FILTERED EQUATIONS AND FILTERING INTEGRATION SCHEMES

Peter Lynch
Meteorological Service
Dublin, Ireland

I still consider the elimination or dampening of noise to be the crucial problem in weather analysis and prediction.
K.H. Hinkelmann, WMO Bulletin, **34**(4), 279, Oct., 1985

1. INTRODUCTION

The spectrum of atmospheric motions is vast, encompassing phenomena having periods ranging from seconds to millennia. The motions of interest to the forecaster have timescales of the order of a day, but the mathematical models used for numerical prediction are quite general, and describe a broader span of dynamical features than those of direct concern. For many purposes these higher frequency components can be regarded as *noise* contaminating the motions of meteorological interest. The elimination of this noise has been achieved by adjustment of the initial fields (a process called *initialization*) or by modification of the governing equations (called *filtering* the equations). There is a close relationship between these two approaches: the diagnostic constraints imposed to initialize the fields may also be used to replace prognostic components of the prediction system, and thus the constraints may be applied throughout the forecast. Filtered equation systems are discussed in §2 below, and their relationship to normal mode initialization is considered.

As an alternative to modification of the equations, a numerical integration scheme may be employed having the property that it selectively eliminates or dampens elements of the solution which are considered to be noise, while simulating the meteorologically significant components accurately. A number of such filtering integration schemes are examined in §3. Following that, §4 introduces the theory of digital filters. Two applications of these filters are described in §5, one to initialization and one to integration, and it is argued there that an integration scheme having a specified frequency response may be constructed using filter theory. The final section (§6) attempts to synthesize the ideas discussed in the preceeding parts, and raises some important unsolved problems.

1.1 Signals and Noise: the Spectrum of Atmospheric Motions

The natural oscillations of the atmosphere fall into two groups (see, for example, Holton, 1975, §2.4). The solutions of meteorological interest have low frequencies and are close to geostrophic balance. They are called rotational modes since their vorticity is greater than their divergence; if divergence is ignored, these modes reduce to the Rossby-Haurwitz waves. There are also very fast gravity-inertia wave solutions, with phase speeds of hundreds of metres per second and large divergence. For typical conditions of large scale atmospheric flow (when the Rossby and Froude numbers are small) the two types of motion are clearly separated and interactions between them are weak. The high frequency gravity-inertia waves may be locally significant in the vicinity of steep orography, where there is strong thermal forcing or where very rapid changes are occurring; but overall they are of minor importance and may be regarded as undesirable noise.

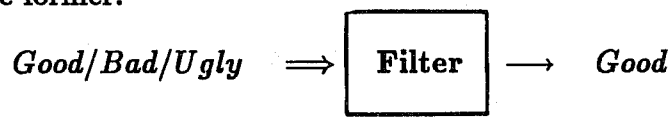
Observations show that the atmospheric pressure and wind fields in regions not too close to the equator are close to a state of geostrophic balance and the flow is quasi-nondivergent. The bulk of the energy is contained in the slow rotational motions and the amplitude of the high frequency components is small. The existence of this geostrophic balance is a perennial source of interest; it is a consequence of the forcing mechanisms and dominant modes of hydrodynamic instability and of the manner in which energy is dispersed and dissipated in the atmosphere. The gravity-inertia waves are instrumental in the process by which the balance is maintained, but the nature of the sources of energy ensures that the low frequency components predominate in the large scale flow. The atmospheric balance is subtle, and difficult to specify precisely. It is delicate in that minor perturbations may disrupt it but robust in that local imbalance tends to be rapidly removed through radiation of gravity-inertia waves in a process known as geostrophic adjustment.

When the primitive equations are used for numerical weather prediction the forecast usually contains spurious large amplitude high frequency oscillations. These result from anomalously large gravity-inertia waves which occur because the balance between the mass and velocity fields is not reflected faithfully in the numerically analysed fields. The problem is that small errors in the initial fields of pressure and wind can

lead to large deviations from a balanced state. As a result, high frequency oscillations of large amplitude are engendered, and these may persist for a considerable time unless strong dissipative processes are incorporated in the forecast model. It was the presence of such imbalance in the initial fields which gave rise to the totally unrealistic pressure tendency of 145 $hPa/6h$ obtained by Lewis Fry Richardson in the first-ever objective numerical weather forecast.

1.2 The Primitive Notion of Filtering

The concept of filtering has a rôle in virtually every field of study, from topology to theology, seismology to sociology. The process of filtering involves the *selection* of those components of an assemblage having some particular property, and the removal or elimination of those components which lack it. A filter is any device or contrivance designed to carry out such a selection. It may be represented by a simple system diagram, having an input with both desired and undesired components, and an output comprising only the former:



We are primarily concerned with filters as used in signal processing. The selection principle for these is generally based on the frequency of the signal components. There are a number of ideal types, lowpass, highpass, bandpass and bandstop, corresponding to the range of frequencies which pass through the filter and those which are rejected. In many cases the input consists of a low-frequency (LF) signal contaminated by high-frequency (HF) noise, and the information in the signal can be isolated by using a lowpass filter which rejects the noise. Such a situation is typical for the applications to meteorology discussed below.

Filter theory originated from the need to design electronic circuits with precise frequency-selective characteristics, for radio and telecommunications. These analog filters were constructed from capacitors and inductors, and acted on continuous time signals. More recently, discrete time signal processing has assumed prominance, and the technique and theory of *digital filtering* has evolved. Digital filters may be implemented in hardware using integrated curcuits, but are more commonly realized in software: the input is processed by a program designed to perform the required selection and compute the output.

2. FILTERED EQUATIONS

2.1 A Little History

Richardson's forecast came to grief through his use of unbalanced data as initial conditions for the primitive equations. The spuriously large projection of this data onto the high frequency gravity waves resulted in an unrealistically large value for the pressure tendency, ruining the forecast. In planning for the first computer forecast, Charney recognised that the problem encountered by Richardson could be avoided in one of two ways. Either the initial data could be adjusted to reduce the high frequency components, or the prognostic equations could be modified to eliminate the solutions corresponding to the noise. The first option is referred to as *initialization* and the second as *filtering the equations.*

In a letter to Philip Thompson dated February 12, 1947, Charney outlined his ideas about filtering the noise. He drew an analogy between a forecasting model and a radio receiver, and argued that the noise could be either eliminated from the input signal or removed by a filtering system in the receiver. He described a method of filtering the equations in a particular case, but concluded "I still don't know what types of approximation have to be made in more general situations". It did not take him long to find out. In a second letter, dated November 4 the same year, he wrote that "The solution is so absurdly simple that I hesitate to mention it ...the motion of *large-scale* systems is governed by the laws of conservation of potential temperature and potential vorticity *and* by the condition that the field of motion is in hydrostatic and *geostrophic* balance. This is the required filter!". [The two letters are reproduced in an article by Thompson in Lindzen *et al.*, 1990.]

Charney (1948) examined the equations using the technique of *scale analysis*, and was able to simplify them in such a way that the gravity wave solutions were completely eliminated. The resulting equations are known as the *quasi-geostrophic* system. The system boils down to a single prognostic equation for the quasi-geostrophic potential vorticity:

$$\left(\frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla\right)\left[f + \zeta + \frac{f_0}{\rho_0}\frac{\partial}{\partial z}\left(\frac{\rho_0}{N^2}\frac{\partial p'/\rho_0}{\partial z}\right)\right] = 0 \qquad (2.1)$$

All that is required by way of initial data to solve this equation is a knowledge of the three-dimensional pressure field.

An alternative and, perhaps, mathematically more rigourous derivation of the quasi-geostrophic system is based on *perturbation theory* (Nayfeh, 1973). This method was first used in meteorology by Kibel' in 1940, when he expanded the equations as power series in a small parameter $\epsilon = V/fL$, now known as the Rossby number. The advantage of this method is the possibility to proceed to higher orders of approximation; the first order gives the quasi-geostrophic system, the next leads to the *balance system.* A hierarchy of successively more accurate approximations may be derived by retaining terms of increasing degree in ϵ. For an account of the history of filtered systems see the article "The Emergence of Quasi-geostrophic Theory", by Phillips in Lindzen *et al.,* 1990.]

2.2 Intermediate Models

Charney (1962) derived a system of equations called the *balance system* by keeping terms to $O(\epsilon^2)$. This system should, in principle, be more accurate than the quasi-geostrophic system. The balance system is highly implicit and difficult to solve. A non-iterative procedure for the integration of the balance equations was presented by Daley (1982). There is a serious flaw in the balance system: these equations have spurious solutions (Moura, 1976) which have no physical counterparts. An alternative system, the *slow equations*, which is also accurate to $O(\epsilon^2)$ but is free from these non-physical solutions, will be derived below.

Hinkelmann (1969) proposed a general method for defining initial data for the primitive equations. He argued that the observed mass and wind fields should be adjusted so that

$$\frac{d^n(\nabla \cdot \mathbf{V})}{dt^n} = 0 \qquad \text{and} \qquad \frac{d^{n+1}(\nabla \cdot \mathbf{V})}{dt^{n+1}} = 0. \tag{2.2}$$

That is, these two conditions should be used to derive diagnostic relationships which the initial data are then required to satisfy. As an alternative, he pointed out that the two conditions could be used to replace two prognostic equations by diagnostic relationships, yielding a general filtered system. The case $n = 0$ yields the quasi-geostrophic equations, while the case $n = 1$ leads to the slow equations (§2.4 below). More general balance conditions based on Hinkelmann's idea have recently been used in connection with the invertibility principle (§2.5 below).

Many investigators have studied systems of equations which are intermediate

in approximation between the primitive equations and the quasi-geostrophic system. Several such systems are reviewed by McWilliams and Gent (1980) and more recently by Allen *et al.* (1990) and a discussion of the advantages and shortcomings of various approximations may be found in these references.

2.3 Normal Mode Initialization and Filtering

Daley (1980) has used the ideas of normal mode initialization (NMI) to develop an efficient integration method for the primitive equations. In effect, his scheme turns a primitive equation model into a filtered model. The model equations can be written in the form

$$\frac{d\mathbf{X}}{dt} + i\mathbf{L}\mathbf{X} + \mathbf{N}(\mathbf{X}) = \mathbf{0} \tag{2.3}$$

with $\mathbf{X}$ the state vector, $\mathbf{L}$ a matrix and $\mathbf{N}$ a nonlinear vector function. If $\mathbf{L}$ is diagonalized, the system separates into two subsystems, for the low and high frequency components:

$$\frac{d\mathbf{Y}}{dt} + i\Lambda_Y \mathbf{Y} + \mathbf{N}_Y(\mathbf{Y},\mathbf{Z}) = \mathbf{0} \tag{2.4}$$

$$\frac{d\mathbf{Z}}{dt} + i\Lambda_Z \mathbf{Z} + \mathbf{N}_Z(\mathbf{Y},\mathbf{Z}) = \mathbf{0} \tag{2.5}$$

where $\mathbf{Y}$ and $\mathbf{Z}$ are the coefficients of the LF and HF components of the flow, referred to colloquially as the *slow* and *fast* components respectively, and Λ_Y and Λ_Z are diagonal matrices of eigenfrequencies for the two types of modes.

Linear NMI imposes the condition $\mathbf{Z} = \mathbf{0}$ at $t = 0$; it has been found that this condition does not produce a noise-free evolution, as nonlinear interractions between the slow modes soon produce HF components. Machenhauer (1977) proposed the balance condition $\dot{\mathbf{Z}} = \mathbf{0}$ at $t = 0$. That is, the initial tendencies of the fast modes are required to vanish. Assuming Machenhauer's criterion to hold throughout the integration, Daley replaced (4) and (5) by the system

$$\dot{\mathbf{Y}} + i\Lambda_Y \mathbf{Y} + \mathbf{N}_Y(\mathbf{Y},\mathbf{Z}) = \mathbf{0} \tag{2.6}$$

$$i\Lambda_Z \mathbf{Z} + \mathbf{N}_Z(\mathbf{Y},\mathbf{Z}) = \mathbf{0} \tag{2.7}$$

giving a prognostic equation for the slow modes and a diagnostic equation for the fast modes. The system (6), (7) may be called the slow equations (in normal mode form). Daley compared an integration using these equations (with $\Delta t = 40$ min) to a control

forecast made with the primitive equations using a leapfrog time scheme and $\Delta t = 10$ min. The integration of the slow system was stable, efficient and accurate.

The Daley scheme has also been implemented in a barotropic version of the HIRLAM Model. A 24 hour forecast with this scheme is shown in Fig. 2.1(a), and a reference forecast based on the primitive equations integrated with a standard leapfrog scheme in Fig. 2.1(b). The two forecasts are very similar; the *rms* differences in height and wind are only $2.67\,m$ and $0.77\,ms^{-1}$.

2.4 The Slow Equations

It is possible to develop a system of equations similar to the system (6), (7) of Daley, but formulated in terms of the physical variables, rather than in normal mode space. Temperton (1987) has devised an initialization scheme which is completely equivalent to the normal mode method, but which operates in physical space. The central idea is to choose a linearization and to configure the equations so that the tendencies can be separated *by inspection* into slow and fast components

$$\dot{\mathbf{X}} = \dot{\mathbf{Y}} + \dot{\mathbf{Z}}. \tag{2.8}$$

Using Machenhauer's criterion, $\dot{\mathbf{Z}} = 0$ at $t = 0$, Temperton derived a technique which he called *implicit* normal mode initialization; the same approach may be used to derive a system of slow equations in physical space.

A general baroclinic system of equations may be separated into a number of systems equivalent to the shallow water equations. Therefore, it is sufficient to consider the shallow water system

$$\dot{\zeta} + f\delta = -N_\zeta \tag{2.9}$$

$$\dot{\delta} - f\zeta + \nabla^2\Phi = -N_\delta \tag{2.10}$$

$$\dot{\Phi} + \bar{\Phi}\delta = -N_\Phi \tag{2.11}$$

where N_ζ, N_δ and N_Φ represent the nonlinear terms and otherwise the notation is conventional. The Coriolis parameter f is variable but the β-terms are included on the *rhs*. An equation expressing conservation of potential vorticity follows directly from the vorticity and continuity equations:

$$\frac{d}{dt}\left[\frac{\zeta + f}{\Phi}\right] = 0 \tag{2.12}$$

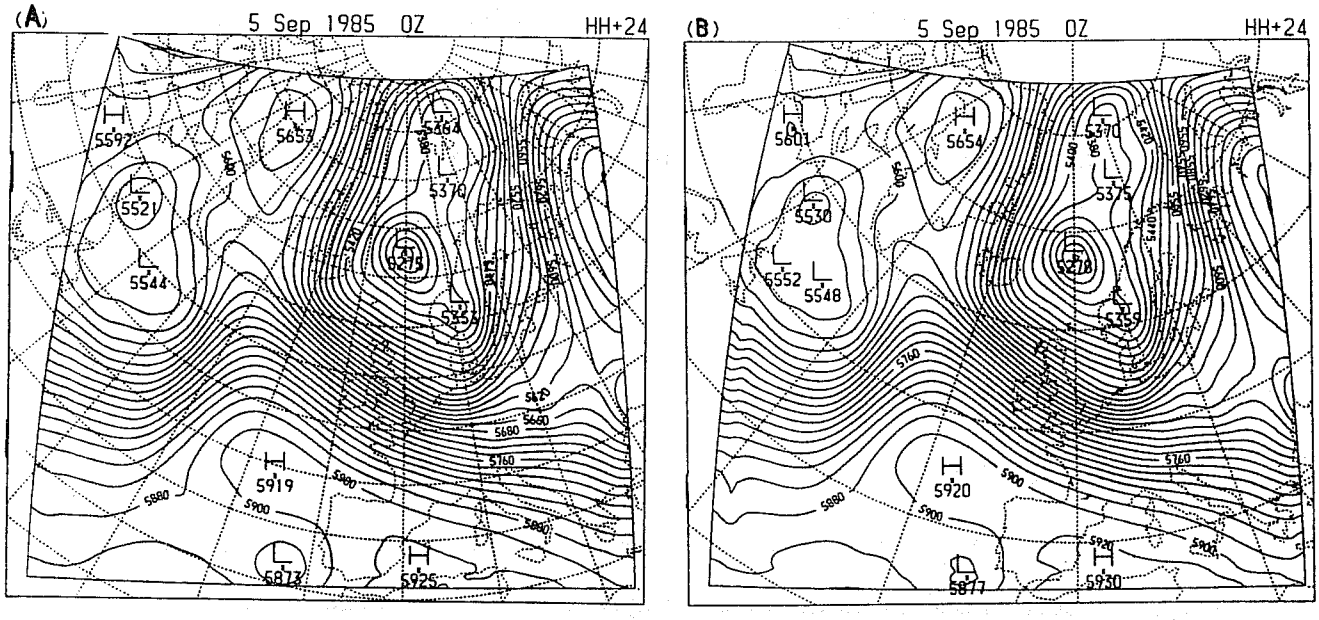


Fig. 2.1. (A) 24 hour forecast with the Daley scheme. (B) Reference forecast based on the primitive equations integrated with a standard leapfrog scheme. The rms differences in height and wind are only 2.67 m and 0.77 m/s.

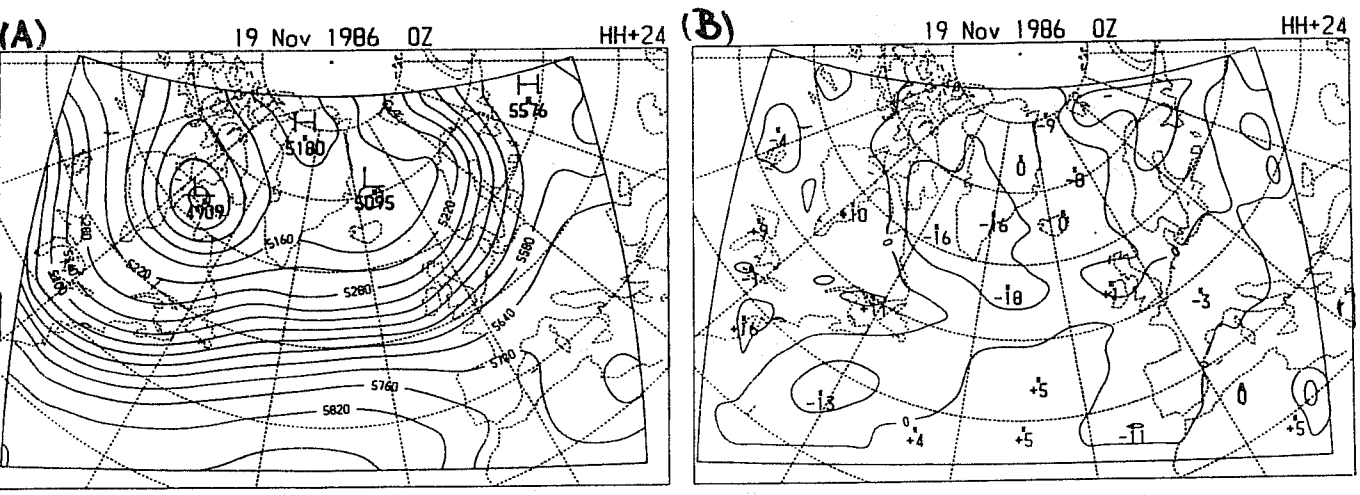


Fig. 2.2 (A) Reference forecast with the primitive equations. (B) Difference between the Slow Equation forecast and the reference. The rms differences in height and wind were 5.7 m and 1.5 m/s.

Alternatively, elimination of δ between these equations yields

$$\frac{\partial}{\partial t}\left(\bar{\Phi}\zeta - f\Phi\right) = -\left[\bar{\Phi}N_\zeta - fN_\Phi\right] \tag{2.13}$$

which is completely equivalent to (12). An equation for the tendency of *geostrophic imbalance* ($\epsilon \equiv \nabla^2\Phi - f\zeta$) is easily derived from (9) and (11):

$$\dot{\epsilon} + \left[\bar{\Phi}\nabla^2 - f^2\right]\delta = -\left[\nabla^2 N_\Phi - fN_\zeta\right] \tag{2.14}$$

The system (10), (13) and (14) is equivalent to the original system (9)–(11). Although the horizontal coordinates are not separable, two crucial properties of the eigenmodes can immediately be deduced: [A] The slow modes are stationary, geostrophic and nondivergent; [B] The fast modes have zero linearized potential vorticity $\left[\bar{\Phi}\zeta - f\Phi\right]$. These properties imply that the tendencies of divergence and imbalance project entirely onto the fast modes and the tendency of potential vorticity entirely onto the slow modes. Thus, the assumption that the gravity wave projections of tendency vanish ($\dot{\mathbf{Z}} = 0$) amounts to dropping the terms $\dot{\delta}$ and $\dot{\epsilon}$ in (10) and (14). The result of applying this assumption is a system with one prognostic and two diagnostic equations:

$$\frac{d}{dt}\left[\frac{\zeta + f}{\Phi}\right] = 0 \tag{2.15}$$

$$\nabla^2\Phi - f\zeta = -N_\delta \tag{2.16}$$

$$\left[\bar{\Phi}\nabla^2 - f^2\right]\delta = -\left[\nabla^2 N_\Phi - fN_\zeta\right] \tag{2.17}$$

This system will be called the *slow equations.* They are equivalent to Daley's equations (6)–(7), But refer to physical variables, obviating the need for transformations to and from Hough space. Further discussion of these equations may be found in Lynch (1989), where it is shown that the system is free from the spurious non-physical solutions of the balance equations.

The slow system (15)–(17) was used to make a 24 hour forecast over a limited area, and the results compared to a reference forecast with the primitive equations. The reference forecast is shown in Fig. 2.2(a) and the difference between the two runs in Fig. 2.2(b). Clearly, the two forecasts are very similar; the *rms* differences in height and wind were $5.7\,m$ and $1.5\,ms^{-1}$. A baroclinic model based on the slow equations has also been developed (Lynch and McDonald, 1990).

2.5 The Invertibility Principle

Rossby based his study of planetary waves on the assumption that the absolute vorticity $\zeta + f$ is conserved. From a knowledge of the absolute vorticity the wind field $\mathbf{V} = (\mathbf{u}, \mathbf{v})$ can be deduced by solution of a Poisson equation:

$$\nabla^2 \psi = \zeta \quad , \quad \mathbf{V} = \mathbf{k} \times \nabla \psi$$

subject to suitable boundary conditions. This is the simplest example of the *invertibility principle*. The absolute vorticity contains full information about the dynamics, provided a suitable balance condition holds; in this case the balance assumption is that the flow is nondivergent.

For divergent barotropic flow, absolute vorticity conservation is supplanted by conservation of potential vorticity (PV), $Q = (\zeta + f)/\Phi$, as expressed in (12). Since the PV involves both the mass and wind fields, it is far from obvious that this field contains essentially all the relevant dynamical information. In terms of the primitive equations it is not possible to disentangle PV to obtain Φ and $\mathbf{V}$. However, if the slow equations (15)–(17) are assumed to describe the dynamics, an equation for Φ follows from the balance equation (16) and the definition of PV:

$$\left[\nabla^2 - fQ\right] \Phi = -\left(N_\delta + f^2\right). \tag{2.18}$$

Assuming appropriate boundary conditions can be prescribed, this Helmholtz equation can be solved for the geopotential. The vorticity ζ follows immediately from PV and Φ. If the divergence is calculated by solving the *imbalance* equation (17), the wind field may then be derived from

$$\nabla^2 \psi = \zeta \quad , \quad \nabla^2 \chi = \delta$$

$$\mathbf{V} = \nabla \chi + \mathbf{k} \times \nabla \psi$$

provided suitable boundary conditions can be specified.

For three-dimensional frictionless, adiabatic motion of a baroclinic, hydrostatic atmosphere conservation of potential vorticity is expressed by the equation

$$\frac{d}{dt}\left[(\zeta_\theta + f)\left(-g\frac{\partial \theta}{\partial p}\right)\right] = 0 \tag{2.19}$$

where $\zeta_\theta = \mathbf{k} \cdot \nabla_\theta \times \mathbf{V}$ is the vorticity in isentropic coordinates. This is the most general form normally used in meteorology, although it has been extended to the non-hydrostatic case by Ertel. The idea that the potential vorticity contains all the essential dynamical information has been formalised in the invertibility principle:

> *If the total mass under each isentropic surface is specified, then a knowledge of the global distribution of potential vorticity on each isentropic surface and of the potential temperature at the lower boundary is sufficient to deduce, diagnostically, all the other dynamical fields, such as winds, temperature, geopotential height, static stability and vertical velocity, under a suitable balance condition.*

(Hoskins, *et al.*, 1985). The last clause is crucial: the accuracy with which the remaining fields can be extricated from the PV depends sensitively on the definition of balance.

McIntyre and Norton (1990) have investigated the inversion of potential vorticity for a hemispheric barotropic model, using a range of balance conditions based on the idea of Hinkelmann as expressed in (2) above. They also considered a hierarchy of balance requirements derived from normal mode initialization theory. Their results show that the information contained on the PV field is remarkably complete. Using a high-order balance assumption, the original mass and wind fields used to construct the PV were recovered with great accuracy; even the divergence and associated vertical velocity fields could be adequately retrieved from the potential vorticity information.

The equation of PV-conservation together with a balance assumption yields a filtered model, the accuracy of which increases with the order of the balance condition. There is, in principle, no limit to the order at which balance may be imposed. Hinkelmann's hierarchy (2) may be applied for any value of n. A balance condition framed in terms of NMI is to require $d^n \mathbf{Z}/dt^n = 0$. For $n = 0$ this is the condition of linear NMI; for $n = 1$ it is Machenhauer's criterion. Lorenz (1980) introduced the concept of *superbalance* defined by the condition

$$\lim_{n \to \infty} \frac{d^n \mathbf{Z}}{dt^n} = 0. \tag{2.20}$$

The objective of combining such a condition with PV-conservation is to obtain a system of equations of the highest possible accuracy, but having no high frequency components in the solution — the ultimate filtered system. Such a system has yet to be explicitly derived.

3. FILTERING INTEGRATION SCHEMES

3.1 Filtering Characteristics of Simple Schemes

The frequency-selection characteristics of time integration schemes can be deduced by analysis of their application to the simple linear advection equation

$$\frac{\partial U}{\partial t} + c\frac{\partial U}{\partial x} = 0 \tag{3.1}$$

which, for a single wave $U(x,t) = u(t)\exp(ikx)$, becomes

$$\frac{du}{dt} + i\omega u = 0 \tag{3.2}$$

with $\omega = kc$. For example, the Euler forward scheme is

$$\frac{u^{n+1} - u^n}{\Delta t} + i\omega u^n = 0$$

which can be solved immediately for u^{n+1}:

$$u^{n+1} = (1 - i\omega\Delta t)u^n = \lambda_0 u^n. \tag{3.3}$$

This is unstable for all frequencies, since $|\lambda_0| = \sqrt{1+\omega^2\Delta t^2} > 1$. Higher frequencies are amplified most by this scheme.

The Euler backward or Matsuno scheme is defined by a two step process:

$$\frac{u^* - u^n}{\Delta t} + i\omega u^n = 0$$

$$\frac{u^{n+1} - u^n}{\Delta t} + i\omega u^* = 0.$$

Solving for u^{n+1}, one obtains

$$u^{n+1} = (1 - \omega^2\Delta t^2 - i\omega\Delta t)u^n = \lambda u^n$$

and the modulus of the *amplification factor* λ is

$$|\lambda| = \sqrt{1 - (\omega\Delta t)^2 + (\omega\Delta t)^4} \tag{3.4}$$

which is 1 for $\omega = 0$, less than 1 for $|\omega\Delta t| < 1$ and greater than 1 for $|\omega\Delta t| > 1$ (solid curve, Fig. 3.1). To ensure stability, Δt is chosen so that $\omega\Delta t$ is less than 1 for the highest frequency occurring.

The Courant-Friedrichs-Lewy criterion requires $\omega\Delta t < 1$ for stability. However, $|\lambda|$ is a minimum for $\omega\Delta t = 1/\sqrt{2}$, and normally Δt is chosen so that this equality holds for the highest frequency occurring. In that case the damping increases monotonically with frequency. It should be noted that the Euler backward scheme is a rather blunt instrument as far as frequency selection is concerned. The amount of damping is controled by Δt; there is no other parameter available to determine a cutoff frequency separating modes to be retained from those to be removed. For $\omega\Delta t$ small, the damping factor is given approximately by

$$\lambda \approx 1 - \tfrac{1}{2}\omega^2\Delta t^2. \tag{3.5}$$

Thus, with $\Delta t = 360s$, motions with periods of 6 hours are damped by a factor of about 0.995 for each timestep. For a period $\tau = 12h$ the value is 0.99875 while for $\tau = 3h$ it is 0.98. Even the highest frequency is damped by a factor of only 0.866; ideally, it should be close to zero.

Kurihara (1965) investigated the properties of the following scheme:

$$\frac{u^* - u^n}{\Delta t/2} + i\omega u^n = 0$$

$$\frac{u^{**} - u^n}{\Delta t} + i\omega u^* = 0.$$

$$\frac{u^{n+1} - u^n}{\Delta t} + i\omega u^{**} = 0.$$

Eliminating intermediate quantities, one obtains for u^{n+1}:

$$u^{n+1} = \left[(1 - \omega^2\Delta t^2) - i\omega\Delta t(1 - \tfrac{1}{2}\omega^2\Delta t^2)\right] u^n = \lambda u^n$$

and the modulus of the amplification factor λ is easily found to be

$$|\lambda| = \sqrt{1 - (\omega\Delta t)^2 + \tfrac{1}{4}(\omega\Delta t)^6}.$$

This response is shown in Fig. 3.1 (dashed curve). It is more frequency-selective than the Euler-backward scheme, but more expensive; for a nonlinear equation, an extra evaluation of the nonlinear terms is required. Many other variations of the Euler backward scheme have been formulated. Regarded as filtering schemes, they tend to be inflexible and to damp the signal along with the noise.

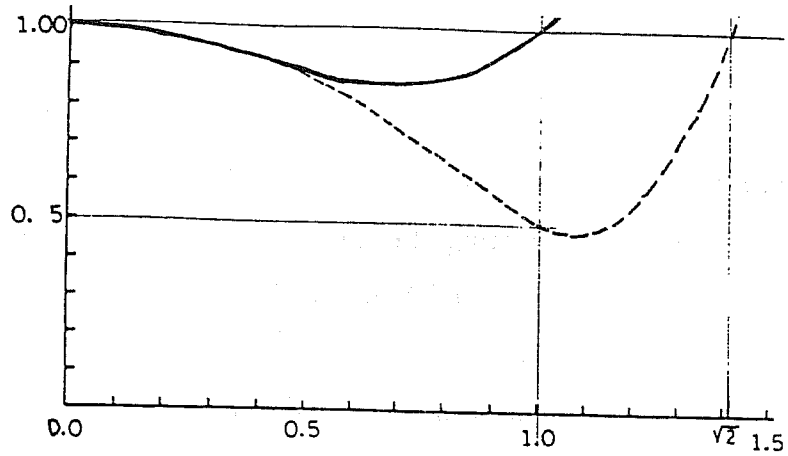
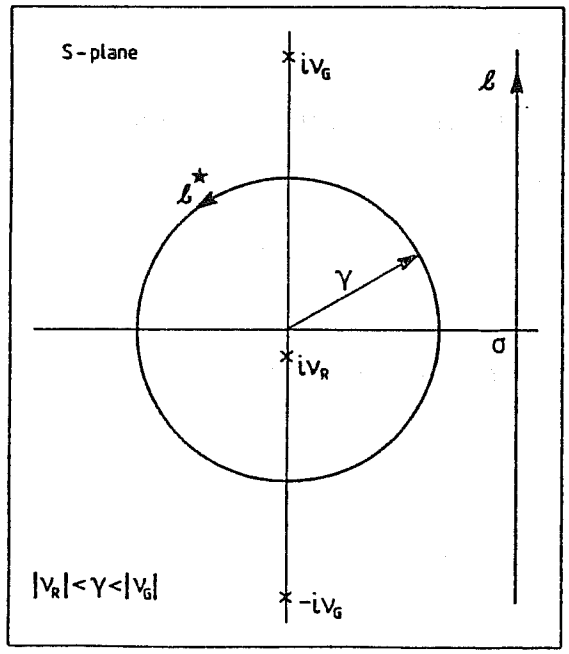
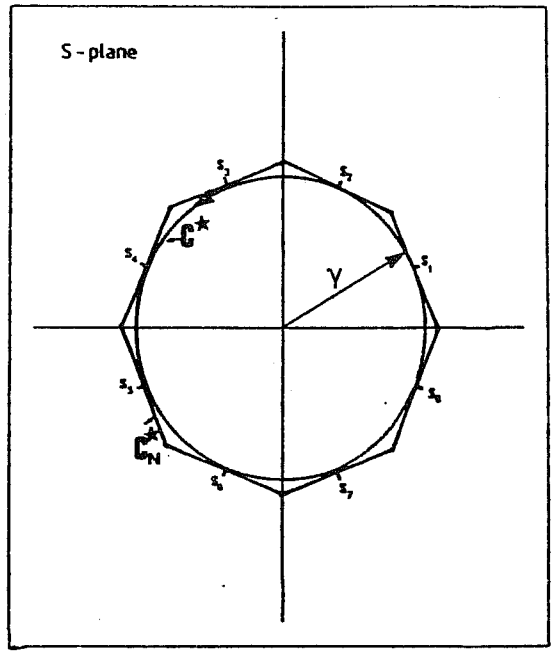


Fig. 3.1. Amplification factor for the Euler backward scheme (solid line) and for the modified Euler backward scheme (dashed line).



Contours in the s plane used for the regular and modified inverse Laplace transform. The value of γ is chosen to separate the rotational frequencies (ν_R) and the gravity-wave frequencies (ν_G).

Fig. 3.2.



Contour C^* used to define the modified operator Ω^*, and circumscribed polygon C_N^* used in approximating the contour integral by a finite sum Ω_N^*.

Fig. 3.3.

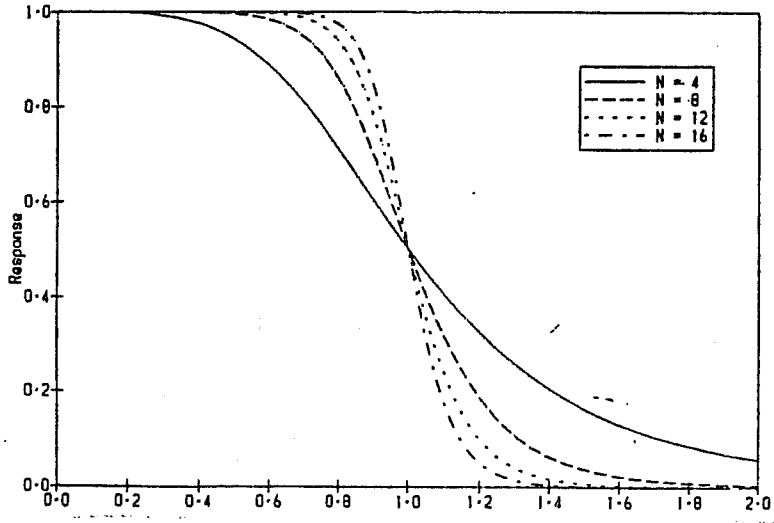


Fig. 3.4. Response or amplification curves for various values of N , for the inversion operator evaluated on an N-gon.

3.2 Laplace and $\mathcal{Z}$-transform schemes

A filtering integration scheme which has arbitrarily sharp frequency selection properties has been formulated using the Laplace transform (LT) technique (Lynch, 1991). This scheme is capable of faithfully simulating the low frequency evolution of the atmosphere whilst eliminating high frequency oscillations. The scheme has also been combined with a Lagrangian treatment of advection, giving stable integrations for long time-steps.

A brief description of the LT scheme will be given here; for fuller details, see Lynch (1991) and references therein. Consider a function $f(t)$ with LF and HF components

$$f(t) = A_r \exp(i\omega_r t) + A_g \exp(i\omega_g t) \quad , \quad |\omega_r| \ll |\omega_g| \tag{3.6}$$

where ω_r represents a Rossby wave frequency and ω_g a gravity wave frequency. The Laplace transform of (6) is

$$\hat{f}(s) = \frac{A_r}{s - i\omega_r} + \frac{A_g}{s - i\omega_g} \tag{3.7}$$

with a pole near the origin (at $s = i\omega_r$) corresponding to the LF component (Rossby wave) and a pole far from the origin (at $s = i\omega_g$) for the HF component (Fig. 3.2). The original function is recovered by applying the inverse LT:

$$f(t) = \mathcal{L}^{-1}\{\hat{f}\} = \frac{1}{2\pi i} \int_{\mathbf{C}} e^{st} \hat{f}(s)\, ds \tag{3.8}$$

where **C** is a line parallel to the imaginary axis and to the right of the origin.

The HF component of $f(t)$ may be eliminated by replacing **C** by a circle $\mathbf{C}^\star$ of radius ω_c, such that $|\omega_r| < \omega_c < |\omega_g|$ (Fig. 3.2):

$$f^\star(t) = \mathcal{L}^\star\{\hat{f}\} = \frac{1}{2\pi i} \oint_{\mathbf{C}^\star} e^{st} \hat{f}(s)\, ds. \tag{3.9}$$

The value of the integral defining $f^\star$ is determined by the residue of the integrand at the pole $s = i\omega_r$ falling *within* $\mathbf{C}^\star$ and corresponding to the Rossby wave. The pole at $s = i\omega_g$, arising from the HF component, falls *outside* $\mathbf{C}^\star$ and contributes nothing. Thus,

$$f^\star(t) = A_r \exp(i\omega_r t) \tag{3.10}$$

so that $\mathcal{L}^\star\mathcal{L}$ acts as an *ideal lowpass filter* with a cutoff frequency ω_c.

In practice the integral (9) must be evaluated numerically. The circle C^* is approximated by a circumscribed N-gon C_N^* and the integrand is evaluated at the midpoints s_n of C_N^* (Fig. 3.3). The discrete modified inverse Laplace transform is then given by

$$\mathcal{L}_N^*\{\hat{f}\} = \frac{1}{2\pi i}\sum_{n=1}^{N} e^{s_n t}\hat{f}(s_n)\Delta s_n . \tag{3.11}$$

The numerical operator $\mathcal{L}_N^*\mathcal{L}$ is no longer an ideal lowpass filter. It can be shown that, if the frequency response $R_N(\omega)$ is defined by

$$\mathcal{L}_N^*\mathcal{L}\{e^{i\omega t}\} = R_N(\omega)\cdot e^{i\omega t}$$

then R_N is given approximately by the following expression:

$$R_N(\omega) \approx \frac{1}{1+(i\omega/\omega_c)^N}. \tag{3.12}$$

For N a multiple of 4, this is the square of the response of a Butterworth filter of order $N/2$. It approximates a step function with a corner point at $\omega = \omega_c$. The slope of the response at the cutoff point, and thus the width of the transition between pass- and stop-bands, can be made arbitrarily sharp by increasing N. Curves of $R_N(\omega)$ for various values of N are shown in Fig. 3.4.

Now consider the application of the LT technique to the simple oscillation equation (2). The Laplace transform of (2) with time-origin at $t_n = n\Delta t$, is

$$s\hat{u} - u_n + i\omega\hat{u} = 0$$

which is immediately soluble for $\hat{u}$, giving

$$\hat{u} = \frac{u_n}{s+i\omega}.$$

Now applying $\mathcal{L}_N^*$ at time $t = t_n + \Delta t$ gives

$$u_{n+1}^* = \mathcal{L}_N^*\left\{\frac{u_n}{s+i\omega}\right\}\Bigg|_{t=\Delta t} = R_N(\omega)\cdot u_n \exp(i\omega\Delta t)$$

which is the analytical solution $u_{n+1} = u_n \exp(i\omega\Delta t)$ multiplied by the response function $R_N(\omega)$. Thus, one obtains

$$u_{n+1}^* \approx u_{n+1} \quad , \quad |\omega| \ll \omega_c$$

$$u^*_{n+1} \approx 0 \quad , \quad |\omega| \gg \omega_c$$

and the scheme preserves LF components and annihilates HF constituents.

The application of the LT method to a general nonlinear system will now be described. Let the state of the system at time t be specified by $\mathbf{X}(t)$, which is governed by (2.3):

$$\frac{d\mathbf{X}}{dt} + i\mathbf{L}\mathbf{X} + \mathbf{N}(\mathbf{X}) = \mathbf{0} \tag{3.13}$$

where $\mathbf{L}$ is a matrix and $\mathbf{N}$ a nonlinear vector function. If the system is in the state $\mathbf{X}^0$ at time $t = 0$ then the LT of this equation is

$$\mathbf{M}\hat{\mathbf{X}} + \hat{\mathbf{N}}(\mathbf{X}) = \mathbf{X}^0$$

where $\mathbf{M}(s) = (s\mathbf{I} + i\mathbf{L})$ with $\mathbf{I}$ the identity matrix. If one considers the evolution of the system over a short time interval $(0, \Delta t)$, and assumes that the nonlinear term does not vary, it follows that

$$\hat{\mathbf{X}} = \mathbf{M}^{-1}\left[\mathbf{X}^0 - \mathbf{N}^0/s\right]$$

where $\mathbf{N}^0 = \mathbf{N}(\mathbf{X}^0)$. To find the solution at time $t = \Delta t$, the inverse LT is applied:

$$\mathbf{X}(\Delta t) = \mathcal{L}^{-1}\left\{\mathbf{M}^{-1}\left[\mathbf{X}^0 - \mathbf{N}^0/s\right]\right\}\Big|_{t=\Delta t}.$$

If only the slowly varying component of $\mathbf{X}$ is of interest, $\mathcal{L}^{-1}$ may be replaced by the modified inverse $\mathcal{L}^*$, which acts to filter out the HF components:

$$\mathbf{X}^1 = \mathbf{X}^*(\Delta t) = \mathcal{L}^*\left\{\mathbf{M}^{-1}\left[\mathbf{X}^0 - \mathbf{N}^0/s\right]\right\}\Big|_{t=\Delta t}.$$

Having the solution at $t = \Delta t$, one may proceed stepwise to extend the forecast: the solution is advanced from $n\Delta t$ to $(n+1)\Delta t$ by

$$\mathbf{X}^{n+1} = \mathcal{L}^*\left\{\mathbf{M}^{-1}\left[\mathbf{X}^n - \mathbf{N}^n/s\right]\right\}\Big|_{t=\Delta t}. \tag{3.14}$$

Numerous other formulations of the timestepping algorithm are also possible.

The LT scheme was implemented in a limited area barotropic model, and compared to a conventional scheme (reference model). The initial data was the same as used for the Daley scheme described in §2.4. The 24 hour LT forecast is shown in Fig.

3.5(a); it is very similar to the reference, which was shown in Fig. 2.1(b). Fig. 3.5(b) shows the difference between the two forecasts (LT–Ref.). The *rms* height and wind differences between the LT and reference forecasts were $3.23\,m$ and $0.65\,ms^{-1}$. These are comparable to the values $2.67\,m$ and $0.77\,ms^{-1}$ obtained for the Daley scheme.

A number of perturbation experiments were also carried out to demonstrate the *shock-proof* nature of the LT scheme (see Lynch, 1991 for details of initial fields). A height perturbation, shown in Fig. 3.6(a) was added to the mass field after one hour of the forecast; it consists of a High and a Low pressure anomaly, both of amplitude $120m$. The wind field was adjusted by the corresponding geostrophic wind perturbation. The time step was 1 hour for both forecasts. The absolute values of the maximum positive (High) and negative (Low) differences between the perturbed and unperturbed forecasts are shown in Fig. 3.6(b) (the LT results are denoted SALT and the reference forecast by SLSI). For both models the High weakens and the Low intensifies, which is consistent with geostrophic adjustment theory. For cyclonic flow the geostrophic wind exceeds the gradient wind, so, if the wind perturbation is assimilated the Low must deepen to maintain gradient balance; by a similar argument the High must weaken. The character of the response differs for the two models: the LT model adjusts rapidly to the inserted data, and the evolution is smooth after HH+02; for the reference model (SLSI) the amplitudes of the High and Low continue to oscillate for several hours. The ability of the LT scheme to assimilate a perturbation without data shock is an attractive feature of the scheme. If data is to be inserted hour by hour, noise from earlier insertions may interfere with quality control and assimilation of later ones. The reference model obviously suffers in this respect; the response of the LT scheme is greatly superior, and the method has considerable potential for asynoptic data assimilation.

A filtering scheme based on the $\mathcal{Z}$-transform (ZT), the discrete analogue of the LT, has been devised, and the two schemes have been compared (Lynch, 1991). Both were found to produce very similar results. The ZT is applied to a system of equations which have already been discretised with respect to time, so it may be slightly easier to adapt an existing model to use this scheme. Properties of the $\mathcal{Z}$-transform are given in the Appendix of Lynch (1991); this transform also plays a central rôle in the analysis of digital filters, to be discussed next.

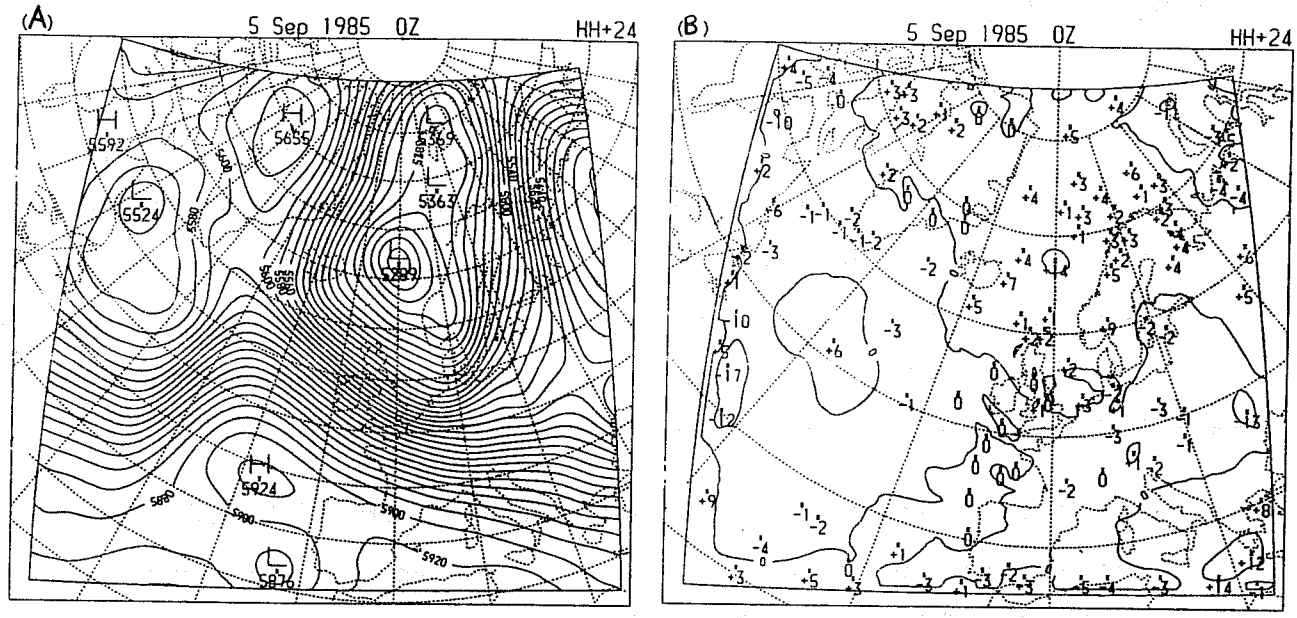


Fig. 3.5. (A) 24 hour forecast with the LT scheme. (B) Difference between the forecast with the LT scheme and the reference (LT—Ref.). The rms height and wind differences between the LT and reference forecasts were 3.23 m and 0.65 m/s.

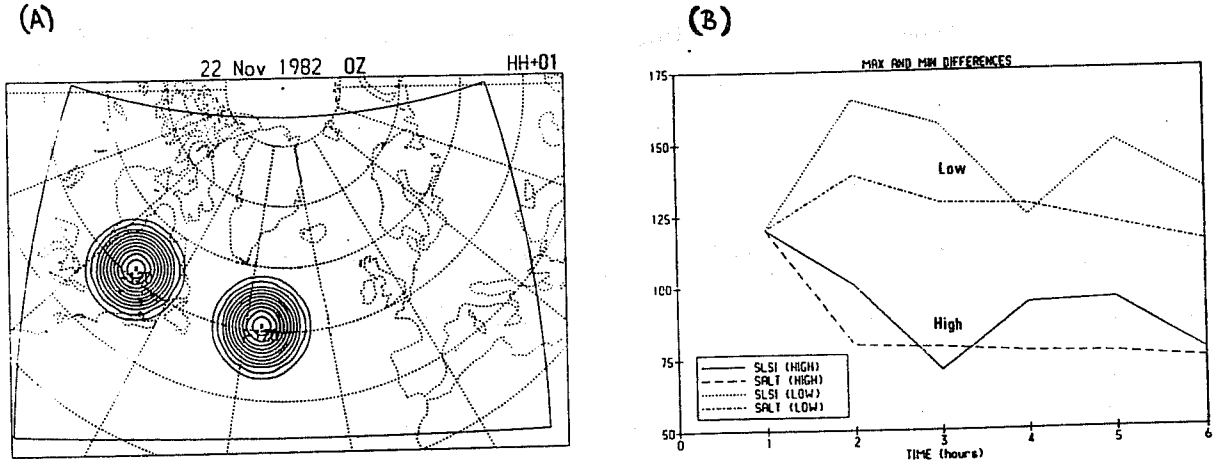


Fig. 3.6 (A) Height perturbation, which was added to the mass field after one hour of the forecast. The wind field was adjusted by the corresponding geostrophic wind perturbation. (B) Absolute values of the maximum positive (High) and negative (Low) differences between the perturbed and unperturbed forecasts (the LT results are denoted SALT and the reference forecast by SLSI).

4. DIGITAL FILTERS

Several filtering integration schemes have been considered above. The Euler-backward scheme is simple to apply but suffers from poor selectivity. The Laplace and $\mathcal{Z}$-transform schemes are highly selective but more complex and computationally demanding. It is desirable to develop schemes formulated in the time domain rather than in some transform space, which are simple to apply, numerically economical and sufficiently selective. Schemes which are both computationally economical and highly discriminative with respect to frequency appear to be lacking at present.

The theory of digital filters can be used to construct a time integration scheme having a specified frequency response. Thus, if the desired amplification or damping is given as a function of frequency, a finite differencing scheme may be designed having a frequency response which approximates the specified transfer function to any accuracy required. The usual method of forming damping schemes has been an empirical process, where a time-stepping algorithm is defined and its frequency response characteristics subsequently examined. In contrast, the digital filtering approach enables one to construct filtering time integration schemes in an efficient, systematic and objective manner.

4.1 Definition of Digital Filters

Given a discrete function of time, $\{x_n\}$, a *nonrecursive* digital filter is defined by

$$y_n = \sum_{k=-N}^{N} a_k x_{n-k} . \tag{4.1}$$

The output y_n at time $n\Delta t$ depends on both past and future values of x_n, but not on other output values. A *recursive* digital filter is defined by

$$y_n = \sum_{k=K}^{N} a_k x_{n-k} + \sum_{k=1}^{L} b_k y_{n-k} . \tag{4.2}$$

The output y_n at time $n\Delta t$ in this case depends on past and present values of the input (for $K = 0$), and also on previous output values. (Occasionally, future input values are also used ($K < 0$), in which case the recursive filter is *non-causal*). Recursive filters are more powerful than non-recursive ones, but can also be more problematical, as the feedback of the output can give rise to instability. The response of a nonrecursive filter

to an impulse $\delta(n)$ is zero for $|n| > N$, giving rise to the alternative name *finite impulse response* or FIR filter. Since the response of a recursive filter to this input can persist indefinitely, it is known as an *infinite impulse response* or IIR filter.

The frequency response of a recursive filter is easily found: let $x_n = \exp(in\theta)$ and assume an output of the form $y_n = H(\theta)\exp(in\theta)$; substituting into (2), the transfer function $H(\theta)$ is

$$H(\theta) = \frac{\sum_{k=K}^{N} a_k e^{-ik\theta}}{1 - \sum_{k=1}^{L} b_k e^{-ik\theta}}. \tag{4.3}$$

For nonrecursive filters the denominator reduces to unity. This equation gives the response once the filter coefficients a_k and b_k have been specified. However, what is really required is the opposite: to derive coefficients (and as few as possible) which will yield the desired response function. This *inverse problem* has no unique solution, and a great variety of techniques have been developed. Only the most elementary design techniques will be considered here; for further information see, for example, Parks and Burrus (1987).

Recursive filters generally have superior performance to nonrecursive filters with the same total number of coefficients. This may be explained by noting that the transfer function (3) can be written

$$H(\theta) = \frac{\sum_{k=K}^{N} a_k z^{-k}}{1 - \sum_{k=1}^{L} b_k z^{-k}} \tag{4.4}$$

where $z = \exp(i\theta)$. For a nonrecursive filter this is a polynomial in $1/z$; for a recursive function it is a rational function in $1/z$, and is more capable of fitting a specified function having sudden changes or narrow features. Against this, the recursive filter will obviously cause problems if the denominator vanishes. It can be shown that a recursive filter is stable if the roots of the *characteristic polynomial*

$$z^L - \sum_{k=1}^{L} b_k z^{L-k} = 0$$

are inside the unit circle $|z| \leq 1$.

Numerous accounts of recursive digital filters are available in publications on digital signal processing (*e.g.*, Strum and Kirk, 1988; Oppenheim and Schafer, 1989).

For a review in the meteorological literature, see Raymond (1991), where another class of filter, the implicit filter, is also discussed.

4.2 Design of Nonrecursive Filters

Consider a function of time, $f(t)$, with low and high frequency components. To filter out the high frequencies one may proceed as follows:

[1] calculate the Fourier transform $F(\omega)$ of $f(t)$;

[2] set the coefficients of the high frequencies to zero;

[3] calculate the inverse transform.

Step [2] may be performed by multiplying $F(\omega)$ by an appropriate weighting function $H_c(\omega)$. Typically, $H_c(\omega)$ is a step function

$$H_c(\omega) = \begin{cases} 1, & |\omega| \leq |\omega_c|; \\ 0, & |\omega| > |\omega_c|, \end{cases} \tag{4.5}$$

where ω_c is a cutoff frequency. These three steps are equivalent to a convolution of $f(t)$ with $h(t) = \sin(\omega_c t)/\pi t$, the inverse Fourier transform of $H_c(\omega)$. This follows from the convolution theorem

$$\mathcal{F}\{(h * f)(t)\} = \mathcal{F}\{h\} \cdot \mathcal{F}\{f\} = H_c(\omega) \cdot F(\omega) \tag{4.6}$$

Thus, to filter $f(t)$ one calculates

$$f^*(t) = (h * f)(t) = \int_{-\infty}^{+\infty} h(\tau) f(t - \tau) d\tau. \tag{4.7}$$

For simple functions $f(t)$, this integral may be evaluated analytically. In general, some method of approximation must be used.

Suppose now that f is known only at discrete moments $t_n = n\Delta t$, so that the sequence $\{\cdots, f_{-2}, f_{-1}, f_0, f_1, f_2, \cdots\}$ is given. For example, f_n could be the value of some model variable at a particular grid point at time t_n. The shortest period component which can be represented with a time step Δt is $\tau_N = 2\Delta t$, corresponding to a maximum frequency, the so-called Nyquist frequency, $\omega_N = \pi/\Delta t$. The sequence $\{f_n\}$ may be regarded as the Fourier coefficients of a function $F(\theta)$:

$$F(\theta) = \sum_{n=-\infty}^{\infty} f_n e^{-in\theta},$$

where $\theta = \omega\Delta t$ is the *digital frequency* and $F(\theta)$ is periodic, $F(\theta) = F(\theta + 2\pi)$. High frequency components of the sequence may be eliminated by multiplying $F(\theta)$ by a function $H_d(\theta)$ defined by

$$H_d(\theta) = \begin{cases} 1, & |\theta| \le |\theta_c|; \\ 0, & |\theta| > |\theta_c|, \end{cases} \tag{4.8}$$

where the cutoff frequency $\theta_c = \omega_c \Delta t$ is assumed to fall in the Nyquist range $(-\pi, \pi)$ and $H_d(\theta)$ has period 2π. This function may be expanded:

$$H_d(\theta) = \sum_{n=-\infty}^{\infty} h_n e^{-in\theta} \quad ; \quad h_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} H_d(\theta) e^{in\theta}\, d\theta. \tag{4.9}$$

The values of the coefficients h_n follow immediately from (8) and (9):

$$h_n = \frac{\sin n\theta_c}{n\pi}. \tag{4.10}$$

Let $\{f_n^\star\}$ denote the low frequency part of $\{f_n\}$, from which all components with frequency greater than θ_c have been removed. Clearly,

$$H_d(\theta) \cdot F(\theta) = \sum_{n=-\infty}^{\infty} f_n^\star e^{-in\theta}.$$

The convolution theorem for Fourier series now implies that $H_d(\theta)\cdot F(\theta)$ is the transform of the convolution of $\{h_n\}$ with $\{f_n\}$:

$$f_n^\star = (h * f)_n = \sum_{k=-\infty}^{\infty} h_k f_{n-k}. \tag{4.11}$$

This enables the filtering to be performed directly on the given sequence $\{f_n\}$. It is the discrete analogue of (7). In practice the summation must be truncated at some finite value of k. Thus, an approximation to the low frequency part of $\{f_n\}$ is given by

$$f_n^\star = \sum_{k=-N}^{N} h_k f_{n-k}. \tag{4.12}$$

Comparing (12) with (1), it is apparent that the finite approximation to the discrete convolution is formally identical to a nonrecursive digital filter.

As is well known, truncation of a Fourier series gives rise to Gibbs oscillations. These may be greatly reduced by means of an appropriately defined "window" function. The response of the filter is improved if h_n is multiplied by the Lanczos window

$$w_n = \frac{\sin\big(n\pi/(N+1)\big)}{n\pi/(N+1)}.$$

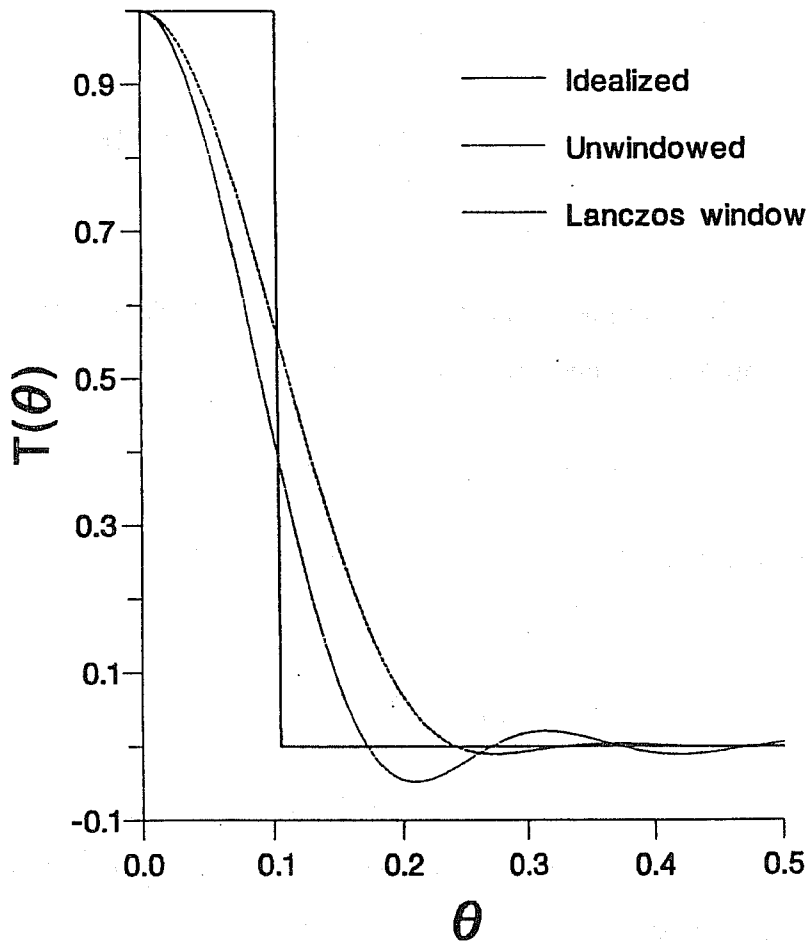


Fig. 4.1.

Transfer functions for the filter defined by the coefficients $h_n = \sin(n\omega_c\Delta t)/n\pi$, where the cutoff period $\tau_c = 2\pi/\omega_c$ is 6 hours, with and without modification by a Lanczos window.

The transfer function $H(\theta)$ of a filter is defined as the function by which a pure sinusoidal oscillation is multiplied when subjected to the filter. For symmetric coefficients, $h_k = h_{-k}$, it is real, implying that the phase is not altered by the filter. Then, if $f_n = \exp(in\theta)$, one may write $f_n^\star = H(\theta) \cdot f_n$, and $H(\theta)$ is easily calculated by substituting f_n in (8):

$$H(\theta) = \sum_{k=-N}^{N} h_k e^{-ik\theta} = \left[h_0 + 2 \sum_{k=1}^{N} h_k \cos k\theta \right]. \qquad (4.13)$$

The transfer functions for a windowed and unwindowed filter are shown in Fig. 4.1. These were calculated for the cutoff period $\tau_c = 6$ hours, span $T_s = 2N\Delta t = 6$ hours and timestep $\Delta t = 360\,s$ (as used in the application to initialization described below). The parameter values are therefore $N = 30$ and $\theta_c = \pi/30 \approx 0.1$. It can be seen that the use of the window decreases the Gibbs oscillations in the stop-band $|\theta| > |\theta_c|$. However, it also has the effect of widening the pass-band beyond the nominal cutoff. For a fuller discussion of windowing see *e.g.* Hamming (1989) or Oppenheim and Schafer (1989).

One of the simplest design methods for nonrecursive filters is the expansion of the desired filtering function, $H(\theta)$, as a Fourier series, and the application of a suitable window function to improve the transfer characteristics. That is the method employed above. An alternative method called frequency sampling fits the desired frequency response by making a selection of values and calculating the inverse discrete Fourier transform to obtain the filter coefficients. A more sophisticated method uses the Chebyshev alternation theorem to obtain a filter whose maximum error in the pass- and stopbands is minimized. This method yields a filter meeting required specifications with fewer coefficients that the other methods. The design of nonrecursive and recursive filters is outlined in Hamming (1989), where several methods are described, and fuller treatments may be found in Parks and Burrus (1987) and Oppenheim and Schafer (1989).

4.3 Design of Recursive Filters

The design of recursive or IIR filters is more difficult than that of nonrecursive or FIR filters. Several techniques are described in the references at the end of the last section; only one such method will be described below. In this approach, a classical analogue lowpass response is specified. A transformation of variables then converts this to discrete time, and the required filter coefficients are deduced.

4.3.1 *Classical Analogue Filter Response*

There are four classical filter functions which may be used as a basis for digital filter design. They are determined by the manner in which the ideal lowpass filter response is approximated in the pass- and stopbands. If a Taylor series approximation truncated at N terms is applied at $\omega = 0$ and $\omega = \infty$, the result is a Butterworth filter. The Chebyshev (type I) filter uses a minimax approximation across the passband and a Taylor series at $\omega = \infty$. The inverse, or type II, Chebyshev filter uses a Taylor series expansion at $\omega = 0$ and a minimax approximation across the stopband. The elliptic filter involves a minimax approximation in both the passband and the stopband.

It is important to choose the type of filter appropriate to the problem. All four types are optimal in one sense or another. The Butterworth filter is the best Taylor series approximation to the ideal lowpass filter magnitude at both $\omega = 0$ and $\omega = \infty$. The Chebyshev filter gives the smallest maximum error over the passband of any filter having similar Taylor series accuracy at $\omega = \infty$. In a complimentary way, the inverse Chebyshev filter minimizes the maximum deviation from zero in the stopband. The elliptic filter involves four parameters (passband ripple, transition width, stopband ripple and filter order) and for given values of any three minimizes the fourth.

The transfer function for the Butterworth filter of order N has a particularly simple form; for cutoff frequency ω_c it is

$$\left|H(i\omega)\right|^2 = \frac{1}{1+(\omega/\omega_c)^{2N}}. \tag{4.14}$$

As a function of s, the parameter appearing in the Laplace transform, the response of the prototype filter (with cutoff frequency $\omega_c = 1$) is

$$H(s)\cdot H(-s) = \left|H(s)\right|^2 = \frac{1}{1+(-s^2)^N}. \tag{4.15}$$

This function has $2N$ poles evenly spaced around the unit circle. To ensure stability of the filter, the N poles in the left half-plane are selected for $H(s)$; $H(-s)$ will then have the remaining poles. There is a simple formula for the poles $s_k = u_k + iv_k$:

$$u_k = -\sin\left[\frac{(2k+1)\pi}{2N}\right] \quad , \quad v_k = \cos\left[\frac{(2k+1)\pi}{2N}\right] \tag{4.16}$$

with $k = 0, 1, 2, \ldots, N-1$. The Butterworth filter is called *maximally flat*, since the first $2N-1$ derivatives vanish at $\omega = 0$.

4.3.2 *Conversion of Analog to Digital Filter*

There are several methods of deriving a digital transfer function from one of the classical analogue expressions. They all involve some mapping from the s-plane to the z-plane, chosen to preserve properties such as optimality of the filter. The matched $\mathcal{Z}$-transform design procedure is conceptually the simplest (although not the most effective). Poles and zeroes of the transfer function $H(s)$ are mapped directly to poles and zeroes of $H(z)$ by a simple substitution. Consider the inverse transform of a simple pole

$$\mathcal{L}^{-1}\left\{\frac{1}{s-\alpha}\right\} = \exp(\alpha t). \tag{4.17}$$

If this is sampled at intervals Δt and the $\mathcal{Z}$-transform calculated, one has:

$$\mathcal{Z}\{\exp(\alpha n \Delta t)\} = \frac{1}{1-e^{\alpha \Delta t} z^{-1}}. \tag{4.18}$$

Now assume that $H(s)$ is rational and is split into factors $(s-\alpha)$. The relations (17) and (18) suggest the substitution $(s-\alpha) \rightarrow (1-\exp(\alpha\Delta t)/z)$. Allowance for a general cutoff frequency is made by the change $s \rightarrow s/\omega_c$. Combining these, the transformation from prototype analog to general digital transfer function is achieved by

$$(s-\alpha) \longrightarrow (1-e^{\alpha\omega_c \Delta t} z^{-1}) \tag{4.19}$$

mapping the analog pole or zero at $s=\alpha$ to the digital pole or zero at $z=\exp(-\alpha\omega_c\Delta t)$. This produces a transfer function $H(z)$ which is a rational function of $1/z$. The filter coefficients can then be ascertained by comparison of $H(z)$ with (4) above.

The matched $\mathcal{Z}$-transform procedure is very easy to apply, but has the disadvantage that an all-pole analog filter becomes an all-pole digital filter, with no zeros to help shape the frequency response. A more powerful technique is the bilinear transformation; the theory of this procedure can be found in the literature on digital signal processing. The definition is as follows: the mapping from the s-plane to the z-plane is given by

$$s = \frac{1}{\tan(\theta_c/2)}\left[\frac{z-1}{z+1}\right] \tag{4.20}$$

where $\theta_c = \omega_c \Delta t$ is the cutoff frequency. Conversion from a prototype analog filter to a digital filter with cutoff frequency θ_c is implemented by the substitution of (20) into the transfer function $H(s)$ to obtain a function of z.

5. APPLICATION OF DIGITAL FILTERS

To illustrate the potential usefulness of digital filtering, two applications will be described. The first is the use of a nonrecursive filter to construct an initialization scheme. Only the simplest design procedure will be considered; it is possible to increase the efficiency of the scheme significantly by using an optimal filter design. The second application is the combination of a recursive filter with a time-stepping algorithm to remove high frequency components of the solution. The classical Robert-Asselin filter, defined empirically, is examined and compared to some alternative recursive filters designed using the methods of §4.3.

5.1 Initialization using a Nonrecursive Filter

An initialization scheme using a nonrecursive digital filter has been developed by Lynch and Huang (1991) for the HIRLAM model. The value chosen for the cutoff frequency corresponded to a period $\tau_c = 6$ hours. With the time step $\Delta t = 6$ minutes used in the model, this corresponds to a (digital) cutoff frequency $\theta_c = \pi/30$. The coefficients were derived by Fourier expansion of a step-function, truncated at $N = 30$, with application of a Lanczos window, and are given by

$$h_n = \left[\frac{\sin\left(n\pi/(N+1)\right)}{n\pi/(N+1)}\right]\left(\frac{\sin(n\theta_c)}{n\pi}\right).$$

The frequency response was depicted in Fig. 4.1. The central lobe of the coefficient function spans a period of six hours, from $t = -3$ hours to $t = +3$ hours. The summation in (4.1) was calculated over this range, with the coefficients normalized to have unit sum over the span. Thus, the application of the technique involved computation equivalent to sixty time steps, or a six hour adiabatic integration.

The uninitialized fields of surface pressure, temperature, humidity and winds were first integrated forward for three hours, and running sums of the form

$$f_F^\star(0) = \tfrac{1}{2}h_0 f_0 + \sum_{n=1}^{N} h_{-n} f_n, \qquad (5.1)$$

where $f_n = f(n\Delta t)$, were calculated for each field at each gridpoint and on each model level. These were stored at the end of the three hour forecast. The original fields were then used to make a three hour 'hindcast', during which running sums of the form

$$f_B^\star(0) = \tfrac{1}{2}h_0 f_0 + \sum_{n=-1}^{-N} h_{-n} f_n \qquad (5.2)$$

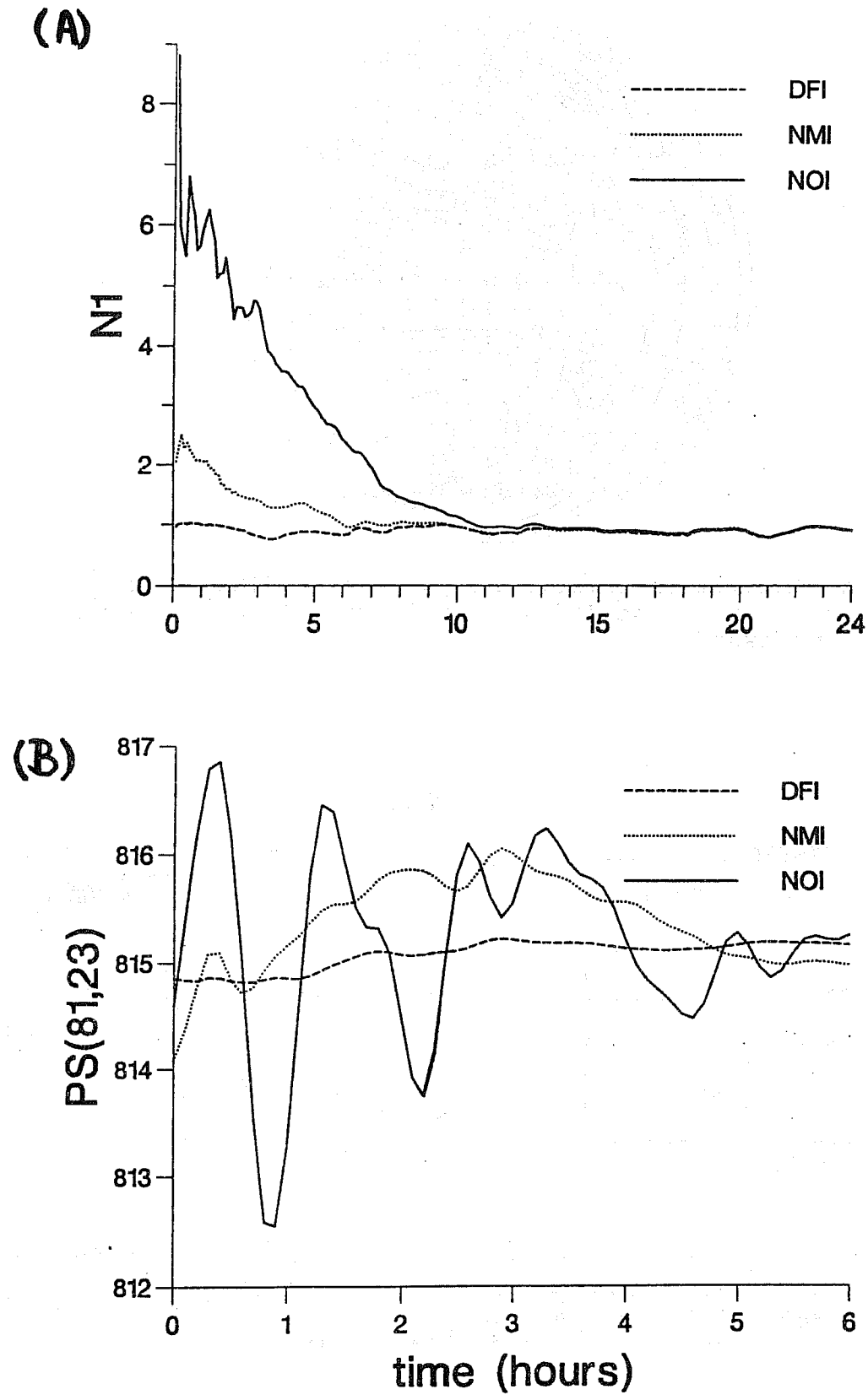


Fig. 5.1. (A) Mean absolute surface pressure tendency, (units are hPa/3h) for a 24 hour forecast starting from uninitialized data (solid curve), from the digitally filtered fields (dashed curve) and from data after normal mode initialization (dotted curve). (B) Surface pressure variation at a gridpoint in the Alps for the first 6 forecast hours (units are hPa).

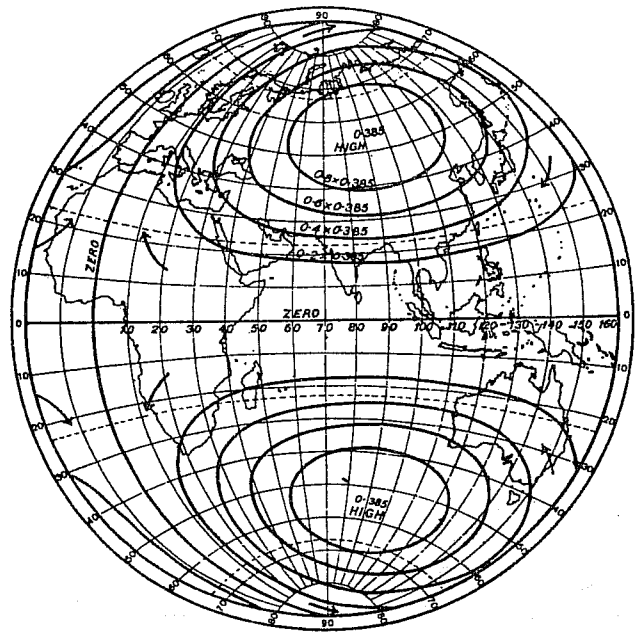


Fig. 5.2. The initial pressure field chosen by Richardson for his barotropic forecast. The initial winds were in geostrophic balance wih the pressure.

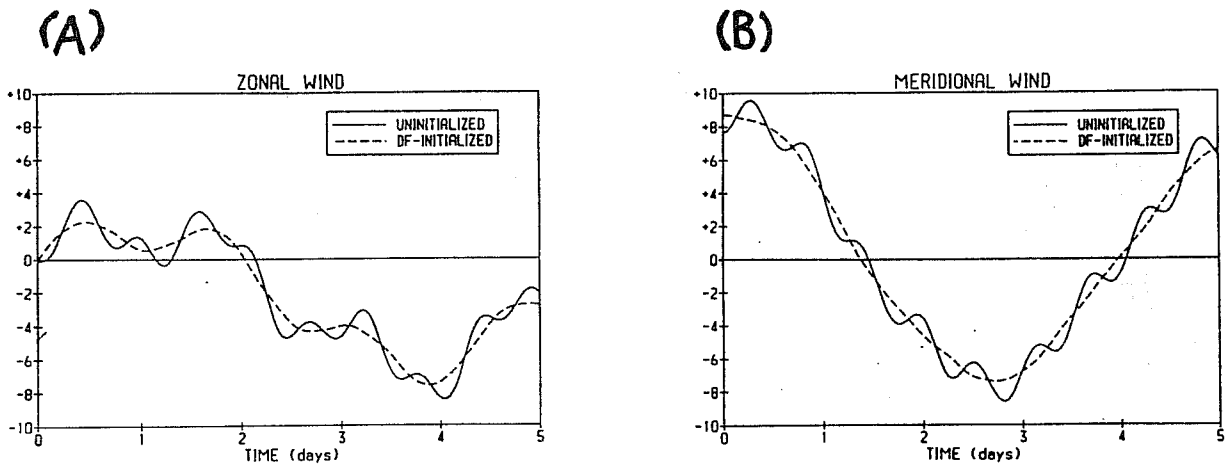


Fig. 5.3. Time evolution of (A) zonal wind and (B) meridional wind at a particular gridpoint for a forecast starting from Richardson's data. Solid curves are for the original data, dashed curves for data initialized by a digital filter.

were accumulated for each field, and stored as before. The two sums were then combined to form the required summations:

$$f^{\star}(0) = f_F^{\star}(0) + f_B^{\star}(0). \tag{5.3}$$

These fields correspond to the application of the digital filter (4.1) to the original data, and will be referred to as the filtered data.

Fig. 5.1 shows the evolution of (a) the areally averaged surface pressure tendency and (b) the surface pressure at a grid-point in the Alps, for a 24 hour forecast starting from uninitialized data (solid curve) and from the digitally filtered fields (dashed curve). For comparison, the result of applying a normal mode initialization is also shown (dotted curve). It is clear that the filtering of the initial fields results in removal of the spuriously large tendencies which are found in the uninitialized run. The difference between the two forecasts was remarkably small: the *rms* surface pressure difference was only $0.07\,hPa$. For further details, see Lynch and Huang (1991).

As a diversion, the digital filtering method was used to initialize the idealized pressure and wind fields which Lewis Fry Richardson (1922) used for his introductory example, in which he integrated the linear shallow water equations. The initial pressure field is depicted in Fig. 5.2; the wind was derived using the geostrophic relationship. Remarkably, the fields chosen by Richardson on considerations of smoothness closely resemble a natural oscillation of the atmosphere: 85% of the energy in these fields is accounted for by a single eigenmode, the so-called five-day wave.

A global barotropic model similar to Richardson's was used to make a five day integration using the idealized data (Lynch, 1992). The variation of the zonal and meridional winds at a point in the English Channel is shown in Fig. 5.3(a) and (b). Superimposed on the predominant variation with a period of five days, higher frequency oscillations can be seen. For the chosen parameter values, the Kelvin wave has a period of 34 hours; this can be seen in the variation of u. The gravest eastward-travelling gravity wave period is 13.5 hours; it can be seen clearly in Fig. 5.3(b) since $v \approx 0$ for the Kelvin wave. To eliminate this "noise", while preserving the Kelvin wave, a cutoff frequency corresponding to a period of 24 hours was chosen and the data initialized with a digital filter. The dashed curves in Fig. 5.3 show the result: the high frequency variation is absent from the initialized forecast.

5.2 Recursive Filters For Use in Time Integration Schemes

Suppose the oscillation equation (3.2) is integrated using a leapfrog timestepping scheme:

$$\frac{u^{n+1} - u^{n-1}}{2\Delta t} + i\omega u^n = 0. \tag{5.4}$$

Seeking a solution $u^n = u^0 \lambda^n$, one immediately finds

$$\lambda^2 + 2i\omega\Delta t\lambda - 1 = 0.$$

If $|\omega\Delta t| < 1$ the roots are unimodular ($|\lambda| = 1$). For $|\omega\Delta t| \ll 1$ they are:

$$\lambda_+ \approx (1 - i\omega\Delta t + \tfrac{1}{2}\omega^2\Delta t^2) \approx \exp(-i\omega\Delta t);$$

$$\lambda_- \approx -(1 + i\omega\Delta t + \tfrac{1}{2}\omega^2\Delta t^2) \approx -\exp(i\omega\Delta t).$$

The first root (λ_+) represents the physical solution; the second (λ_-) is spurious, and represents a computational mode having no counterpart in the original equation (3.2). The leapfrog scheme is *neutral* (for $|\omega\Delta t| < 1$), so the computational mode is undamped.

Robert (1966) introduced a time filter for use in conjunction with the leapfrog scheme. An analysis of the filter was later carried out by Asselin (1972). The filter is defined by the following operation (overbars denote output values):

$$\overline{u^n} = u^n + \tfrac{1}{2}\epsilon\left(u^{n+1} - 2u^n + \overline{u^{n-1}}\right)$$

or, in notation more consistent with that used above (x for input, y for output),

$$y^n = \left(a_0 x^n + a_{-1} x^{n+1}\right) + b_1 y^{n-1} \tag{5.5}$$

where $a_0 = (1 - \epsilon)$, $a_{-1} = \epsilon/2$ and $b_1 = \epsilon/2$. The response function $H(z)$ is

$$H(z) = \frac{(1 - \epsilon) + (\epsilon/2)z}{1 - (\epsilon/2)z^{-1}} \tag{5.6}$$

(this follows from (4.4) above). It has a pole at $z = \epsilon/2$ and zeros at $z = 0$ and $z = 2 - (2/\epsilon)$. For $\epsilon = \frac{2}{3}$, $H(-1) = 0$ and the highest frequency component ($\theta = \pi$ and $\tau = 2\Delta t$) is completely removed. The filter is stable if the pole falls within the unit circle, *i.e.* if $|\epsilon| < 2$.

Next, two alternative recursive filters will be defined, based on the second order Butterworth analog filter. From (4.15), this has a squared amplitude response

$$|H(s)|^2 = \frac{1}{1+s^4} = \left[\frac{1}{s^2+\sqrt{2}s+1}\right]\cdot\left[\frac{1}{s^2-\sqrt{2}s+1}\right] \tag{5.7}$$

The first term has two poles in the left half-plane so, to ensure stability of the filter, $H(s)$ is defined in terms of these:

$$H(s) = \left[\frac{1}{s^2+\sqrt{2}s+1}\right] = \left[\frac{1}{(s-s_1)(s-s_2)}\right] \tag{5.8}$$

where $s_1 = (-1+i)/\sqrt{2}$ and $s_2 = (-1-i)/\sqrt{2}$. If the matched $\mathcal{Z}$-transform approach is used, $H(s)$ is converted to $H(z)$ by means of (4.19), resulting in

$$H(z) = \frac{z^2}{[z-\exp(s_1\theta_c)][z-\exp(s_2\theta_c)]} \tag{5.9}$$

with $\theta_c = \omega_c \Delta t$. This has two poles within the unit circle, and a double zero at the origin. It corresponds to a filter of the form

$$y^{n+1} = a_0 x^{n+1} + \left(b_1 y^n + b_2 y^{n-1}\right)$$

(the coefficients are found by comparison with (4.4)) which is different to the Robert-Asselin filter (5) but requires the same amount of storage.

Alternatively, if the bilinear transformation (4.20) is applied to (5.8), the resulting transfer function is

$$H(z) = \frac{\mu_c^2(z+1)^2}{(z-1)^2+\sqrt{2}\mu_c(z^2-1)+\mu_c^2(z+1)^2} \tag{5.10}$$

where $\mu_c = \tan(\theta_c/2)$. Since there is a (double) zero at $z=-1$, the highest frequency is completely annihilated, which is a desirable feature for a lowpass filter. The transfer function corresponds to a filter of the form

$$y^{n+1} = \left(a_0 x^{n+1} + a_1 x^n + a_2 x^{n-1}\right) + \left(b_1 y^n + b_2 y^{n-1}\right)$$

which requires the storage of more values than before.

The final filter to be considered is obtained by a simple frequency transformation applied to the elementary two-point average

$$y^{n+1} = \tfrac{1}{2}\left(x^{n+1} + x^n\right). \tag{5.11}$$

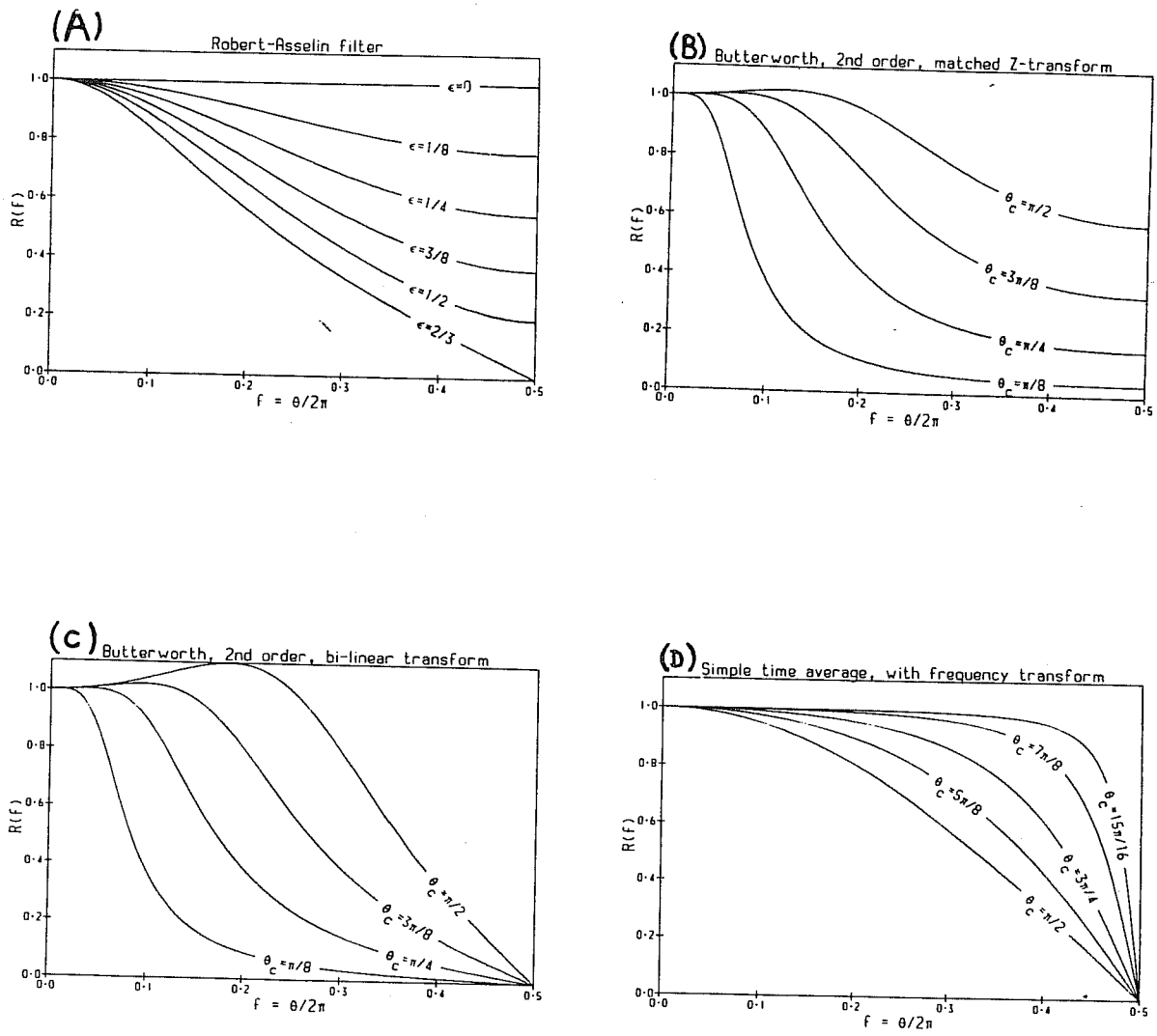


Fig. 5.4. Frequency response (magnitude of transfer function) for four recursive digital filters. (A) Robert-Asselin filter. (B) Second order Butterworth, digitized by the matched Z-transform method. (C) Second order Butterworth, digitized by the bilinear transformation. (D) Simple two-time averaging filter, with frequency transformation. (Details of the design techniques are given in the text).

The transfer function of this nonrecursive filter is

$$H(z) = \frac{z+1}{2z} \tag{5.12}$$

(this follows from (4.4) above) which has a magnitude

$$R(\theta) = \sqrt{\frac{1+\cos\theta}{2}}.$$

The cutoff is at $\theta = \pi/2$, where the gain is $R = 0.707$. This cutoff may be changed to an arbitrary value θ_c by the following transformation (see Strum and Kirk, 1988, *p.*690)

$$z \longrightarrow \frac{z-\alpha}{1-\alpha z} \quad , \quad \alpha = \frac{\sin(\pi/4 - \theta_c/2)}{\sin(\pi/4 + \theta_c/2)} \tag{5.13}$$

The transfer function now assumes the form

$$H(z) = \frac{(1-\alpha)(z+1)}{2(z-\alpha)}. \tag{5.14}$$

Since this has a zero at $z = -1$, the highest frequency is still removed; however, the transformation has changed a nonrecursive filter to a recursive one: (14) has a pole at $z = \alpha$. The form of the filter is

$$y^{n+1} = \left(a_0 x^{n+1} + a_1 x^n\right) + \left(b_1 y^n\right)$$

which is different in form but comparable in complexity to the Robert-Asselin filter.

The magnitude of the response for each of the four filters (6), (9), (10) and (14) is shown in Fig. 5.4. The Robert-Asselin filter (6), shown in Fig. 5.4(a), attenuates high frequencies more than low, and the degree of damping increases as ϵ increases in the range $[0, \frac{2}{3}]$. If the highest frequency is to be completely removed, $\epsilon = \frac{2}{3}$ must be chosen and substantial damping of lower frequencies endured. To a first approximation the damping of this filter increases linearly with frequency (which is a far cry from the ideal step-function). The Butterworth filters (9) and (10), shown in Fig. 5.4(b) and (c), tend to have flatter response curves for the lowest frequencies, which is desirable, and their decrease (roll-off) is steeper than that of the Robert-Asselin filter. However, for nominal cutoff frequency $\theta_c > \pi/4$ they amplify some frequencies (assuming response normalized by $H(0) = 1$); this is generally unacceptable, so that θ_c must be chosen to avoid it. The bilinear transform filter (Fig. 5.4(c)) ensures complete removal of the

highest frequency. Finally, Fig. 5.4(d) shows the response of the frequency transformed time-average filter (14). If the main aim is to eliminate the computational mode, which has a period $\tau = 2\Delta t$, this would seem to be an attractive filter: the highest frequency is completely removed, while a choice of θ_c close to π ensures that damping of lower frequencies is negligible.

The choice of a filter must be governed by the specific requirements of the application. It is not possible to rank filters in an order of suitability to cover all circumstances. The purpose of the above comparison was to indicate the rich variety of options available to the modeller. The Robert-Asselin filter has proved immensely popular, and has been widely used for over 20 years. However, it is not the last word; perhaps some of the alternatives considered above might merit further investigation. It is worth noting that, with the larger timesteps made possible by semi-Lagrangian advection schemes, the cutoff between signal and noise occurs at a higher *digital* frequency (because, for ω_c fixed, $\theta_c \propto \Delta t$). Thus, a filter with a sharp cutoff at a high frequency would appear to be needed; the filter depicted in Fig. 5.4(d) fulfils this requirement.

It was noted in §3.2 that the Laplace transform integration scheme had damping characteristics similar to a Butterworth filter. Using the design methods described above, it is possible to construct a recursive filter with a transfer function approaching a specified ideal (to any degree of approximation, provided the order of the filter is arbitrarily large). Thus, it should be feasible to construct an integration scheme, operating in the time domain, with properties similar to the LT scheme. This might provide a means of achieving the desired filtering in a computationally more economical way.

In the foregoing, only the amplitudes of the transfer functions have been discussed. Since these functions are complex, there is also a phase change induced by the filters. Space prohibits further discussion here; however, it is essential that the phase characteristics of a filter be studied before it is considered for use. Ideally, the phase-error should be as small as possible for the low frequency components which are meteorologically important. The error in the high frequency stopband is unimportant. It is salutary to recall that phase-errors are amongst the most prevalent and pernicious problems faced by the forecaster.

6. CONCLUSION

6.1 General Remarks

Two approaches to filtering have been considered: modification of the governing equations to eliminate HF solutions and numerical integration schemes which selectively dampen these components. The use of filtered equations has the attraction that a solution completely free from gravity-wave noise is guaranteed; damping integration schemes keep such noise under control but do not remove it entirely. This difference would favour filtering the equations. However, no filtered system has yet been shown to leave the rotational motions unscathed, whereas integration schemes can be designed to have negligible attenuation at low frequencies. There are arguments for both approaches, and it is not possible to make a conclusive choice at this stage.

The very notion of eliminating what is called "noise" is open to debate. There is no doubt as to the presence of high frequency motions in the atmosphere, and some evidence suggests that they may have a function in the development of meso-scale systems. If the feedback from HF components to the meteorologically significant motion is found to be important in certain circumstances, the application of filtering may be injudicious. Thus, removal of gravity waves cannot be unequivocally justified; the problem becomes all the more acute as model resolution increases.

There has been active investigation recently of the existence of a *slow manifold.* The state of the atmosphere, or of a model, can be represented by a point in a phase-space, $\mathcal{X}$. The slow manifold, $\mathcal{S}$, is a hypothetical invariant subset of $\mathcal{X}$, of lower dimension than the full space, in which the solution evolves without any high frequency components (Leith, 1980). Invariance implies that a flow which starts in $\mathcal{S}$ remains therein for all time. The concept of a slow manifold is very useful as a descriptive tool in understanding the process of initialization. To what extent it has a more fundamental rôle, is a subject of continuing investigation, and several problems remain to be solved.

Under what circumstances does a slow manifold exist? How can the governing equations be reduced on such a manifold? Is the manifold stable to perturbations? Is it truly invariant? It appears now to be very doubtful if such an invariant subspace exists in the atmosphere. There is evidence that nonlinear interactions between the slow modes inevitably result in the generation of freely propagating gravity waves. This

phenomenon, called *spontaneous emission,* has been discussed recently by McIntyre and Norton (1991). Lorenz and Krishnamurthy (1987) considered a simple nonlinear system, and concluded that no slow manifold exists. Jacobs (1991) presents a proof that there is a slow manifold for this model. Such disagreement serves to demonstrate the complex nature of the problem, even for a relatively simple system of five *o.d.es.* The problem for the full primitive equations is wide open.

6.2 Digital Filtering and Superbalance

When the model equations are separated into slow and fast components, the latter are governed by equation (2.5) above:

$$\frac{d\mathbf{Z}}{dt} + i\Lambda_Z \mathbf{Z} + \mathbf{N}_Z(\mathbf{Y}, \mathbf{Z}) = \mathbf{0}. \tag{6.1}$$

If $|\dot{\mathbf{Z}}| \ll |\Lambda_Z \mathbf{Z}|$ a formal solution can be obtained by an iterative process (Picard):

$$\mathbf{Z}^{(n)} = i\Lambda_Z^{-1} \left[\mathbf{N}_Z + \dot{\mathbf{Z}}^{(n-1)}\right]$$

and is expressible as an infinite series:

$$\mathbf{S}(t) = (i\Lambda_Z^{-1}) \sum_{s=0}^{\infty} (i\Lambda_Z^{-1})^s \frac{d^s \mathbf{N}_Z}{dt^s}.$$

The general solution can then be written in the form

$$\mathbf{Z}(t) = \mathbf{A} \exp(-i\Lambda_Z t) + \mathbf{S}(t). \tag{6.2}$$

The first-order balance condition $\dot{\mathbf{Z}} = \mathbf{0}$ at $t = 0$ (Machenhauer, 1977) yields the solution

$$\mathbf{Z}(t) = -i\Lambda_Z^{-1} \dot{\mathbf{S}}(0) \exp(-i\Lambda_Z t) + \mathbf{S}(t)$$

in which the coefficient of the HF component is small but nonzero. Applying the digital filtering initialization (DFI) technique to (2) annihilates the first *rhs* term and gives

$$\mathbf{Z}(0) = \mathbf{S}(0) \tag{6.3}$$

which implies $\mathbf{A} = \mathbf{0}$. The solution is then given by

$$\mathbf{Z}(t) = \mathbf{S}(t) \tag{6.4}$$

which has no HF component. Taking the s-th time derivative of (1) yields

$$\frac{d^s \mathbf{N}_Z}{dt^s} = -\left[i\Lambda_Z \frac{d^s \mathbf{Z}}{dt^s} + \frac{d^{s+1}\mathbf{Z}}{dt^{s+1}}\right].$$

Substituting this into (3) with the summation truncated at $p-1$ terms gives

$$\mathbf{Z}(0) = -\left(i\Lambda_Z^{-1}\right)\left[\sum_{s=0}^{p-1}\left(i\Lambda_Z^{-1}\right)^s\left(i\Lambda_Z \frac{d^s \mathbf{Z}}{dt^s}\right) + \sum_{s=0}^{p-1}\left(i\Lambda_Z^{-1}\right)^s\left(\frac{d^{s+1}\mathbf{Z}}{dt^{s+1}}\right)\right].$$

This can be simplified to a single term and, when the summation is carried out to its infinite limit, it becomes

$$\lim_{p\to\infty}\left(i\Lambda_Z^{-1}\right)^p \frac{d^p \mathbf{Z}}{dt^p} = 0$$

which is just the *superbalance* condition (Lorenz, 1980).

Despite the superlative appellation, the superbalance condition may not yield a solution free from HF components. Even if the initial conditions (6.3) were used (and this would require an infinite amount of calculation), spontaneous emission would presumably still be expected. There is one sure way to get a slowly evolving function: let $\mathbf{X}(t)$ be the solution starting from the original conditions $\mathbf{X}(0)$, and $\mathbf{X}^\star(t)$ the result of convolution of this solution with the filtering function. $\mathbf{X}^\star$ is obviously *slow* but, alas, it is not generally a solution of the governing equations. Now let $\mathbf{X}_S(t)$ be the solution starting from the superbalance conditions (6.3). It would be nice to know the connection between $\mathbf{X}^\star$ and $\mathbf{X}_S$. For a linear system they are equal, but in general it does not seem that much can be said, even if $\mathbf{X}_S$ is assumed to be slow. The question needs further examination.

Acknowledgements

Thanks to Nils Gustafsson for providing the results of the integrations using the Daley time-stepping scheme, and to Jim Hamilton for assistance in preparation of the graphics.

References

Allen, J.S., J.A. Barth and P.A. Newberger, 1990: On intermediate models for barotropic continental shelf and slope flow fields. Part I: Formulation and comparison of exact solutions. *J. Phys. Oceanog.*, **20**, 1017–1042.

Asselin, R., 1972: Frequency filter for time integrations. *Mon. Weather Rev.*, **100**, 487–490.

Charney, J.G., 1948: On the scale of atmospheric motions. *Geofys. Publ.*, **17**(2), 17pp.

Charney, J.G., 1962: Integration of the primitive and balance equations. Pp 131–152 in *Proc. Intl. Symp. on NWP.* Japanese Meteorological Agency.

Daley, R., 1980: The development of efficient time integration schemes using normal mode models. *Mon. Weather Rev.*, **108**, 100–110.

Daley, R., 1982: A non-iterative procedure for the time integration of the balance equations. *Mon. Weather Rev.*, **110**, 1821–1830.

Daley, R., 1991: *Atmospheric Data Analysis.* Cambridge, 457pp.

Hinkelmann, K.H., Primitive Equations. Pp. 306–375 in *Lectures in numerical short range weather prediction.* Regional Training Seminar, Moscow. WMO No. 297, (1969).

Hamming, R.W., 1989: *Digital Filters*. Prentice-Hall International, 284pp.

Holton, J.R., 1975: *The Dynamic Meteorology of the Stratosphere and Mesosphere.* American Met. Soc., 218pp.

Hoskins, B.J., M.E. McIntyre and A.W. Robertson, 1985: On the use and significance of isentropic potential vorticity maps. *Q. J. Roy. Meteor. Soc.* **111**, 877–946.

Jacobs, S.J., 1991: Existence of a slow manifold in a model system of equations. *J. Atmos. Sci.*, **48**, 893–901.

Kurihara,Y., 1965: On the use of implicit and iterative methods for the time integration of the wave equation. *Mon. Weather Rev.*, **93**, 33–46.

Leith, C.E., 1980: Nonlinear normal mode initialization and quasi-geostrophic theory. *J. Atmos. Sci.*, **37**, 958–968.

Lindzen, R.S., E.N Lorenz and G.W. Platzman, Eds., 1990: *The Atmosphere — A Challange. The Science of Jule Gregory Charney.* American Met. Soc., 321pp.

Lorenz, E.N, 1980: Attractor sets and quasi-geostrophic equilibrium. *J. Atmos. Sci.*, **37**, 1685–1699.

Lorenz, E.N, and V. Krishnamurthy, 1987: On the nonexistence of a slow manifold. *J. Atmos. Sci.*, **44**, 2940–2950.

Lynch, Peter, 1989: The slow equations. *Q. J. Roy. Meteor. Soc.* **115**, 201–219.

Lynch, Peter, 1991 : Filtering integration schemes based on the Laplace and Z transforms. *Mon. Weather Rev.*, **119**, 653–666.

Lynch, Peter, 1992: Richardson's barotropic forecast: a reappraisal. To appear in *Bull. Amer. Meteor. Soc.*, **73**, 1, (January issue).

Lynch, Peter, and Xiang-Yu Huang, 1991: Initialization of the HIRLAM model using a digital filter. Report DM–57, Dept. of Meteor., Stockholm University. (Revised version to appear in *Mon. Weather Rev.*)

Lynch, Peter, and Aidan McDonald, 1991: A multi-level limited-area slow-equation model: application to initialization. *Q. J. Roy. Meteor. Soc.* **116**, 595–609.

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

McIntyre, M.E. and W.A. Norton, 1991: Potential vorticity inversion on a hemisphere. Submitted to *J. Atmos. Sci.*,

McWilliams, J.C. and P.R. Gent, 1980: Intermediate models of planetary circulations in the atmosphere and ocean. *J. Atmos. Sci.*, **37**, 1657–1678.

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

Nayfeh, A.H., 1973: *Perturbation Methods.* John Wiley & Sons, Inc., 425pp.

Oppenheim, A.V. and R.W. Schafer, 1989: *Discrete-Time Signal Processing.* Prentice-Hall Intl., Inc., 879pp.

Parks, T.W. and C.S. Burrus, 1987: *Digital Filter Design.* John Wiley & Sons, Inc., 342pp.

Raymond, W.H., 1991: A review of recursive and implicit filters. *Mon. Weather Rev.*, **119**, 477–495.

Richardson, L.F., 1922: Weather Prediction by Numerical Process. Cambridge Univ. Press, 236 *pp.* Reprinted by Dover Publications, New York, 1965.

Robert, A.J., 1966: The integration of a low-order spectral form of the primitive meteorological equations. *J. Met. Soc. Japan,* 44, 5, 237–245.

Strum, R.D. and D.E. Kirk, 1988: *Discrete Systems and Digital Signal Processing.* Addison-Wesley, 848pp.

Temperton, C., 1988: Implicit normal mode initialization. *Mon. Weather Rev.*, **116**, 1013–1031.