KONINKLIJK NEDERLANDS METEOROLOGISCH INSTITUUT

WETENSCHAPPELIJK RAPPORT

SCIENTIFIC REPORT

W. R. 85 - 4

P. Lynch

Forecast updating:
theory and application to a simple model

De Bilt, 1985

Publikatienummer : K.N.M.I. W.R. 85-4 (DM)

Koninklijk Nederlands Meteorologisch Instituut
Dynamische Meteorologie
Postbus 201
3730 AE DE BILT

The Netherlands

U.D.C. : 551.509.313

ISSN : 0169-1651

FORECAST UPDATING

Theory and Application to a Simple Model

by

Peter Lynch *

April, 1985

* on leave from the Irish Meteorological Service

Abstract

The World and Regional Meteorological Centers are currently producing and disseminating high-quality forecasts for several days ahead. It is becoming increasingly difficult for the smaller National Centers to produce competitive forecasts in the traditional way. Therefore, it seems worthwhile to seek alternative methods which use the available data and forecasts in a complementary rather than competitive way. One such approach is described in this report.

The basic idea of Updating is to see where the early forecast is going wrong, and to amend later forecasts accordingly. Suppose we have a series of forecasts starting at time t_0. Using later data, at time t_1, we can calculate the error at that time. Then, if we know how the error evolves, we can estimate it at a later time t_2, and use it to amend or update the t_2-forecast.

If a forecast is analysed into its normal modes, the error dynamics of each component are relatively simple. By making reasonable assumptions about the behaviour of nonlinearities, it is possible to estimate the errors, and to amend each component separately. The components are then resynthesized to give an updated forecast.

A long-standing problem with all forecasting models is their rapid initial error growth-rate. It will be shown that in certain circumstances the errors in updated forecasts grow more slowly than the errors in normal forecasts starting from the same data time. Since updating is computationally undemanding, the technique can provide short-range forecasts which are both more accurate and more economical.

The updating procedure is applied in the context of a simple one-dimensional model. It is found that updated forecasts using data valid

6 or 12 hours after the initial time are more accurate than corresponding forecasts starting at these times, right out to 48 hours after the initial time. Their small initial error growth-rates are in accordance with theory.

The application of updating in a general context is considered, and difficulties which can be foreseen are discussed. More work is required before the operational feasibility of the technique can be assessed.

Acknowledgements

It is a pleasure to thank Joop Bijlsma for helpful discussions, Bronno de Haan for producing the graphs, Fons Baede for reading the manuscript and Birgit Kok for typing the report.

CONTENTS

I INTRODUCTION

I.1 Motivation

Predicting the future course of the weather from its present state has been the central goal of synoptic meteorology since its beginnings. Vilhelm Bjerknes first recognized, early this century, that the physical principles governing the atmosphere could be used to deduce its evolution from a known initial state. The hydrodynamic and thermodynamic equations expressing these physical laws form a determinate system which allows us, in principle, to derive a solution from known initial conditions.

The initial value approach is at the heart of present-day numerical weather prediction (NWP). The equations of motion are integrated forward in time from specified initial conditions to predict the future state of the atmosphere. Such integrations are now carried out routinely at a number of forecasting centres. The larger national and international centres (ECMWF, UKMO, NMC, etc.) are currently producing and disseminating high-quality forecasts for several days ahead.

The production of numerical forecasts is extremely expensive in computational terms. The large centres have computers capable of hundreds of million operations per second. Computational resources in the smaller national meteorological centres (NMC's) are much more modest. Since forecast accuracy is strongly dependent upon computer power, and since the large centres frequently upgrade their computational facilities, it is becoming increasingly more difficult for the smaller national services to produce competitive forecasts in the traditional way (i.e. by solving an initial-value problem using an NWP-model).

There are several excellent reasons why national services should continue to engage in numerical prediction using Limited Area Models (LAM's); these are discussed, for example, by Janjić (1984). Nevertheless, the increasing availability of high-quality products from the World- and Regional-Centres is a reality which provides us with cogent and pressing arguments for reviewing our forecasting methods. Specifically, we should seek, as alternatives to the initial-value approach, methods which use the available observations and forecasts less extravagantly and with greater effect. One such alternative is discussed in this report.

I.2 Methodology

The forecasts available internationally are of high quality; nevertheless, they do contain errors. Observations which are made later than the initial time of a forecast contain new information which has not been used in preparing the forecast. The forecast and later observations are available concurrently in the national weather services. It seems that it should be possible to combine these two types of information in such a way as to produce a better forecast. This can be done either subjectively (by a human forecaster) or objectively (by a computer); we will discuss each method in turn.

(a) Holistic Amendment of Forecasts

Consider how the forecaster amends an old forecast in the light of new data. He compares today's analysis with yesterday's forecast for today. He makes a 'holistic analysis' of the weather chart into 'components' such as low pressure systems, fronts, jet streams, etc. He assesses the error of each component: low centre misplaced or not deep enough; cold front moved too slowly; wave overdeveloped; jet stream too slack. With this knowledge he can re-draw tomorrow's forecast, making 'holistic amendments': move the front further on; deepen the low; tighten the jet. He does not think purely in dynamical terms, but more in terms of patterns or isobaric geometry, and mostly in two dimensions. Nevertheless, he can make significant improvements to a forecast by this process of kinematic extrapolation, provided he has access to more up-to-date data.

The forecaster uses the following tools: (1) His synoptic experience, which embodies knowledge of atmospheric statistics, together with details of many specific case studies; (2) Various rules-of-thumb, and mechanistic models, e.g. the Norwegian model of a frontal depression; (3) Personal intuition (no two forecasters will produce identical prognoses). The laws of atmospheric dynamics enter only indirectly - there are no 'holistic equations of motion' which describe the dynamics of a low or a front as a single component or entity.

(b) Automation of the Amendment Procedure

It seems worthwhile to try to simulate the activities of the forecaster in an automatic procedure. Since the computer operates quantitatively, can extrapolate more accurately and can assimilate more observational information, it should be possible to improve upon the subjective approach, and to 'beat the forecaster at his own game'.

The computer must operate in a 'reductionistic' way, since it cannot recognize patterns such as highs, fronts, etc. (at least, not with the data stored in conventional spectral or gridpoint form). But the reduction of the data to components of some sort (normal modes, EOF's, gridpoint values, etc.) allows us to amend each component separately. The amendment may be based on dynamical, statistical or empirical rules goverining the errors, and all the new observational data can be utilized. Thus, we may hope that the automatic amendment may be superior to the subjective one.

First let us consider a direct simulation of the human approach. Given the gridpoint values, we can calculate the error field at a particular data-time, t_1. Individual 'patches of error' may then be assumed to advect along some trajectory with an appropriately chosen velocity, until a later time, t_2. The resulting error field can then be used to amend the t_2-forecast. A system of this sort is under development by Roodenburg (1984). The advantages of the approach are its simplicity and its local nature: we can concentrate on a specific region of interest. The main drawback is that we do not have quantitative information about the behaviour of the errors: the equations governing the errors at a gridpoint are more complicated than the original forecast equations! Therefore, we can do little more than advect the errors passively with an assumed 'steering' flow.

An alternative approach, considered in detail in this report, is to analyse the meteorological fields into components which behave in a simple and predictable way, and to amend each component separately. The <u>normal modes</u> of a system are the basic components of its linear dynamics, and are governed by (formally) simple equations. The structure of these equations suggests appropriate approximations which make it possible to derive a relation for the error in each component. These relations can be used to amend the individual components. The residual error can be examined quantitatively; in the case of linear dynamics the method yields exact results. Allowance is made for both amplitude- and phase-errors. The method is discussed in more detail in the following section. A schematic comparison between the subjective and automatic amendment techniques is illustrated in Box 1.

I.3 Forecast Updating: Normal Mode Approach

The basic idea of Updating is to see where the early forecast is going wrong, and to amend later forecasts accordingly. Suppose we have a series of forecasts starting at time t_0. Using later data, at time t_1, we can calculate the error at that time. Then, if we know how the error evolves, we can estimate it at a later time t_2, and use it to amend or update the t_2-forecast. (We will refer to the three times as the initial time, t_0, the data time, t_1, and the update time, t_2).

We consider a simple case first: a system consisting of a single component which is governed by a linear equation. The error in the model simulation of this system is also governed by a simple linear equation. If we know the error at two times, say t_0 and t_1, it can be calculated at any other time. Thus, we can correct or update the model solution exactly at any later time, t_2.

We next consider a single-component system, governed by a nonlinear equation. The error dynamics are also nonlinear. However, by making a hypothesis relating the true and model nonlinearities we can derive a relation for the error, which is formally similar to that in the linear

Method	subjective/holistic	objective/reductionistic
Elements	Mental analysis of chart (2-dimensional) into components (Lows, fronts, jets, etc.)	Spectral analysis of atmosphere (3-dimensional) into eigenmodes (Hough-functions, EOFs; etc)
Correction	Amendment of each component (Movement of a front; deepening of a low; etc)	Amendment of each component (e.g. correction of the phase and amplitude of a mode)
Guiding Principles	(1) Experience, (2) Rules-of-thumb (3) Intuition	(1) Dynamical, (2) Statistical, (3) Empirical rules governing errors.
Critique	Very little dynamical content. There are no 'holistic equations of motion' which describe the dynamics of a low or front as a single entity. Intuition is unreliable.	Quantitative method with sound theoretical basis. The components are governed by known equations. Analysis of residual errors is possible.

Box 1 Comparison between subjective and automatic methods of amending forecasts.

case. This allows us to update later forecasts. Because of the approximations made, the updated forecasts are not perfect, but under certain circumstances their errors grow more slowly than those of a normal forecast starting at the data time, t_1. In these conditions it is better to update the old forecast (from t_0) than to do a fresh forecast (from t_1). It is also computationally much cheaper.

Numerical forecasting models have many degrees of freedom. We cannot expect that a particular value (e.g. the surface pressure at one gridpoint) is governed even approximately by a simple linear equation. However, the linearized model equations have simple solutions called normal modes. A general solution can be analyzed into its normal mode components. We can apply the updating procedure, outlined above, to each individual component, and then re-synthesize an updated forecast from them.

I.4 Overview of this Report

The general theory of updating is developed in Chapter II, and the behaviour of the residual errors is examined. A simple one-dimensional model, DYNAMO, is described in Chapter III, and its normal modes are derived. The application of the updating technique is formulated. The results of this application are presented in Chapter IV. Finally, Chapter V, comprises a review, and suggestions for future work.

Readers who are satisfied with the general description of updating given above, and who are not concerned with the mathematical technicalities, may wish to skip now to the results described in Chapter IV.

II THEORY OF UPDATING

II.1 A Simple Example

Consider a one-dimensional linear advection equation governing non-dispersive propagation on a periodic domain:

$$\frac{\partial \phi}{\partial t} + c\,\frac{\partial \phi}{\partial x} = 0, \quad \phi(0,t) = \phi(L,t). \tag{1}$$

Since c is constant the spatial dependence can be spectrally analysed, and the wavenumber, k, is quantized by the periodicity. The coefficient $X = X(t)$ of a single component, exp (ikx), is governed by the equation

$$\dot{X} + i\lambda X = 0, \tag{2}$$

where $\lambda = \lambda(k) = kc$ is the frequency. If X_o is the initial value of X, the solution of (2) is

$$X = X_o \exp(-i\lambda t). \tag{3}$$

Now suppose that the numerical counterpart of (1) involves a phase-error, such that wavenumber k is advected at a speed $c' = c'(k)$. The corresponding coefficient of the model solution, X', will be governed by the equation

$$\dot{X}' + i\,\lambda'\,X' = 0 \tag{4}$$

where $\lambda' = kc'$, so the solution with initial value X'_o may be written

$$X' = X'_o \exp(-i\lambda' t). \tag{5}$$

We define the <u>error ratio</u> E by

$$E(t) = X(t)/X'(t). \tag{6}$$

Using equations (2) and (4) it is easily shown that E is governed by the equation

$$\dot{E} = \frac{X}{X'}\left(\frac{\dot{X}}{X} - \frac{\dot{X}'}{X'}\right) = -i(\lambda-\lambda')\,E \tag{7}$$

so that the solution with initial value E_o is

$$E = E_o \exp[-i(\lambda-\lambda')t]. \tag{8}$$

Suppose now that we are given the initial values X_o and X'_o and the values X_1 and X'_1, at time t_1, of the true and model solution. If the model solution at t_1 is amended by the change

$$X'_1 \rightarrow E_1 X'_1$$

we retrieve the true solution. Given the model solution at some later time, t_2, how can we amend it so as to obtain to true solution? From (8) it is clear that

$$E(t) = E_o(E_1/E_o)^{t/t_1}. \tag{9}$$

This allows us to amend the component at any time. In this simple

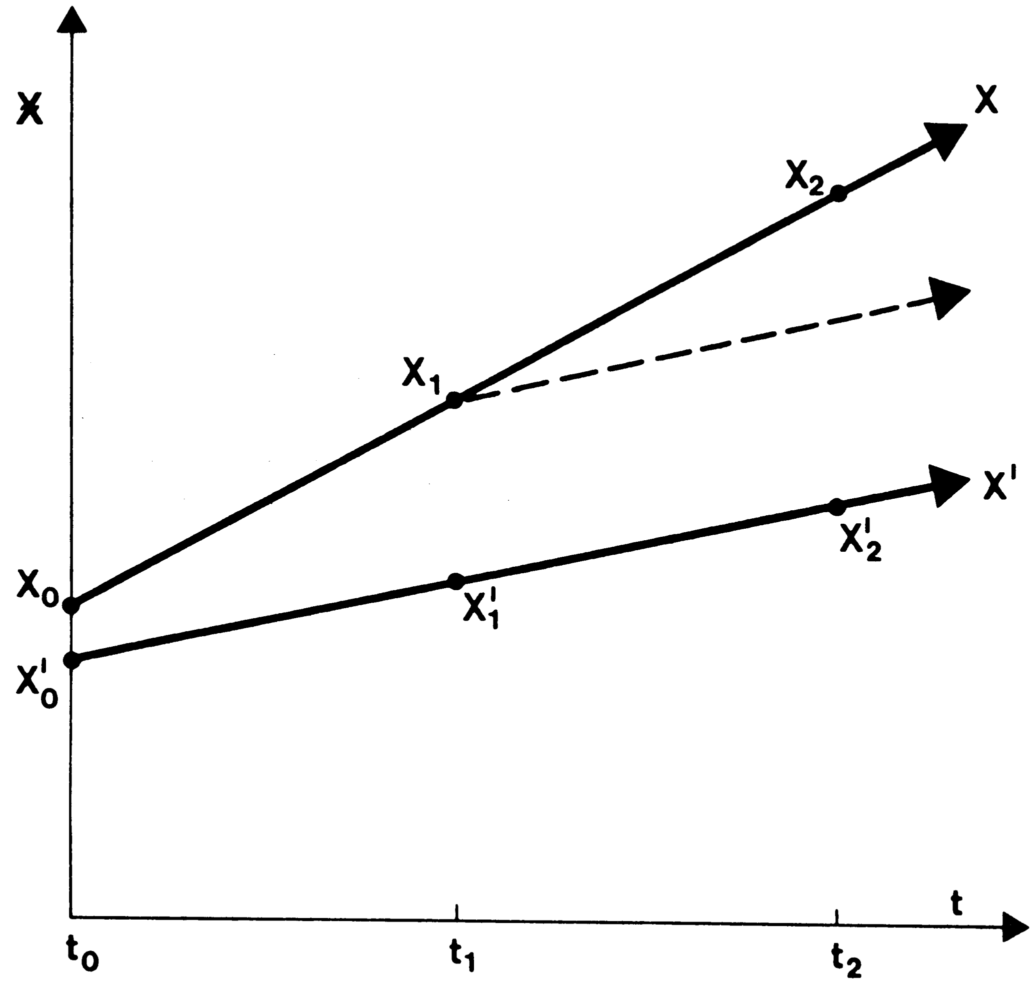

Figure 1 Schematic depiction of the true solution, X, and the model solution, X', governed by the linear equations (2) and (4).

linear case we retrieve the exact solution X. Note that we have not used the values of λ and λ' explicitly in (9).

In figure 1 the state of the system is represented schematically by a point in a (here 1-dimensional) state-space, $\mathfrak{X}$, which moves linearly in time. We assume that X_o, X'_o, X_1, X'_1 and X'_2 are known. The problem is to estimate X_2. We write

$$E_o = X_o/X'_o;\ E_1 = X_1/X'_1;\ E_2 = E_o(E_1/E_o)^{t_2/t_1}$$

and amend the model value at t_2 by

$$X'_2 \longrightarrow E_2\, X'_2.$$

This turns out to be the exact solution.

If each component of the model solution is amended as above, we retrieve the true solution of (1). Note from figure 1 that the amended solution is <u>better</u> than the model solution which starts from the correct value at t_1 (the dashed line in figure 1). The phase of each component is corrected. The systematic error inherent in the model has been overcome, and its dispersive nature has been counteracted.

II.2 More General Systems

We can represent the state of the atmosphere by a vector $\underset{\sim}{X}(t)$ whose evolution is governed by a nonlinear vector equation

$$\dot{\underset{\sim}{X}} + \underset{\sim}{L}\underset{\sim}{X} + \underset{\sim}{N}(\underset{\sim}{X}) = 0, \tag{10}$$

where $\underset{\sim}{L}$ is a constant linear operator and $\underset{\sim}{N}$ is a nonlinear vector function. If the eigenvectors of $\underset{\sim}{L}$ are known, the system can be diagonalized, and a single component, X, will be governed by an equation of the form

$$\dot{X} + i\lambda X + N(\underset{\sim}{X}) = 0. \tag{11}$$

We write the numerical counterpart of (11) as

$$\dot{X}' + i\lambda' X' + N'(\underset{\sim}{X}') = 0, \tag{12}$$

where X' is the component of the numerical solution. There is an error $(\lambda'-\lambda)$ in the frequency and also an error in the nonlinear term.

Suppose we consider the linear case, where $N \equiv N' \equiv 0$. The components X and X' of the true and model solutions then behave as wave-like normal modes with well-defined phase-speeds, governed by equations like (2) and (4). Thus, the error ratio is given by equation (9), and has an exponential behaviour in time (this would not be so, for example, if X, X' were grid-point values). So, in the linear case, each component can be corrected just as for the simple example in section II.1.

Consider now the nonlinear case: we write equations (11) and (12) in the form

$$\dot{X} + i\mu(t)\, X = 0, \quad \mu \equiv \lambda + N(\underset{\sim}{X})/iX \tag{13}$$

$$\dot{X}' + i\mu'(t) X' = 0, \quad \mu' \equiv \lambda' + N'(\underset{\sim}{X}')/iX'. \tag{14}$$

We define mean values for the quantities μ and μ':

$$\overline{\mu} = \frac{1}{t}\int_0^t \mu(\tau)d\tau; \quad \overline{\mu'} = \frac{1}{t}\int_0^t \mu'(\tau)d\tau.$$

Then, the solutions of equations (13) and (14) can be written

$$X(t) = X_o \exp(-i\bar{\mu}t); \quad X'(t) = X'_o \exp(-i\bar{\mu}'t)$$

where X_o, X'_o are the initial values. (In fact, these are not explicit solutions, but nonlinear integral equations for X and X', since the integrals $\bar{\mu}$ and $\bar{\mu}'$ involve the unknown quantities. However, they may be treated formally as though they were explicit solutions.)

If we now define

$$\mu_n = \bar{\mu}(t_n); \quad \mu'_n = \bar{\mu}'(t_n)$$

for n = 1 and 2, the true and model solutions are

$$\begin{aligned} X_n &\equiv X(t_n) = X_o \, e^{-i\mu_n t_n} \\ X'_n &\equiv X'(t_n) = X'_o \, e^{-i\mu'_n t_n} \end{aligned} \qquad (15)$$

From these we see immediately that

$$X_2 = X_o \left(\frac{X_1}{X_o}\right)^{(\mu_2 t_2/\mu_1 t_1)} \qquad (16)$$

$$X'_2 = X'_o \left(\frac{X'_1}{X'_o}\right)^{(\mu'_2 t_2/\mu'_1 t_1)} \qquad (17)$$

To procede, we make a hypothesis about the relative changes in $\bar{\mu}$ and $\bar{\mu}'$ with time: it is reasonable to assume that

$$\frac{\mu_2}{\mu_1} \doteq \frac{\mu_2'}{\mu_1'} \quad \text{or equivalently} \quad \frac{\mu_2}{\mu_2'} \doteq \frac{\mu_1}{\mu'_1} \tag{18}$$

Using this to eliminate (μ_2/μ_1) in (16), and dividing by (17) we get

$$E_2 \equiv \left(\frac{X_2}{X_2'}\right) \doteq E_o\left(\frac{E_1}{E_o}\right)^{(\mu'_2 t_2/\mu'_1 t_1)} \equiv \check{E}_2 \tag{19}$$

Now from (15) we know that

$$X_1'/X_o' = e^{-i\mu_1' t_1}; \quad X_2'/X_o' = e^{-i\mu_2' t_2}$$

from which it follows immediately that the exponent in (19) is

$$\rho \equiv \left(\frac{\mu_2' t_2}{\mu_1' t_1}\right) = \frac{\log(X_2'/X_o')}{\log(X_1'/X_o')} \; . \tag{20}$$

The equations (19) and (20) allow us to calculate the estimate $\check{E}_2$ of E_2 from the quantities $(X_o, X_1; X_o', X_1', X_2')$. We can then estimate the true value X_2 at t_2 by the amendment

$$X_2' \rightarrow \check{E}_2 X_2' \equiv \check{X}_2 \tag{21}$$

This will be exact ($\check{X}_2 = X_2$) if (18) holds exactly. Allowance has been made for the nonlinear terms, and also for their non-constancy in time; the approximation (18) involves essentially an assumption that the mean functions $\overline{\mu}$, $\overline{\mu}'$ change with time in the same way.

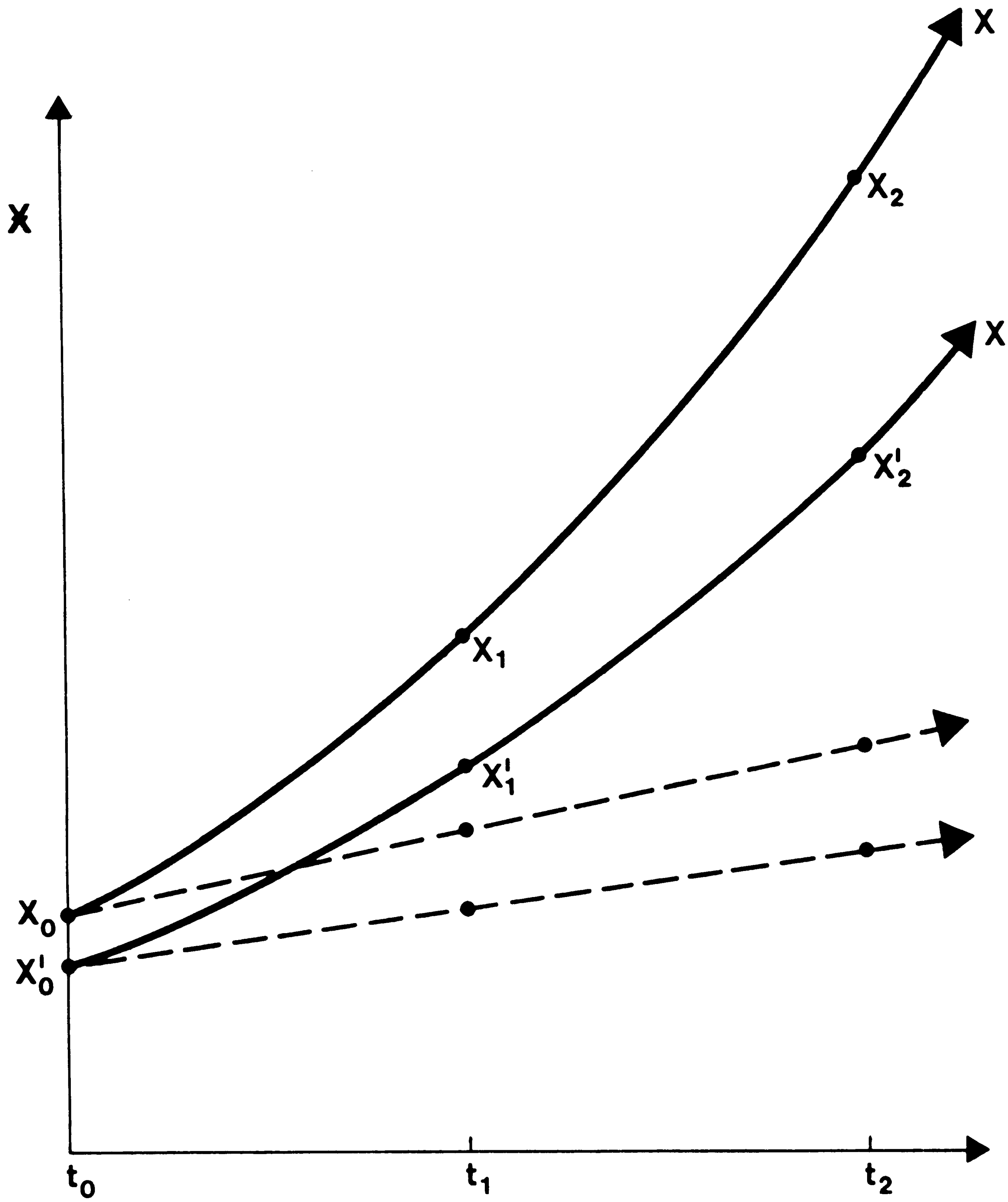

Figure 2 Schematic depiction of the true solution, X, and the model solution, X', governed by the nonlinear equations (11) and (12). (Dashed lines indicate the corresponding linear solutions.)

Note that in the above updating formulae, (19) and (20), no use has been made of the values λ, λ' and no explicit evaluation of nonlinear terms is required.

In figure 2 we represent the true and model solutions schematically by points in a 1-dimensional state-space, $\mathfrak{X}$. The phase-speeds are not constant in time, so that simple extrapolation of errors is insufficient; we must allow for the time-dependent nonlinear terms.

II.3 The Growth of Forecast Errors

A simple way to measure the accuracy of a forecast is to calculate the root-mean-square (RMS) difference between the actual and forecast fields. Usually the error grows quickly at first, and gradually levels off as the forecast progresses. We will try to explain why this is so, and how the updating process can lead to a slower initial error growth.

(a) Errors in the original forecast

Consider a single wave, forecast by a model with an inherent phase-error. We have seen in section 2 that the error-ratio is

$$E = \exp[-i\ \kappa\ t]$$

where $\kappa = (\lambda-\lambda')$ is the frequency error, and the initial error is zero ($E_0 = 1$). Thus, the phase-error is κt; the true and model waves move into and out of phase with a period $(2\pi/\kappa)$. The RMS error, σ, is easily calculated, and is proportional to $|\sin \frac{1}{2}\kappa t|$ (see figure 3a). In particular, note that the initial growth of the error is linear in time; we can show that

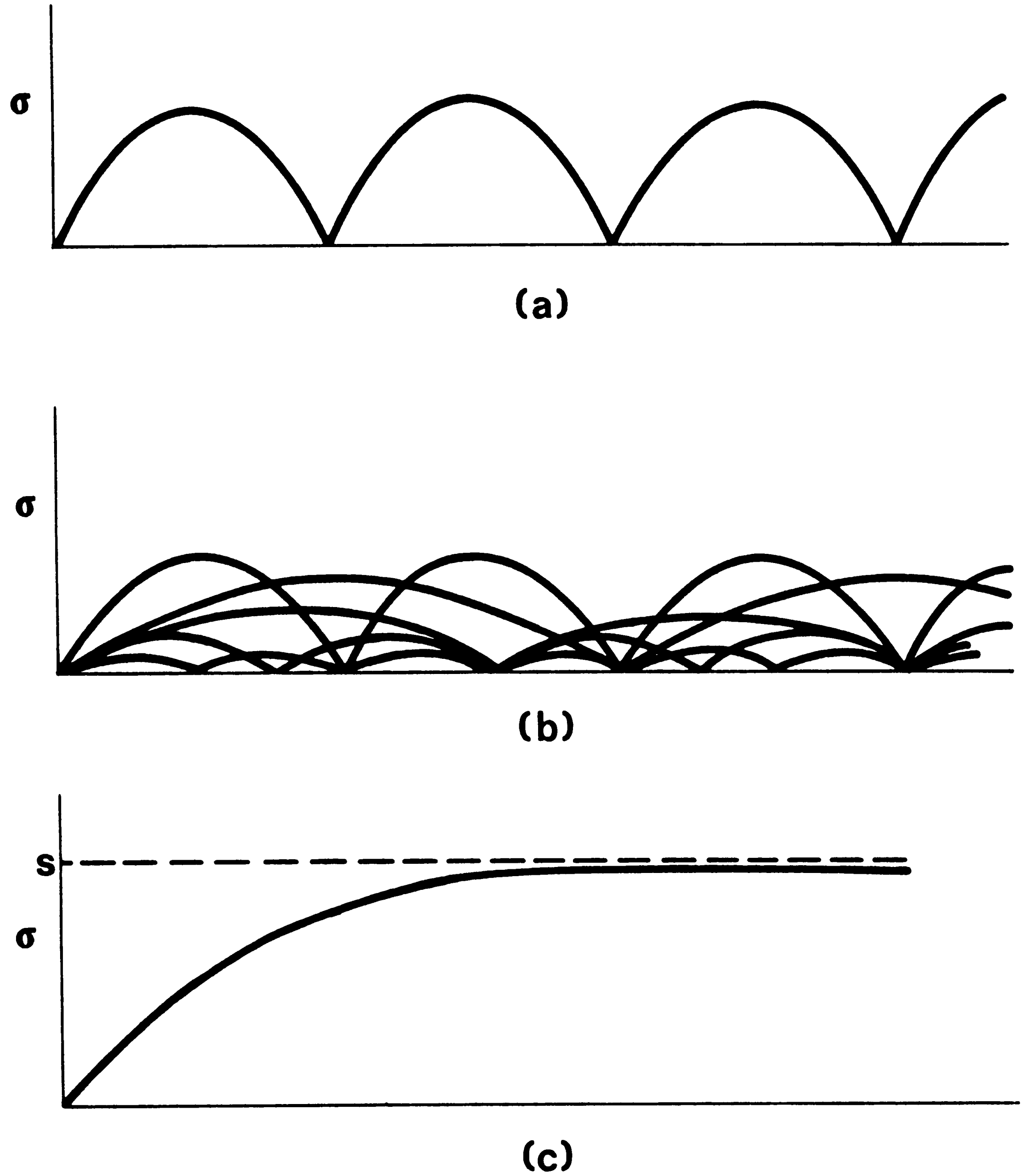

Figure 3 Behaviour of the RMS error in time for (a) a single component, (b) Many components with different frequencies and phase-errors, and (c) the nett resulting error for a many-component system. (Note that the variances (σ^2) are additive componentwise.)

$$\sigma \doteq \left(\frac{X_o}{\sqrt{2}}\right) \kappa t, \text{ for small } t \qquad (22)$$

where X_o is the amplitude of the wave.

In general there are many wave components present, each with a different phase error (Figure 3b). The initial error will still be approximately linear, dominated by the components for which σ in equation (22) is largest. After a time the components will tend alternately to cancel and reinforce each other in a more-or-less random manner (the variances σ^2 are additive componentwise). Therefore, we would expect the error σ to level off at some saturation level, and not to change much when t is large (see figure 3c). The shape of the σ-curve may be approximated by a function of the form

$$\sigma \doteq S(1-e^{-\alpha t})$$

where S is the saturation level and $(S\alpha)$ is the initial slope.

The discussion in the previous paragraph was naīve in that non-linear interactions between wave components were ignored. In reality these interactions result in the transfer of errors between different scales, and may result in more rapid growth of initial errors. A glance ahead to figure 5(b) on page 30 confirms this: the dashed line is the error for a linear model run and the solid line is the corresponding error for a non-linear run. The latter has much more rapid error growth rate. On the other hand, Boer (1984) has examined the errors of the Canadian Meteorological Centre operational forecast system, and has separated the consequences of initial and model errors. In Figure 4 (redrawn from Boer's paper) we show the total error at various times, and the estimated contribution due to the initial error (ϵ_o), the model error (shaded), and the non-linear production term. Clearly, the model error dominates the growth of error in the early stages. At later times the nonlinear production term becomes more important. The best way to reduce the error in short-range forecasts is to reduce the model error; reducing the initial error ϵ_o alone would not have much effect. The updating technique is designed primarily to counteract the effects of model error; the error remaining after updating is discussed next.

(b) Errors in the Updated Forecast

We write the true and model solutions in the form

$$X = X_o \exp(-i\mu(t).(t-t_o)) \ ; \ X' = X'_o \exp(-i\mu'(t).(t-t_o)) \quad (23)$$

(where the overbars on μ and μ' have been dropped for notational simplicity). The original error ratio is

$$E = E_o \exp\left(-i[\mu(t)-\mu'(t)](t-t_o)\right) \quad (24)$$

So, the initial phase-error grows linearly with time. The basic approximation in the updating technique is equation (18), which we write here as

$$\mu/\mu' \doteq \mu_1/\mu'_1. \quad (25)$$

The exact error ratio (24) is estimated using (25), resulting in equation (19) which we write here as

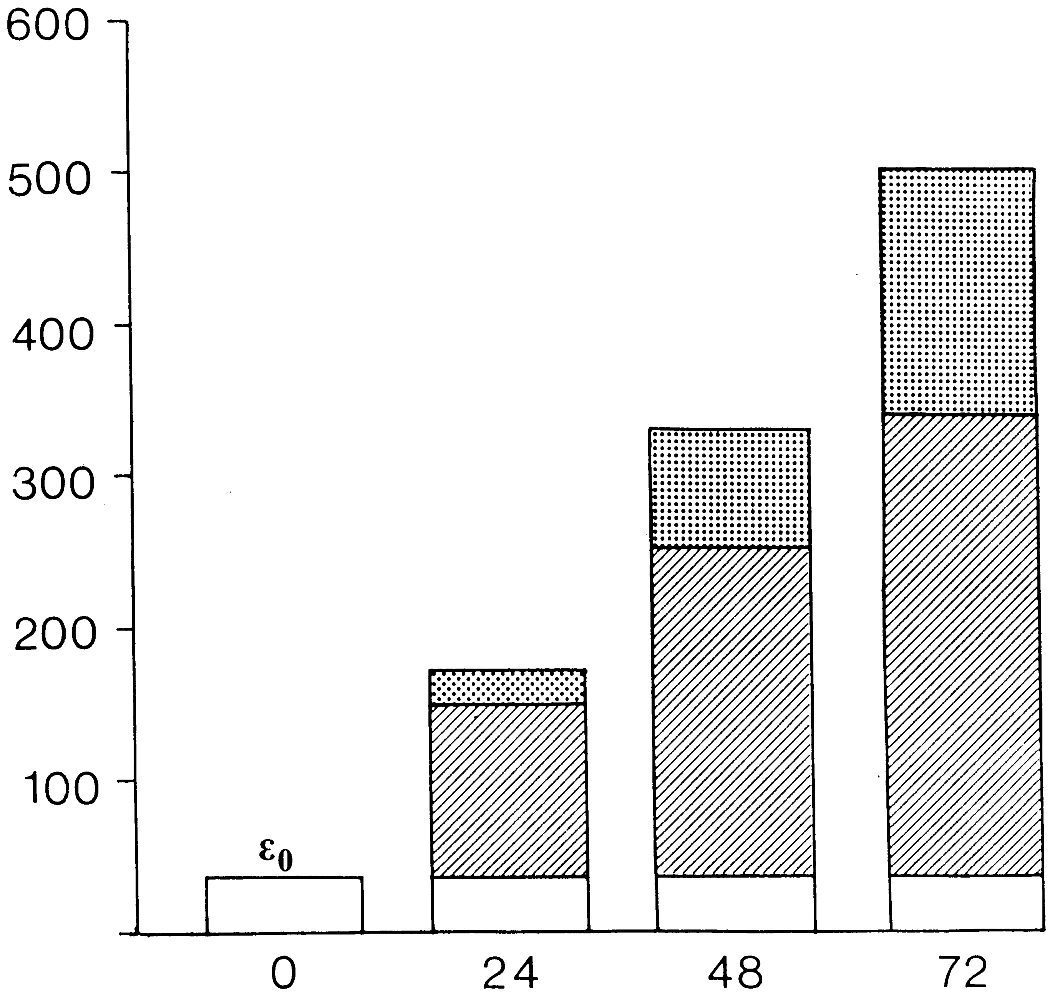

Figure 4 Total error (10^3 J m^{-2}) of the Canadian Meteorological Centre forecasting model, and contributions due to the initial error (ϵ_o), the source term or model error (shaded), and the non-linear production term (stippled). (After Boer, 1984).

$$\check{E} = E_o \exp\left(-i[(\frac{\mu_1}{\mu_1'})\mu' - \mu'](t-t_o)\right), \qquad (26)$$

and the updated value is defined by $\check{X} = \check{E}\, X'$.

We can now estimate the residual error, i.e. that error which remains after the updating process has been performed. We have

$$\varepsilon \equiv X/\check{X} = EX'/\check{E}X' = E/\check{E},$$

and using (24) and (26) this is

$$\varepsilon = \exp\left(i[(\frac{\mu_1}{\mu_1'})\mu' - \mu](t-t_o)\right) \qquad (27)$$

Clearly, if (25) is exactly true, E is equal to unity, i.e. no error remains.

Suppose now that $(t-t_1)$ is small; in words, the update time is close to the data time. We can expand $\mu(t)$ in a Taylor series about t_1:

$$\mu(t) = \mu_1 + (t-t_1)\frac{d\mu_1}{dt} + O(t-t_1)^2,$$

and similarly for $\mu'(t)$. Then the exponent in (27) becomes

$$[(\frac{\mu}{\mu'})\frac{d\mu'}{dt} - \frac{d\mu}{dt}]_{t_1} \cdot (t-t_1).(t-t_o) \qquad (28)$$

where squares and higher terms in $(t-t_1)$ have been dropped.

The initial phase-error in the updated forecast, $\check{X}$, is proportional to $(t-t_1)(t-t_o)$. When (t_1-t_o) is small, this is approximately quadratic in time, in sharp contrast to the linear error growth of the original model error (compare equations (24) and (28)). Thus, we may hope that the updated forecast value, $\check{X}$, is not only better than the original one, X', but even better than a fresh forecast starting at t_1, at least during the initial hours (when $(t-t_1)$ is small).

When the data-time t_1 is much later than the initial time we have initial error growth at t_1 proportional to

$$(t-t_1)(t-t_o) \doteq (t-t_1)(t_1-t_o)$$

and since (t_1-t_o) is large, the error will grow rapidly. Thus, updating from a late data-time is unlikely to be useful.

The theoretical behaviour of the errors outlined above will be seen to be in accordance with the actual results described in chapter IV.

III THE ONE-DIMENSIONAL MODEL

In order to test the usefulness of the idea of updating forecasts, we apply it in the simplest context: two forecasts are made using a one-dimensional model, with two different grid-resolutions. The fine-grid forecast is regarded as representing the 'true atmosphere', and is used as a reference against which the coarse grid or 'model' forecast is measured.

In order to apply the method of updating we need to know the normal modes of the model. These are described in section III.1. The effects of discretization are also discussed. The formulation of the updating is developed in section III.2.

III.1 The model DYNAMO; Normal Modes

The model used in this study is a simple one-dimensional shallow-water model. It is described in detail in Lynch (1984). The momentum equations are differentiated to form vorticity and divergence equations: this makes the β-effect explicit, and all further latitude dependence is supressed.

The basic equations, which are nonlinear, may be written

$$\zeta_t + (u\zeta)_x + f\delta + \beta v = 0 \qquad (29)$$

$$\delta_t + (u\delta)_x - f\zeta + \beta u + \phi_{xx} = 0 \qquad (30)$$

$$\phi_t + (u\phi)_x + \phi\delta \qquad = 0 \qquad (31)$$

(all notation is conventional - see Lynch (1984) for details). The vorticity and divergence are related to the winds by

$$\zeta = v_x \quad ; \quad \delta = u_x .$$

To investigate the simple types of wave-motion supported by the above system the equations are linearized about a state of rest and the perturbation quantities are assumed to be harmonic in x and t:

$$\begin{bmatrix} \zeta' \\ \delta' \\ \Phi' \end{bmatrix} = \begin{bmatrix} \hat{\zeta} \\ \hat{\delta} \\ \hat{\Phi} \end{bmatrix} \cdot \exp[ik(x-ct)]$$

Then (29), (30), and (31) become three homogeneous equations for the amplitudes $(\hat{\zeta}, \hat{\delta}, \hat{\phi})$. The condition for a non-trivial solution is that the system determinant should vanish. This gives a cubic equation for the phase-speed:

$$c(c+\beta/k^2)^2 - c(\overline{\phi}+(f/k)^2) - (\beta/k^2)\overline{\phi} = 0 \qquad (32)$$

The three roots may be estimated by making simple assumptions about the magnitude of the phase-speed. These assumptions can then be justified *a posteriori*.

If c is small the cubic term in (32) is neglected, giving

$$c = c_R = -(\beta/k^2)/[1+f^2/k^2\overline{\phi}]$$

This is the Rossby wave phase-speed. Equations (29) and (30) then tell us that this solution is in approximate geostrophic balance for v and that u is much smaller than v, i.e. the wave is quasi-nondivergent. The Rossby waves always travel westward relative to the mean flow.

If we assume that $|c| \gg |c_R|$ the constant term in (32) is negligible and we get the two roots:

$$c = \pm \sqrt{(\bar{\phi}+f^2/k^2)}$$

which are the phase-speeds of the gravity-inertia waves. The gravity waves travel in both directions with relatively large phase-speeds. They are divergent motions and typically have fairly small vorticity.

There is an eigenvector, $\underset{\sim}{Z}_k^\tau = (\delta_k^\tau,\ \zeta_k^\tau,\ \phi_k^\tau)^T$, associated with each eigenvalue $c = c_k^\tau$, where τ indexes the root of (32) and k indexes the wavenumber. Its components are related by:

$$\delta_k^\tau = (ikc/\bar{\phi})\phi_k^\tau \ ; \ \zeta_k^\tau = (fc/\bar{\phi}(c+\beta/k^2))\phi_k^\tau. \qquad (33)$$

An inner product is defined as follows:

$$\underset{\sim}{Z}_1 \cdot \underset{\sim}{Z}_2 = \int_o^L [\bar{\phi}(\overset{*}{\delta}_1 \delta_2 + \overset{*}{\zeta}_1 \zeta_2) + \frac{\partial \overset{*}{\phi}_1}{\partial x} \cdot \frac{\partial \phi_2}{\partial x}]\ dx$$

where asterisks denote complex conjugates and the integral ranges over a periodic channel [0, L]. Then the eigenfunctions $\underset{\sim}{Z}_k^\tau$ can be shown to be orthogonal:

$$\underset{\sim}{Z}_k^\tau \cdot \underset{\sim}{Z}_{k'}^{\tau'} = \delta_{\tau'}^\tau\ \delta_{k'}^k \left\| \underset{\sim}{Z}_k^\tau \right\|^2.$$

We can expand a general state, $\underset{\sim}{Z} = (\delta, \zeta, \phi)^T$, in the eigenmodes

$$\underset{\sim}{Z} = \sum_k \sum_\tau a_k^\tau \underset{\sim}{Z}_k^\tau$$

and the coefficients are given by the generalized Fourier coefficients

$$a_k^\tau = (\underset{\sim}{Z} \cdot Z_k^\tau)^* / \left\| Z_k^\tau \right\|^2. \tag{34}$$

In the application of the model, (32) is solved numerically for the eigenvalues, and the eigenfunctions are constructed using (33) and stored on disk.

We discretize the periodic channel as follows

$$\{x_o = 0,\ x_1 = \Delta x, \ldots,\ x_n = n\Delta x, \ldots,\ x_N = N\Delta x = L\}.$$

Then ζ, δ and ϕ are specified on this grid, and u and v at points half way between these gridpoints. A function $q(x_n)$ has N degrees of freedom (note that $q_o = q_N$ by periodicity) and may be expanded in a Fourier series

$$q(x_n) = \sum_{m=-M}^{M} q_m \exp(2\pi i m x_n / L)$$

where $M = [N/2]$. If q is real, as in the present case, then $q_{-m} = q_m^*$ and only half the coefficients are needed.

The finite difference scheme which approximates the equations (29), (30) and (31) is described in Lynch (1984). The differencing and averaging operations involve discretization errors. If we repeat the normal mode analysis as in the continuous case, but take account of the spatial discretization, we find the results are identical provided only that we make the substitutions

$$k \rightarrow k' = \sin(\tfrac{1}{2}k\Delta x)/(\tfrac{1}{2}\Delta x)$$

$$\beta \rightarrow \beta_k = \cos(\tfrac{1}{2}k\Delta x).\beta \qquad (35)$$

Then the eigenvalues are obtained from (32) as before, and the eigenvectors from (33).

It was found that the use of the continuous eigenmodes (those for $\Delta x = 0$) evaluated at the gridpoints led to results very similar to those obtained after using (35). Which set to use appears to be a matter of expediency.

III.2 Formulation of the Updating technique

The model DYNAMO was used to make two parallel series of forecasts. The forecast domain was a channel of length $L = 10^4$ km. The fine grid has 50 gridpoints with $\Delta x = 200$ km, and the coarse grid has 25 points with $\Delta x = 400$ km. Forecast duration was 48 hours. An Adams-Bashforth timestepping scheme was used; to avoid errors associated with time-discretization, a very short timestep ($\Delta t = 15$ sec.) was used for all runs.

A number of nondimensional combinations occur in the model; we define length and velocity scales L and V and form

$$R_o = (V/fL) : R_\beta = (\beta L/f); R_F = \overline{\phi}/(fL)^2.$$

The choice of parameters is such that they take the values

$$R_o = 10^{-1}; R_\beta = 1.6 \times 10^{-1}; R_F = 10.$$

The initial conditions were defined by setting

$$\phi_n^o = \phi^o(x_n) = \sum_{k=1}^{10} a_k \cos\{(2\pi k.n\Delta x/L) + \varphi_k\} \qquad (36)$$

where the amplitude a_k is determined by requiring that the fields have a specified spectral density and the phase φ_k is chosen randomly. The results presented below are for $a_k = a_o k^{-5/6}$, giving a spectrum of kinetic energy-per-unit wavenumber proportional to $k^{-2/3}$. Other choices were $a_k = a_o k^{-4/3}$ and $a_k = a_o k^{-2}$ giving the minus five-thirds and minus three spectra more typical of atmosphere flows (see e.g. Tennekes, 1978).

Geostrophic winds were derived from

$$v_{n-\frac{1}{2}}^o = (\phi_n^o - \phi_{n-1}^o)/\Delta x; \; u_{n-\frac{1}{2}}^o = 0.$$

The fields were then initialized using the Laplace transform technique (Lynch, 1985). This led to slightly different fields on the two grids.

The initial fields and the forecast values after each hour were stored on disk, to serve as input files to the updating program. The root-mean-square (RMS) error in the coarse grid geopotential, vorticity and divergence fields were calculated each hour (by error we mean the difference between the two forecasts, with the fine-grid values regarded as the truth). Plots of the errors in ϕ and ζ are shown in Figure 5. The solid curves are the errors for the full nonlinear forecasts and the dashed lines show the scores when the nonlinear terms are switched off in both cases. The initial error in ϕ is due to the initialization acting differently on the two grids; since rotational modes are unchanged by initialization, there is little difference between the initial vorticity fields. The divergence errors are somewhat noisy, but very small relative to the vorticity errors, and will not be considered. The errors in the linear runs grow relatively slowly, and at a more-or-less constant rate; the nonlinear errors have a greater initial growth rate, and reach saturation level somewhere between one and two days.

The updating was done in the following way: the initial time t_o and "data-time" t_1 were chosen, and the fine-grid and coarse-grid fields valid at these times were spectrally analysed into components using the model normal modes described in section III.1. The initial error in the m^{th} component is defined by

$$E_o^m = X_o^m / X_o'^m$$

where X^m represents the fine-grid value and X'^m the coarse grid coefficient. The error at time t_1, the data-time, is

$$E_1^m = X_1^m / X_1'^m.$$

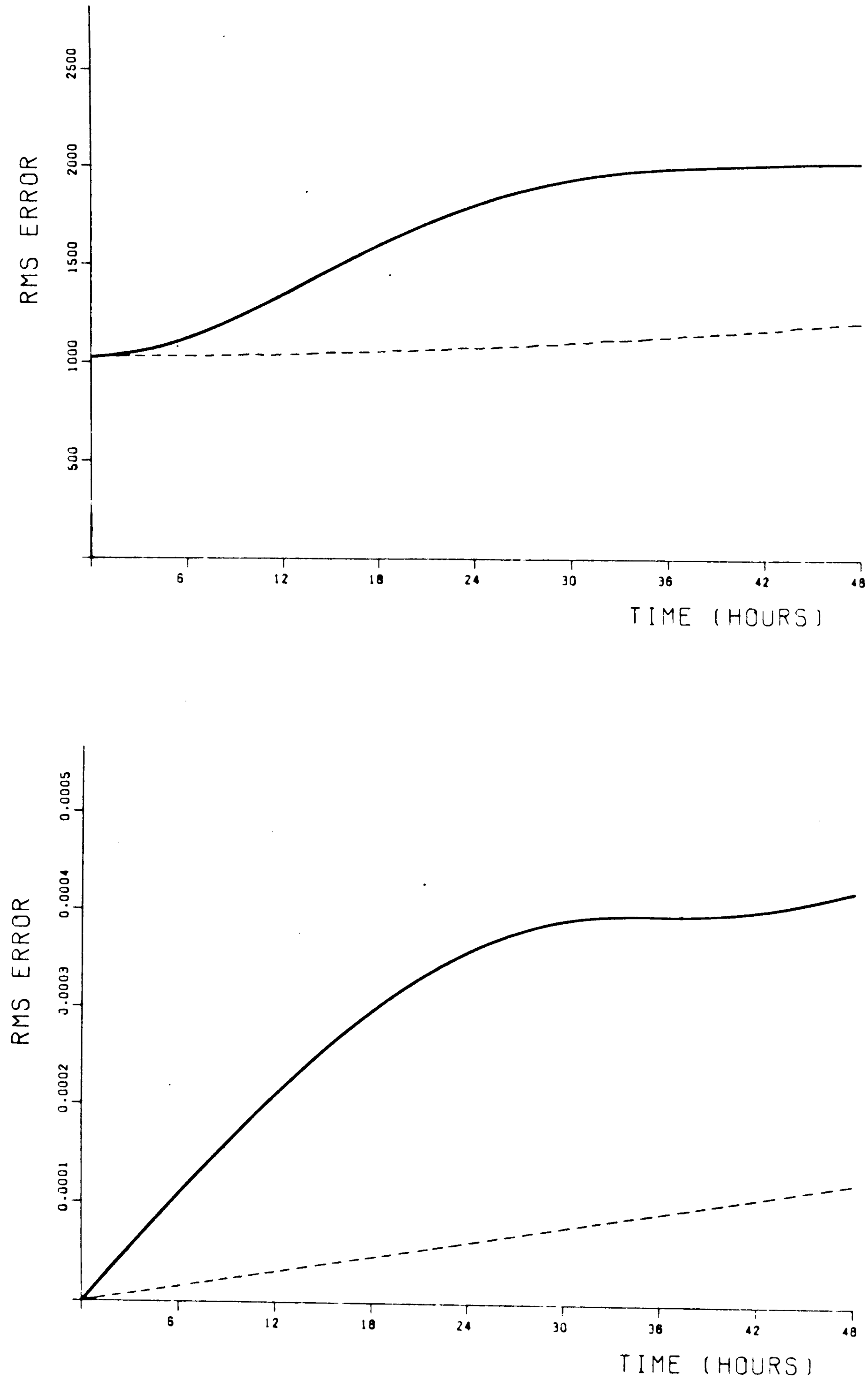

Figure 5 RMS errors in (a) the geopotential and (b) the vorticity fields, for the coarse-grid forecast (measured against the fine-grid forecast). Solid curves show errors for the full nonlinear run; dashed curves are for a linear run.

We wish to estimate the true, or fine-grid, value at a later time, t_2. To do this, we estimate $E^m_2 = E^m(t=t_2)$ by the method described in section II.2. The exponent ρ_m is calculated using only coarse-grid values at the initial time, data-time and update time:

$$\rho_m = \log(X'^m_2/X'^m_o)/\log(X'^m_1/X'^m_o)$$

and equation (19) is then applied as follows

$$\check{E}^m_2 = E^m_o(E^m_1/E^m_o)^{\rho_m} . \qquad (37)$$

The updated value is then given by the modification

$$X'^m_2 \rightarrow \check{E}^m_2 \, X'^m_2 \equiv \check{X}^m_2 . \qquad (38)$$

The actual values used in this procedure are contained in the set $\{X^m_o, X'^m_o; X^m_1, X'^m_1; X'^m_2\}$. Note that, despite the nonlinear nature of the equations, no explicit evaluation of non-linear terms is required; we do not even need to know the form of these terms.

The updated fields are now constructed by synthesizing the updated components:

$$\underset{\sim}{\check{X}}_2 = \sum_m \check{X}^m_2 \, \underset{\sim}{Z}_m \qquad (39)$$

where $\underset{\sim}{Z}_m$ are the eigenfunctions of the system. In the linear case $\check{X}_2$ will be equal to X_2. The approximations made in the nonlinear case mean that some error still remains. However, as we saw, the growth of this error may be of a different character to that of a normal forecast error.

The updating procedure is applied for a series of 'update-times' t_2 in the range (t_1, t_{max}), and the error of $\check{X}_2$ relative to X_2 is calculated. This is compared to the original error, and also to the error in a hypothetical forecast starting from the data-time t_1. The latter is assumed to be identical in form to the former, but translated in time by $(t_1 - t_o)$.

IV RESULTS

We consider the case of initial data with a minus two-thirds kinetic energy density spectrum (so that the amplitudes of the components of the geopotential fall off as $a_k = a_o k^{-5/6}$). The initial fields are balanced by the Laplace Transform initialization technique (Lynch, 1985). This causes differences to arise between the initial fields on the two grids; these differences are clearly seen in the error of the geopotential, but hardly affect the vorticity field (see Figure 5). Since they are of peripheral interest in this study, we will concentrate mainly upon the errors in the vorticity.

The data time t_1, is taken to be 12 hours after t_o (t_o and t_1 are the only times at which fine-grid information is used in the update procedure). The coarse grid forecasts for each hour between t_1 = 12 hours and t_{max} = 48 hours are now updated by the method of section III.2, and the scores (or errors relative to the fine-grid values) of the updated forecast are calculated.

We will discuss the root-mean-square (RMS) error, σ, in the vorticity. In Figure 6 the error in the original forecast is shown by curve 1. The error grows rapidly during the early forecast, and levels off between one and two days. The initial error-growth is linear, in accordance with the theoretical discussion in section II.3(a). We assume that a fresh forecast starting from time t_1 would behave in a similar way, as shown by curve 2 (this is curve 1 translated in time by 12 hours). The errors in the updated forecasts for $t_1 < t \leqslant t_{max}$ are shown by curve 3. We see that they grow much more slowly at first (for t close to t_1) and that curve 3 remains below the other curves out to 48 hours. This small initial error-growth is consistent with the theory outlined in section II.3(b), and is of crucial importance from a practical viewpoint. It has the

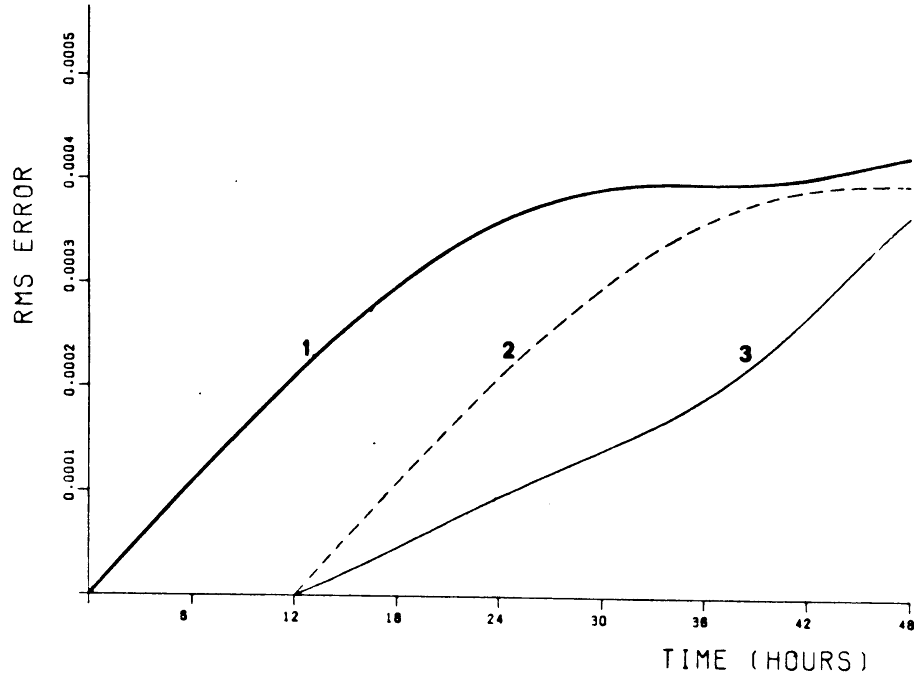

Figure 6 Error curves for an update from t_1 = 12 hours. Curve 1: RMS error in original forecast. Curve 2: Error in forecast beginning at t_1 (this is curve 1 translated by t_1). Curve 3: Error in updated forecasts for $t_1 < t \leqslant t_{max}$ = 48 hours.

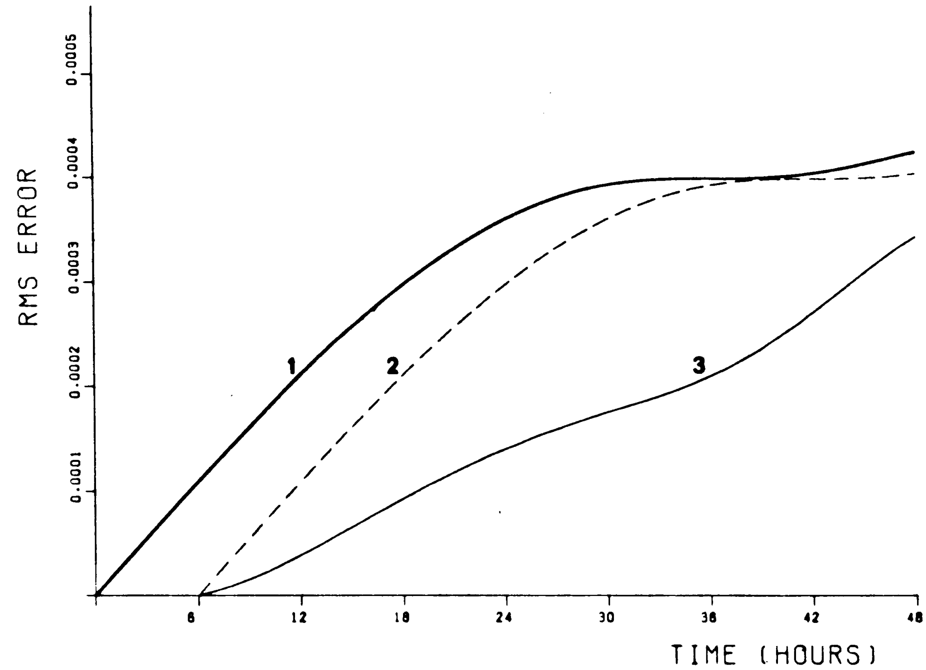

Figure 7 As Figure 6, with t_1 = 6 hours.

following implication: if we are given the original series of forecasts ($X'(t)$), and the new data at 12 hours ($X(t_1)$), then the updated forecast for any time between 12 and 48 hours is more accurate than a normal forecast starting from the data at 12 hours. This is so because the normal forecast error grows rapidly during the initial hours. So, it is better to update the old forecast than to do a fresh one. It is also computationally much cheaper.

When the data time t_1 is chosen to be six hours after t_0, the results are as shown in Figure 7. They are qualitatively similar to those for t_1 = 12 hours. Note, however, that the initial growth of the error in the updated forecasts is even smaller in this case. From the results of section II.3(b) we may argue that the initial growth-rate tends to zero as the data time t_1 approaches the initial time t_0.

In Figures 8 and 9 we show the scores for forecasts updated from data times 24 and 36 hours respectively. The curves are numbered as before. The updated forecasts have faster error growth in these cases. With $t_1 = 24$ hours updating still has an advantage over a fresh forecast out to about 44 hours (after t_0), and improves the original forecast (from t_0) out to almost 48 hours. When $t_1 = 36$ hours, the errors in the updated forecasts grow faster initially than those of a standard forecast starting at that time. Updating appears to be of little use in this case.

The behaviour of the updated forecast errors is seen clearly in Figure 10, where we show the curves for data times every six hours from 6 to 42 hours after t_0. The earlier the data-time, the smaller the initial error growth-rate; thus, we obtain the most advantage by updating using data which is valid as soon as possible after the inital time of the original forecast. There is not much to be gained from further updates using later data.

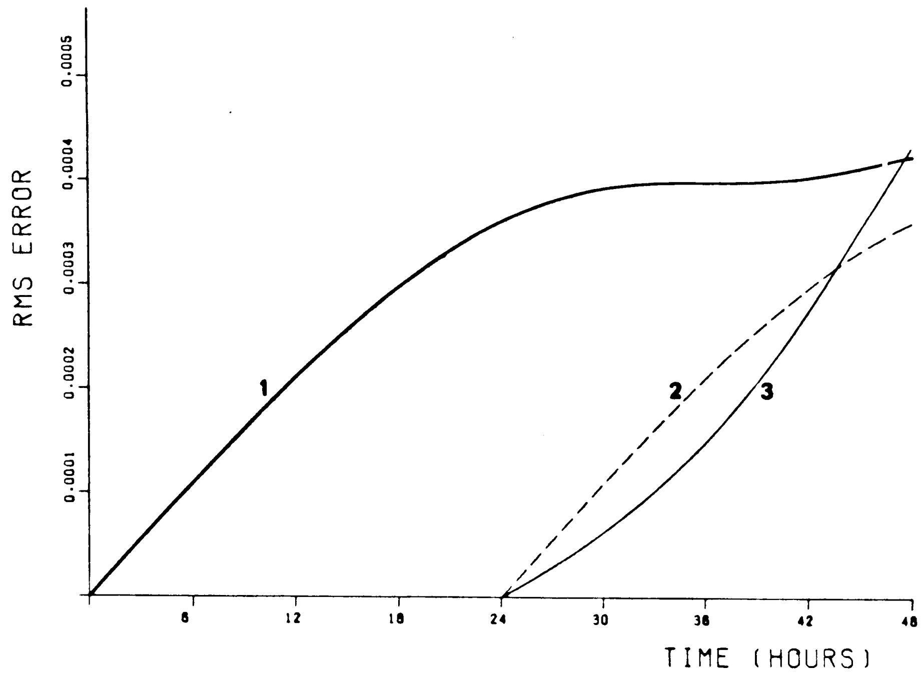

Figure 8 As Figure 6, with t_1 = 24 hours.

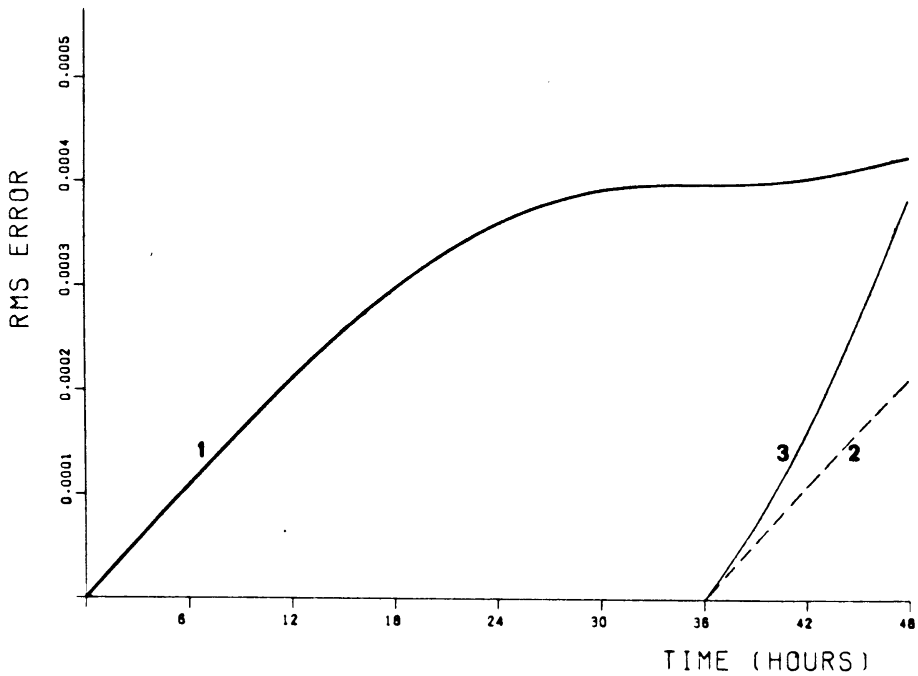

Figure 9 As Figure 6, with t_1 = 36 hours.

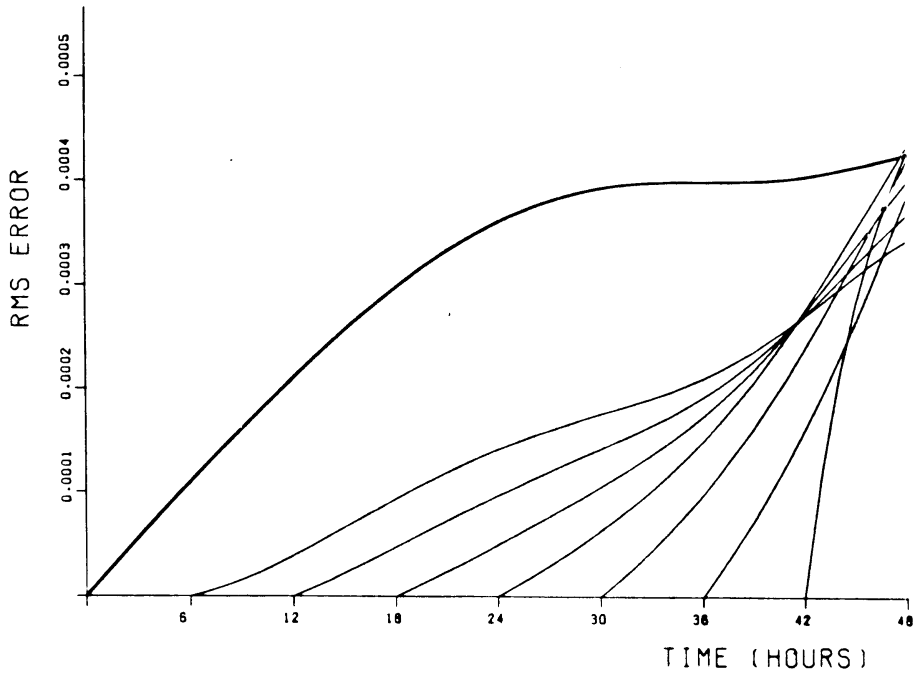

Figure 10 Error curves for the original forecast (heavy) and for the updated forecasts for various data-times (every 6 hours, for $6 \leqslant t_1 \leqslant 42$ hours).

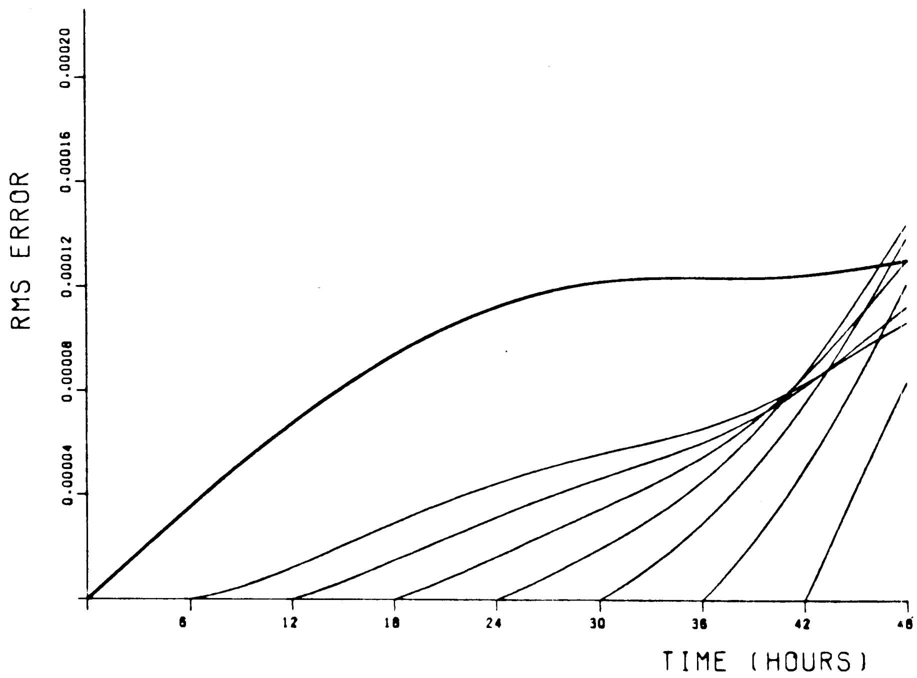

Figure 11 As Figure 10, for initial data with spectral density proportional to -5/3.

To check the representativity of the above results, the numerical experiments were repeated with different initial data. The results for initial fields with a minus five-thirds kinetic energy density spectrum ($a_k = a_o k^{-4/3}$) are shown in Figure 11. Since the smaller scale components are less important in this case, and their nonlinear interactions weaker, the error growth rates are smaller than those in the previous case. However, the general character of the error-curves is the same as before.

V DISCUSSION AND REVIEW

V.1 General Remarks

The changing international scene has a direct bearing on forecasting activities in the National Meteorological Centres. Continuing improvements to the models of the large forecasting centres make it more and more difficult to compete locally. New methods are needed, which use the available data and forecasts in more effective ways.

Locally-run Limited Area Models offer a number of advantages. The integration domain is reduced so that higher resolution may be possible, yielding more refined short-range forecasts in the smaller synoptic scales. Such autonomous systems can be run several times a day, with earlier availability of the output. Still, these systems are expensive to run. Furthermore, they suffer from the same drawback as the global models, viz. rapid initial error growth-rate. This seriously curtails the advantages of more timely forecasts.

Forecast Updating, as described in this report, offers an alternative means of providing short-range forecast guidance. Updating is computationally undemanding, may have better error-growth characteristics, and can remove systematic model errors. Improvements resulting from advances in global forecasting techniques are automatically reflected in the update: perfect input forecasts result in perfect output.

It is a far cry from the results presented in this report to a demonstration of the operational feasibility of updating. Nevertheless, the results are sufficiently encouraging to make a more thorough investigation worthwhile. Some possibilities for future research are suggested below.

V.2 Future Work

The results of the updating experiments described in chapter IV are impressive enough if taken at face value. However, the context in which they were derived is open to criticism; a simple one-dimensional model is a pale shadow of the atmosphere. It cannot be claimed that results for such a case have any direct implications for more complex situations. Therefore, rather than make extravagant claims for updating, we should apply it to the output of a realistic forecasting model.

The European Centre (ECMWF) has been producing medium-range forecasts daily for several years. The global analysis and forecast fields are routinely archived in the form of spherical harmonic coefficients. Since the normal modes of the stream-function are of this form for horizontally non-divergent flow, these coefficients may serve adequately as input to a global one-level updating system. Alternatively, the Hough-mode coefficients may be derived and the updating done in three dimensions. (Since the results in the case of DYNAMO were not unduly sensitive to the precise modes used, nor to the effects of discretization, the simple use of spherical harmonics may well suffice).

In general, forecasts are required for a relatively small area. Medium range forecast models are, of necessity, global; but updating need only be done locally. It is not trivial to adapt the normal mode approach to a limited domain (this difficulty also arises in the case of initialization for LAMs); neither is it obvious how well such a system will function in this context.

The mathematical formulation in section II is not unique. As an alternative, we have treated the nonlinearities as forcing terms, so that they appear in the solution as convolution integrals.

Approximating these integrals in one way or another, we arrive at other estimates of the error ratio. These were found to be inferior to the approach finally adopted. Furthermore, they involved explicit use of the eigenfrequencies, and calculation of the nonlinear terms. Still, we should not abandon the search for other simpler and more accurate formulations of the updating technique.

Finally, we must make a remark about analysis. Most currently operational analysis systems use optimal interpolation; this technique is essentially local in character, and it has been shown (Cats and Wergen, 1982) that such a method is not especially suited to the analysis of large scale normal modes. In an operational updating system it may prove necessary to modify the analysis method so that it provides the required spectral components in an optimal way. Little more can be said without further investigation.

References

(The works listed below were all found to be helpful in the present study, although not all are referred to explicitly in the report.)

Bengtsson, L., (1978): Growth rate and vertical propagation of the initial error in baroclinic models. Tellus, 30, 323-334.

Boer, G.J. (1984): A spectral analysis of predictability and error in an operational forecast system. Mon.Wea.Rev., 112, 1183-1197.

Cats, G.J. and W. Wergen (1982): Analysis of large scale normal modes by the ECMWF analysis system. Report of Workshop on Current Problems in Data Assimilation, ECMWF, November 1982.

Janjić, Z. (1984): General introduction to Limited-Area Modelling. In PSMP Publication Series No. 8, WMO.

Koo, C.C. (1959): On the equivalency of formulations of weather forecasting as an initial-value problem and as an evolution problem. "The Atmosphere and the Sea in Motion". Rossby Memorial Volume, Rockefeller Institute Press. 505-509.

Lambert, S.J. and P.E. Merilees (1978): A study of planetary wave errors in a spectral numerical weather prediction model. Atmos. Ocean., 16, 197-211.

Lorenz, E., (1982): Atmospheric predictability experiments with a large numerical model. Tellus, 34, 505-513.

Lynch, P. (1984): DYNAMO - A one-dimensional primitive equation model. Tech. Note No. 44. Irish Meteorological Service, Dublin.

Lynch, P. (1985): Initialization using Laplace transforms. Q.J.Roy.Met.Soc., 111, 243-258.

Roodenburg, J. (1984): Semi-interactieve correctiemethode voor ECMWF grondprognoses. Memorandum DM-84-3. KNMI. (Unpublished manuscript)

Tennekes, H. (1978): Turbulent flow in two and three dimensions. Bull. Am. Met. Soc., 59, 22-28.

Wallace, J.M. and J.K. Woessner (1981): An analysis of forecast error in the NMC hemispheric primitive equation model. Mon. Wea. Rev., 109, 2444-2449.