing terms. Then we will ask the variational scheme to recover both the initial state and the correct forcing parameters from the observations. This is basically a combination of the variational continuous assimilation (Derber, 1989), which seeks to optimize a correction to the model equations and of the familiar variational assimilation, which seeks to optimize the initial state.

We define a modified state vector z' which does not only hold the predictive variables ψ, χ, ϕ, but also the forcing terms Q.

$$\boldsymbol{Z}' = [\psi, \chi, \Phi, Q_\zeta, Q_{\mathrm{D}}, Q_\Phi]^{\mathrm{T}} \tag{26}$$

In order to project into normal mode space we write:

$$\boldsymbol{Z}' = \underline{\underline{S}}'\boldsymbol{Y}', \tag{27}$$

where $\underline{\underline{S}}'$ is a 6×6 block diagonal matrix with block elements $\underline{\underline{S}}$. In analogy to the derivation in Subsection 2.2, the model equations in normal mode form are given by:

$$\boldsymbol{Y} = \underline{\underline{D}}'\boldsymbol{Y}'(0). \tag{28}$$

$\underline{\underline{D}}'$ is a 3×6 matrix made up of $\underline{\underline{D}}$ and 3 extra columns with elements

$$d'_{3,3+j} = \frac{i}{\Lambda_j}(1 - \mathrm{e}^{i\Lambda_j t}). \tag{29}$$

The predictive variables z at any particular time can now be expressed as function of the initial state and of the forcing parameters through

$$\boldsymbol{Z} = \underline{\underline{S}}\,\underline{\underline{D}}'\underline{\underline{S}}'^{-1}\boldsymbol{Z}'(0). \tag{30}$$

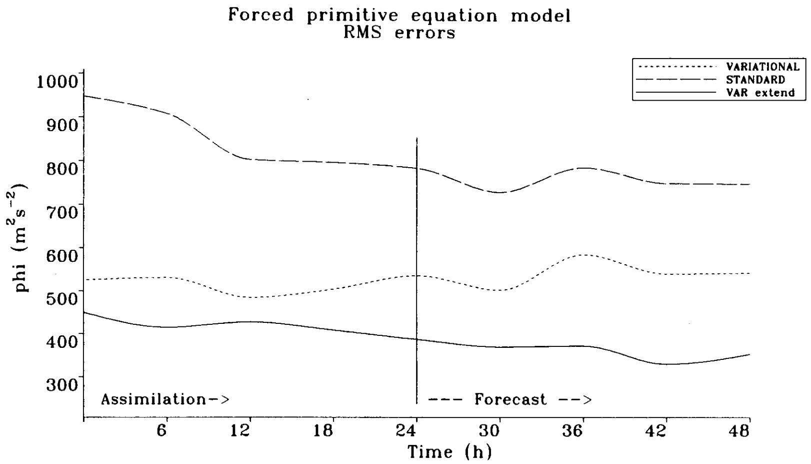

Fig. 8. RMS geopotential error in $\mathrm{m^2\,s^{-2}}$ as a function of time for standard (dashed), variational (dotted) and extended variational (full) assimilation. The latter recovers both the initial state and the tuning parameters. A 24-h assimilation was followed by a 1-day forecast. $\alpha = 10^{-3}$.

All the remaining equations can be derived with (30) accordingly.

7.1. Variational tuning

In a preliminary series of experiments (not shown) only the forcing terms were recovered from the observations. These experiments can be regarded as variational tuning, where the tunable parameters of the model are the control variables. Even with noisy observations, the forcing terms were recovered to an acceptable degree of accuracy.

Fig. 8 shows the RMS errors as a function of forecast time for a run where the initial state and the forcing terms were recovered from the data. The observations were generated with Q_ζ and Q_D set to zero and $Q_\phi = 3.4\ m^2\ s^{-3}$. Initialization was switched on so that only the Rossby mode projections were extracted from the observations. Clearly, the extended variational scheme has the best fit to the truth during the assimilation. This advantage is kept during the forecast which also benefits from the recovered forcing terms. As the forcing terms have been reconstructed, the assimilation and the forecast are done with the correct model, so that most of the results from Subsection 6.1 apply. Because of the missing forcing, the standard variational scheme (dashed) can only fit the observations to a lesser degree. It is, however, considerably better than the intermittent assimilation (dashed), which would need more time for a better fit.

The fact that the extended variational assimilation can in principle serve to determine the tunable parameters of a model together with the initial state has implications for at least two areas of

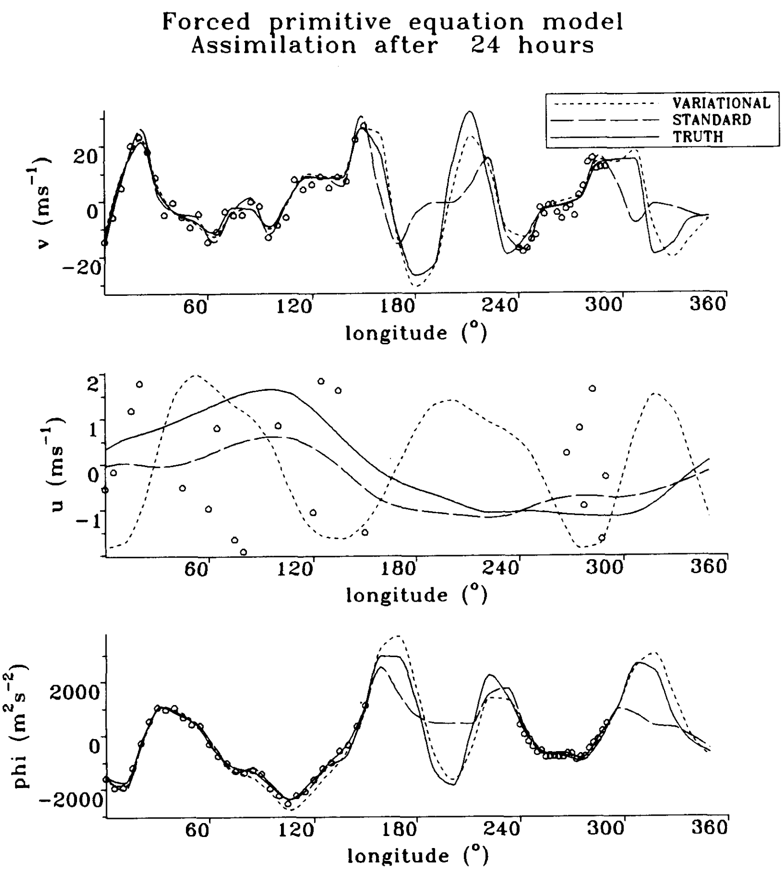

Fig. 9. Same as Fig. 4, but variational assimilation with the model used as weak constraint. $\alpha = 10^{-4}$, $\gamma = 10^{-3}$.

current research in numerical weather prediction. The first concerns parameterization studies. They are usually done on a one-by-one trial and error basis. This means that the setting of a certain tunable parameter is based on a limited number of experiments in which only the parameter in question is varied. There is often little consideration of the interaction between different parameterized processes. For instance, a retuning of the boundary layer scheme might have consequences for the convection scheme through a change in the low-level convergence patterns. The variational approach allows in principle to account for all possible interactions. If successful, it would yield the optimal specification of all the tunable parameters in one go. Application of this method in different flow regimes would give some indication of the reliability of the results. Of course, the schemes have to be applied with great care, as they might yield "unphysical" values which could lead to the right results but for the wrong reasons.

The second area of potential benefit of the extended variational assimilation is the long standing problem of separating forecast errors into analysis- and model errors. Studies in this area are mostly based on statistical concepts (Arpe et al., 1985) relying on various assumptions and approximations. Observation quality permitting, the proposed method could in principle contribute to a solution of this problem, as it relies on the evolution of the error pattern in space and time. Only if model- and analysis errors lead to the same spatial and temporal evolution of the error fields will the method yield ambiguous results.

7.2. *Assimilation with weak constraint*

All the experiments in Section 6 were done with the model assumed to be correct. There has been no account for possible model errors. This is the strong constraint approach (Sasaki, 1970). We will now discuss how the extended variational assimilation can also be interpreted as assimilation with weak constraint. Following Sasaki (1970), we do not strictly enforce the model equations, but require that they are obeyed to only up to an arbitrary "misfit" term, which we also subject to the minimization. Thus, we interpret the term Q_ϕ in (3) as misfit term to be minimized in the assimilation. For simplicity, the modification is limited to the mass field equation. The corresponding definition of the cost function is given by:

$$J_3 = J_2 + \gamma \sum_m Q_{\Phi m} Q^*_{\Phi m} \qquad (31)$$

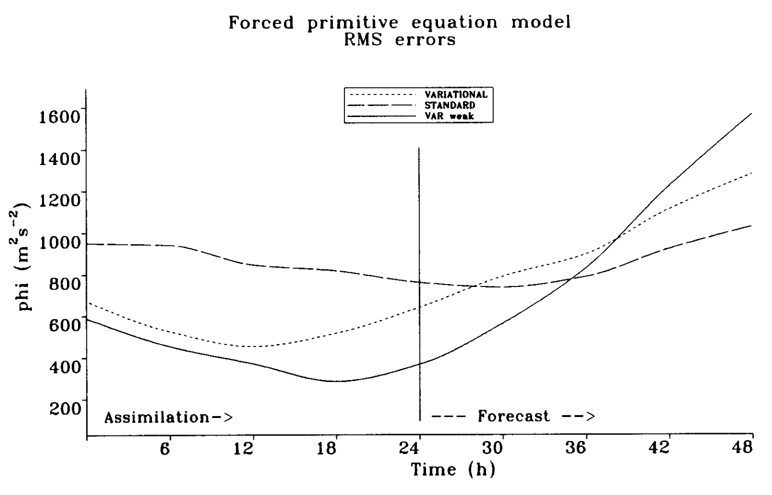

Fig. 10. RMS geopotential error in $m^2 s^{-2}$ as a function of time for standard (dashed), variational (dotted) and variational assimilation with weak constraint (full). A 50% phase speed error for the Rossby modes was introduced in the model. One-day assimilation followed by one-day forecast. $\alpha = 10^{-4}$, $\gamma = 10^{-3}$.

where γ is an arbitrary weight to be determined experimentally. This means that we can apply the formalism from Subsection 7.1 together with an additional loading of the corresponding diagonal terms in matrix $\underline{\underline{V}}$.

Fig. 9 shows the analysed fields after 24 h of assimilation with γ set to 10^{-3}. As in Section 6, a 50% phase speed reduction was introduced for the Rossby modes. The weak constraint assimilation (dotted) has largely managed to compensate for this error. The geopotential curves can be directly compared to Fig. 4, the experiment with strong constraint clearly showing the phase lag of the variational solution. This problem is no longer apparent in Fig. 9, which shows a good fit to the truth in the rotational wind and the geopotential. The fit in the divergent wind is less satisfactory. From this, a weak constraint assimilation would seem to alleviate some of the problems associated with erroneous models in variational assimilation.

There is, however, a snag. It concerns the quality of the subsequent forecasts. Fig. 10 shows the RMS geopotential errors as a function of time. Again, the superiority of the weak constraint scheme (full) is confirmed for the assimilation. But things change dramatically during the forecast. The errors grow much faster than in the variational assimilation with strong constraint (dotted) and in the standard intermittent scheme (dashed). After 24 h, the standard scheme has the smallest errors.

This result is a typical example of getting right answers for the wrong reasons. As long as observations are available, the weak constraint assimilation manages to keep the error level low. It does so at the cost of a completely erroneous spectrum of the Rossby modes together with large amplitudes of the misfit terms. In fact, the fit is surprisingly good during the assimilation if one remembers that the misfit terms have to make up for a phase speed error. From (12) it is clear that this is bound to fail. It is precisely here, where the fundamental problem with the weak constraint approach lies. Every assumption about the functional form of the misfit term implies a certain modification of the solution. In lack of any knowledge about the required corrections, one might easily obtain unwanted results. In any case, it is essential to assess a particular analysis not only by the fit to the data but also by the quality of the ensuing forecasts.

8. Discussion

At first sight, the results of the present study might seem puzzling. After all, at the theoretical level it can be shown that four-dimensional variational assimilation (4D-Var) is equivalent to Kalman-Bucy filtering (KBF) (Ghil and Malanotte-Rizzoli, 1991). The standard assimilation as applied in the present paper, on the other hand, is only a suboptimal version of KBF, in which the true forecast error covariance matrices explicitly calculated by KBF are replaced by much simpler prespecified formulations. How can this simple scheme nonetheless give better results in the case of an erroneous model?

There are number of reasons for this apparent discrepancy. First and foremost, 4D-Var is only equivalent to KBF if the model is not strictly enforced but only used as additional constraint in the definition of the cost function. Secondly, statistical information on the model- and observation errors consistent with the formulations in KBF have to be incorporated into 4D-Var. How to do this and how best to solve the resulting equations is still an open research problem (Ghil and Malanotte-Rizzoli, 1991). The present paper shows that it is a very important issue.

Most of the recent applications implicitly assume the model to be correct (e.g., Courtier and Talagrand, 1990). It is only through this assumption that the constrained variational problem for the state vector over the entire assimilation period can be reduced to an unconstrained problem for the lower dimensioned state vector at the initial time. This so-called "reduction of the control variable" (Le Dimet and Talagrand, 1986) renders variational assimilation computationally affordable. Of course, statistical informations on forecast errors are irrelevant in this method. As shown in Subsection 6.2, this approach can lead to serious problems if the model has errors.

Relaxing the perfect model assumption is possible by introducing extra "misfit" terms in the equations and subjecting them to the minimization together with the deviation from the observations. This is a simple form of the weak constraint formalism (Sasaki, 1970). Misfit terms act as extra forcing in the equations making sure that the model trajectory stays closer to the observations than in the strong constraint approach. The control variable would thus be augmented by the extra forcing terms; essentially doubling the computational costs. As discussed in Section 7, this approach can improve the variational solution for the assimilation interval but may have detremental effects on the ensuing forecasts. This will always happen, if the extra forcing terms are unable to compensate for the real model errors.

As opposed to 4D-Var, the standard OI-based scheme as applied in the present paper does use statistical information on the forecast- and observation errors without interfering with the dynamics of the model. As shown in Section 6, incorporating this knowledge into the analysis process is essential for the performance of the schemes. Apparently, the crude assumptions about these errors as currently used in most OI applications are preferable to the perfect model assumption as used in 4D-Var, if the model has errors. KBF theory could be used as a guide line for further cost-effective refinements of the statistics. The links between 4D-Var and KBF could open up new avenues in further research on variational assimilation. Simple models as the one used in the present study could play an important role in this research.

It must be stressed, however, that the present study focusses only on one aspect of variational assimilation, namely on the incorporation of the time domain into the analysis process. It could well be that the other advantages of the method, e.g., use of observed quantities nonlinearly related to the models variables or other nonlinear aspects as discussed in Lorenc (1988) would justify an operational implementation even if the above reservations still hold.

REFERENCES

Arpe, K., Hollingsworth, A., Tracton, M. S., Lorenc, A. C., Uppala, S. and Kallberg, P. 1985. The response of numerical weather prediction systems to FGGE level III b data. Part II: Forecast verification and implications for predictability. *Q. J. Meteorol. Soc. 111*, 67–101.

Bengtsson, L. 1975. *4-dimensional assimilation of meteorological observations.* GARP Publication Series No. 15, World Meteorological Organization, Geneva, Switzerland, 75 pp.

Courtier, P. and Talagrand, O. 1987. Variational assimilation of meteorological observations with the adjoint vorticity equation (II). Numerical results. *Q. J. R. Meteorol. Soc. 113*, 1329–1347.

Courtier, P. and Talagrand, O. 1990. Variational assimilation of meteorological observations with the direct and adjoint shallow-water equations. *Tellus 42A*, 531–549.

Derber, J. C. 1989. A variational continuous assimilation technique. *Mon. Wea. Rev. 117*, 2437–2446.

Ghil, M. and Malanotte-Rizzoli, P. 1991. Data assimilation in meteorology and oceanography. *Adv. Geophys. 33*, 141–266.

Heckley, W. H. 1985. Systematic errors of the ECMWF operational forecasting model in tropical regions. *Q. J. R. Meteorol. Soc. 111*, 709–738.

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

Lorenc, A. C. 1981. A global three-dimensional multivariate statistical interpolation scheme. *Mon. Wea. Rev. 109*, 701–721.

Lorenc, A. C. 1986. Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc. 112*, 1177–1194.

Lorenc, A. C. 1988. Optimal nonlinear objective analysis. *Q. J. R. Meteorol. Soc. 114*, 205–240.

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

Sasaki, Y. 1970. Some basic formalisms in numerical variational analysis. *Mon. Wea. Rev. 98*, 875–883.