Spectral characteristics of Kalman filter systems with application to the equatorial beta plane.

Roger Daley and Richard Ménard
Canadian Climate Centre and McGill University

1. Introduction

The primary purposes of any data assimilation system are: 1 - to extract the signal from noisy observations - *filtering;* 2 - to interpolate in space and time to obtain estimated values of state variables at each analysis gridpoint at each time step - *interpolation;* and 3 - to reconstruct state variables that are missing in the observational data - *completeness.* Due to its inherent filtering properties, it is possible to regard any data assimilation system as an elaborate filter. For certain special situations, it is possible to eliminate the issues of completeness and interpolation, and examine the filtering properties (i.e. spectral characteristics) of the assimilation scheme in isolation.

The theory is first developed for a very simple model. The model chosen is the one dimensional linear advective-diffusive equation on a regular analysis/forecast grid. The observation network is specified to coincide with the analysis grid, permitting the complete diagonalization of the second moment error statistics of the Kalman filter system. It is then possible to

obtain the complete covariance structure at any time from knowledge of the observation and model error covariances and the initial forecast error covariances. The stationary solutions, convergence rates and asymptotic solutions can be found. The not strictly positive-definite and multivariate cases can also be considered.

The procedure is then applied to the linearized shallow water equations on an equatorial beta plane. The system or model error is constructed from the slow eigenmodes of the model and is based on an expansion in parabolic cylinder functions. The slow (low frequency) and fast (high frequency) error statistics are examined separately for both the optimal and sub-optimal cases.

2. Theory and results for a one dimensional model.

The basic theory follows Daley and Ménard (1993). The standard Kalman filter equations are derived in Daley (1992). The evolution equations of the forecast and analysis error covariance and the equation for an optimal gain matrix are given by equations (2.3, 2.8 and 2.7) of Daley (1992), viz,

$$\mathbf{P}^f_{n+1} = \mathbf{M}_n \mathbf{P}^a_n \mathbf{M}^T_n + \mathbf{Q}_n \quad , \tag{1}$$

$$\mathbf{P}^a_n = [\mathbf{I} - \mathbf{K}_n \mathbf{H}_n]\mathbf{P}^f_n \quad , \tag{2}$$

$$\mathbf{K}_n = \mathbf{P}^f_n \mathbf{H}^T_n [\mathbf{H}_n \mathbf{P}^f_n \mathbf{H}^T_n + \mathbf{R}_n]^{-1} \quad , \tag{3}$$

where $\mathbf{M}_n$ is the (linear) assimilation model at time t_n and $\mathbf{H}_n$ is the (linear) forward interpolation operator from the analysis grid to the observation network. $\mathbf{P}_n^f$ and $\mathbf{P}_n^a$ are the forecast and analysis error covariances which are to be determined, $\mathbf{R}_n$ is the observation error covariance and $\mathbf{Q}_n$ is the model error covariance between t_n and t_{n+1} which must be specified. We have assumed that $\mathbf{Q}_n$, $\mathbf{R}_n$, $\mathbf{M}_n$ and $\mathbf{H}_n$ have been specified correctly and thus $\mathbf{K}_n$, the gain (or weight) matrix, is optimal.

The filtering properties of the Kalman filter system can be investigated by application of the error covariance evolution equations (1-2) and the optimal gain calculation (3) to an observation network which coincides with the analysis grid. When the variable that is forecast is the same variable that is observed then the forward interpolation matrix $\mathbf{H}_n$ becomes the identity matrix. We also assume that the *a priori* statistics $\mathbf{Q}_n$ and $\mathbf{R}_n$ are stationary and that the model $\mathbf{M}_n$ is time independent. This permits the Kalman filter statistics to become stationary. Under these assumptions, equations (1-3) become,

$$\mathbf{P}_{n+1}^f = \mathbf{M}\mathbf{P}_n^a\mathbf{M}^T + \mathbf{Q} \quad , \tag{4}$$

$$[\mathbf{P}_n^a]^{-1} = [\mathbf{P}_n^f]^{-1} + \mathbf{R}^{-1} \quad , \tag{5}$$

upon substitution of (3) into (2).

Define the model to be the one dimension advective/diffusive equation,

$$\frac{\partial h}{\partial t} + U\frac{\partial h}{\partial x} - \nu\frac{\partial^2 h}{\partial x^2} = 0 \quad , \tag{6}$$

where U is a constant advection velocity, ν is a dissipation coefficient and h is (say) the geopotential height. Equation (6) is applied over the one dimensional periodic domain $-\pi a \le x \le \pi a$, and $h(-\pi a,t) = h(\pi a,t)$. The model and analysis gridpoints and the observation network are defined as $x_j = -\pi a + (j-1)\Delta x$, $1 \le j \le J$, where J is the number of gridpoints and $\Delta x = 2\pi a/J$. All matrices in (4-5) are J x J matrices.

Assume the model and observation error covariances are homogeneous. Then it can be shown (following Daley and Ménard, 1993), that for the model (6) on this network that equations (4-5) can be diagonalized. Then if $\mathbf{P}_0^f$ is homogeneous, we can write for each wavenumber p, $1 \le p \le (J-1)/2$,

$$f_{n+1}^2(p) = m^2(p)a_n^2(p) + q^2(p) \quad , \tag{7}$$

$$\frac{1}{a_n^2(p)} = \frac{1}{f_n^2(p)} + \frac{1}{r^2(p)} \quad , \text{ or } \quad a_n^2(p) = \frac{r^2(p)f_n^2(p)}{r^2(p) + f_n^2(p)} \quad , \tag{8}$$

with $m^2(p) = \exp(-2\nu\Delta t p^2/a^2)$. Here $r^2(p)$, $q^2(p)$, $f_n^2(p)$, $a_n^2(p)$ are the spectral variances for wavenumber p in the observation, model, forecast and analysis errors. Substitution of (8) into (7) yields a Ricatti equation,

$$f^2_{n+1}(p) = \frac{m^2(p)r^2(p)f^2_n(p)}{r^2(p) + f^2_n(p)} + q^2(p) \quad . \qquad (9)$$

In equations (7-9), each wavenumber is decoupled and there is no propagation of covariances. In particular, equations (7-9) are independent of U. Propagation of the covariances would occur if either **R** or **Q** were not homogeneous, or if the observation network did not coincide with the analysis grid, or if the model had variable coefficients.

Stationary solutions

Suppose now that equations (7-9) are integrated from some initial choice $f^2_0(p)$. (As noted above, we presume that the initial forecast error covariance $\mathbf{P}^f_0$ is homogeneous). For the moment, assume that we are concerned with a fixed wavenumber p and re-write (9) as,

$$f^2_{n+1} = G(f^2_n) \quad , \text{ with } G(f^2) = m^2r^2f^2/(r^2+f^2) + q^2 \quad , \qquad (10)$$

and references to p have been dropped. Equation (10) can be studied using the theory of iterated maps. This equation has two stationary solutions or fixed points,

$$\bar{f}^2_+ = \alpha - r^2 + \sqrt{\beta} \quad , \qquad (11)$$

$$\bar{f}^2_- = \alpha - r^2 - \sqrt{\beta} \quad , \qquad (12)$$

where $\alpha = 0.5(q^2 + r^2(m^2 + 1))$ and $\beta = \alpha^2 - m^2r^4$. We assume that **R** and **Q** are positive definite and thus, r^2 and q^2 are positive for all p. Since $\mathbf{P}^f$ is also a covariance matrix, f^2 should also be positive. Thus, the positive root (11) is the

physical solution, while (12) is non-physical.

Stability properties

Denote a stationary solution of equation (10) as $\bar{f}^2$. Then, $\bar{f}^2$ will be stable/unstable to small perturbations if $|dG/d\bar{f}^2|$ is less than/greater than one. The stability analysis shows that the physical solution (11) is always stable, while the unstable solution (12) is always unstable (assuming q^2, $r^2 > 0$).

Figure 1 is a map of the inviscid case ($m^2 = 1$) with q^2 and r^2 both specified to be equal to 2. The abscissa is $G(f^2)$ and the ordinate is f^2 with $-15 \le f^2 \le 15$. The curved solid line is $G(f^2)$ and the singularity in (12) at $f^2 = -2$ is obvious. The two fixed points occur at the intersection of $G(f^2)$ and the dashed straight line.

An interesting limit occurs when $\beta \to 0$; the two stationary solutions (11-12) coalesce and the fixed points **U** and **S** in Figure 1 merge. In this case,the fixed point becomes neutral. This can occur for the inviscid , perfect case ($m^2 = 1$, $q^2 \to 0$) or for the case ($r^2 \to 0$, $q^2 \to 0$).

Also shown on Figure 1 are two trajectories. They are calculated from the time dependent solutions,

$$f_n^2 = \frac{[f_0^2 - \bar{f}_+^2][\bar{f}_+^2(c_+^n - 1) - 2\sqrt{\beta}\, c_+^n] - 2\sqrt{\beta}\, \bar{f}_+^2}{[f_0^2 - \bar{f}_+^2][c_+^n - 1] - 2\sqrt{\beta}}. \quad (13)$$

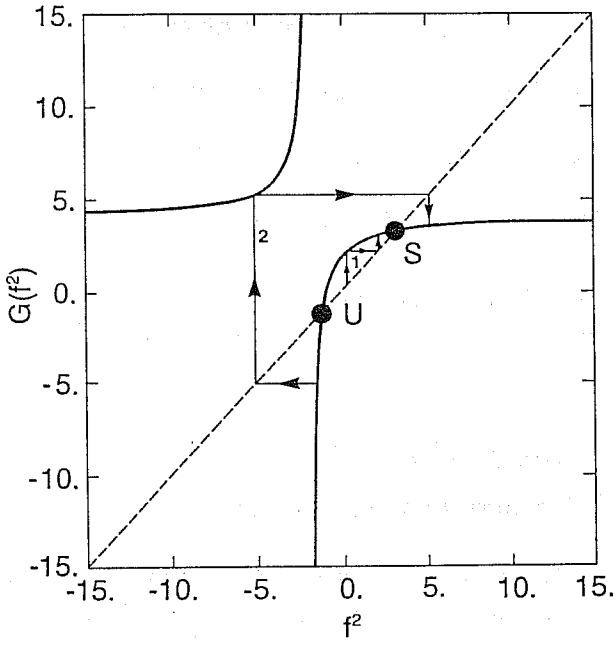

Fig. 1 Map of the inviscid case ($m^2 = 1$) with q^2 and r^2 both specified to be equal to 2. The abscissa is $G(f^2)$ and the ordinate is f^2 with $-15 \leq f^2 \leq 15$

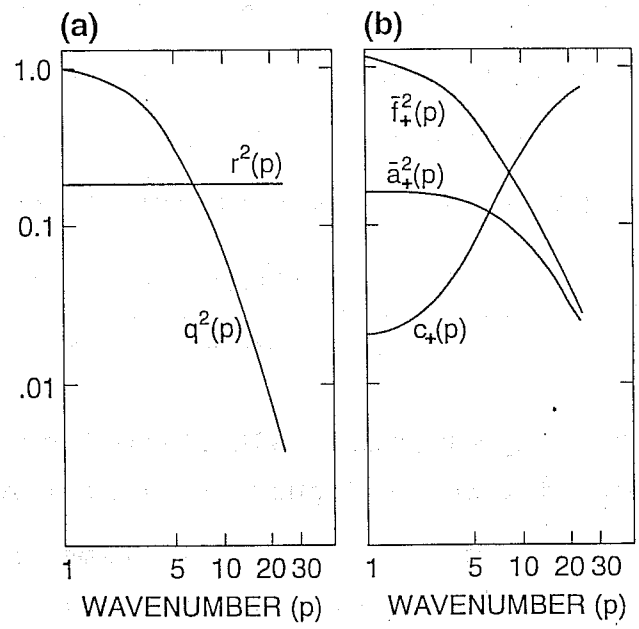

Fig. 2 a)g raph of $q^2(p)$ and $r^2(p)$ as a function of wavenumber p, b) graph of $\bar{f}_+^2(p)$, $\bar{a}_+^2(p)$ and $c_+(p)$ as a function of wavenumber p

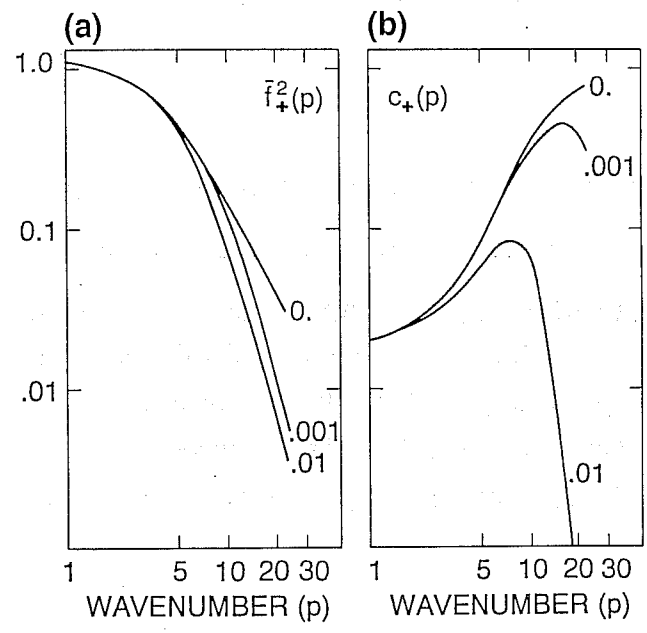

Fig. 3 a) graph of $\bar{f}_+^2(p)$ for v = 0, 3 x 10^5 and 3 x 10^6 m^2s^{-1}, b) graph of $c_+(p)$ for v = 0, 3 x 10^5 and 3 x 10^6 m^2s^{-1}

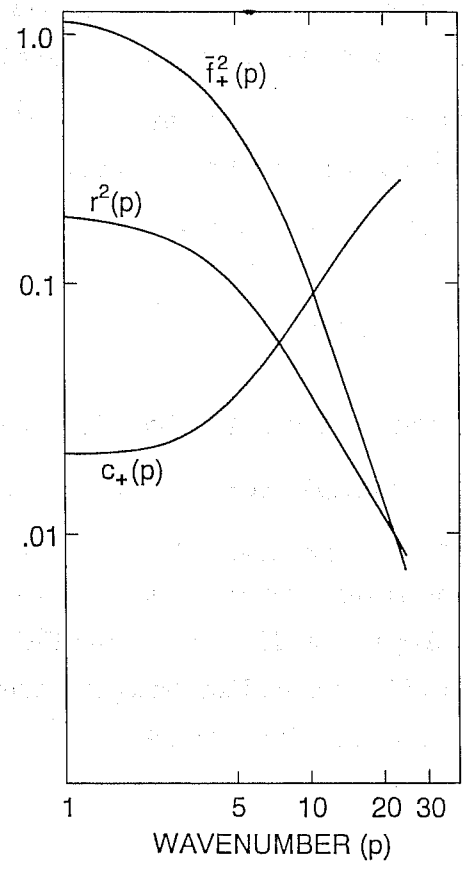

Fig. 4 Graph of r^2, $\bar{f}_+^2$, c_+ as a function of wavenumber p for α = 2500 km g = 9.8 ms^{-2}, $\bar{f}$ = $10^{-4}s^{-1}$, $E_h^q = E_h^r$ = 10m and the wind observation error E_v^r = $2ms^{-1}$

where $c_+ = (\alpha - \sqrt{\beta})/(\alpha + \sqrt{\beta})$. $0 \leq c_+ < 1$ is the convergence rate toward stationarity - larger values indicating rapid convergence and small values indicating slow convergence. When $c_+ = 1$, (13) is undefined. A form derivable by L'Hôpital's Rule is,

$$f_n^2 = \frac{[f_0^2 - \alpha + r^2][r^2 + (n+1)(\alpha - r^2)] + \alpha[\alpha - r^2]}{n[f_0^2 - \alpha + r^2] + \alpha} \quad . \quad (14)$$

Two trajectories (calculated from 13 with $m^2 = 1$, $q^2 = r^2 = 2$) are plotted on Figure 1 as a series of perpendicular lines marked with arrows. Each timestep consists of one horizontal and one vertical segment. Trajectory 1 commences with $0 < f_0^2 < \bar{f}_+^2$ and converges to the fixed point $\bar{f}_+^2$. The other trajectory (2) starts with a non-physical value $f_0^2 < 0$. However, we note that trajectory 2 also converges to $\bar{f}_+^2$, but it crosses the singularity twice. All trajectories are attracted towards the stable fixed point except those beginning at the unstable fixed point, or for which $f_n^2 = -2$ for some n. However, if $f_n^2 > 0$, for some time t_n, then f^2 will remain greater than zero for all later time. Thus, the long term behaviour of the filter does not lead to negative variances.

Error spectra

We discuss results for the case a = 2500 km, which means that the total channel length $2\pi a \approx 15{,}700$ km. The number of gridpoints J = 49, corresponding to a gridlength $\Delta x = 320$ km. Because of the homogeneity assumption, the results are independent of $U\Delta t$ but $\nu\Delta t/a^2$ will be specified subsequently. In this section we will only consider height observations (the univariate case). We assume that the model $q^2(p)$ and observation $r^2(p)$ error spectra have the functional form,

$$(1 + p^2\ell_q^2)^{-2} \quad \text{and} \quad (1 + p^2\ell_r^2)^{-2} \quad , \qquad (15)$$

respectively. Here ℓ_q and ℓ_r are specified length scales for the model and observation error. The model and observation error variances are denoted $(E_h^q)^2$ and $(E_h^r)^2$ respectively.

Inviscid case

We first consider the inviscid case, $\nu = 0$ or $m^2(p) = 1$, for all p. In Figure 2a, we illustrate $q^2(p)$ and $r^2(p)$ as a function of wavenumber p. In this case, $E_h^q = E_h^r = 10$ m, $\ell_q = a/6 = 417$ km and $\ell_r = 0$. Thus, the observation error spectrum is white, consistent with spatially uncorrelated observation error. The ordinate is logarithmic and has been defined such that $q^2(0) = 1$. Since $E_h^q = E_h^r$, the sum (over p) of the variances is the same for the model and observation errors.

In Figure 2b, in the same format, are shown corresponding curves for $\bar{f}_+^2(p)$ and $\bar{a}_+^2(p)$ as obtained from (11 and 8). We note from (13 and 10), that in the inviscid case, as $r^2(p) \to 0$, then $\bar{f}_+^2(p) \to q^2(p)$ and $\bar{a}_+^2(p) \to r^2(p)$. At the other limit, as $q^2(p) \to 0$, equations (9, 10 and 13) show that $\bar{a}_+^2(p) = \bar{f}_+^2(p) \to \sqrt{q^2(p)r^2(p)}$

Also plotted in Figure 2b is $c_+(p)$ for this case. It can be seen that when $m^2(p) = 1$, as $r^2(p) \to 0$, then $c_+(p) \to 0$. Conversely,, when $q^2(p) \to 0$, then $c_+(p) \to 1$. This→ behaviour is evident in Figure 2b. Convergence is very slow for an inviscid, perfect model.

The rate of convergence to the stationary solution can be considered further by examination of equations (15-16). For $c_+ < 1$, (equation 15), the distance from the fixed point (for n large), is proportional to c_+^n . On the other hand, for the inviscid, perfect model $c_+ = 1$ and equation (16) becomes,

$$f_n^2 = f_0^2 r^2 / (n f_0^2 + r^2) \quad , \qquad (16)$$

which means $f_n^2 \rightarrow r^2/n$ as $n \rightarrow \infty$. This implies a very small change in forecast error variance between data insertions at big n. The behaviour at small n may differ, however. If $f_0^2 >> r^2$, then (16) implies $f_1^2 \approx r^2$, which means a large initial reduction in error, followed by r^2/n behaviour as n increases. However, if $f_0^2 << r^2$, then $f_1^2 \approx f_0^2$ and there is relatively little change in forecast error for small n.

Figure 2 indicates that for univariate height data assimilation with an inviscid model, uncorrelated observation error and a red model error spectrum; forecast and analysis error spectra are much whiter than the model error. Moreover, at smaller scales, convergence to the stationary state is rather slow, which means that these scales tend to "remember" the initial condition $f_0^2(p)$ much longer. If E_h^q is fixed under these conditions, the asymptotic limits (discussed above) suggest that decreasing E_h^r will tend to redden the forecast error spectrum, while increasing E_h^r would whiten it. It can also be seen that as the resolution of the model increases (i.e. small Δx), there would be smaller and smaller scales in the problem and convergence in these scales would be even slower.

Effect of dissipation

Now consider the effect of dissipation (for the univariate case).

We again choose $\ell_r = 0$ (spatially uncorrelated observation error) and specify three values for the dissipation parameter, $\nu\Delta t/a^2 = 0$, 0.001 and 0.01. If Δt = 6 hours and a = 2500 km, then these values correspond to $\nu = 0$, 3×10^5 and 3×10^6 m^2s^{-1}.

In Figure 3a, we plot $\bar{f}_+^2(p)$ for these choices of ν as a function of p. It is clear that the addition of dissipation reddens the forecast error spectrum. When $m^2(p) < 1$, equation (11 and 8) show that as $r^2(p) \to 0$, then $\bar{f}_+^2(p) \to q^2(p)$ and $\bar{a}_+^2(p) \to r^2(p)$ as in the inviscid case. However, when $q^2(p) \to 0$, then $\bar{f}_+^2(p) = \bar{a}_+^2(p) \to q^2(p)[1 - m^2(p)]^{-1}$. It is clearly evident from Figure 3a, that as p gets large, $m^2(p) \to 0$ and $\bar{f}_+^2(p) \to q^2(p)$.

In Figure 3b is plotted $c_+(p)$ for the same choices of ν. It is clearly evident from Figure 3b that positive ν speeds up convergence in the shorter waves. From (13) it can be seen that as $r^2(p) \to 0$, then $c_+(p) \to 0$ as in the inviscid case. However, as $q^2(p) \to 0$, $c_+(p) \to m^2(p)$. For a perfect, viscous model, the stronger the dissipation, the more rapdily will the system converge towards its fixed point.

The multivariate case

In the derivation of equations (7-8), it was assumed that the observed and forecast variables were both h. If h is the height field, then this derivation can be generalized to include wind observations as well. Suppose that at every observation

station/analysis gridpoint j, $1 \leq j \leq J$ we observe the geostrophic wind v as well as the height field. We define the geostrophic wind as,

$$v = \frac{g}{\tilde{f}} \frac{\partial h}{\partial x} \quad , \tag{17}$$

where g is the gravitational constant and $\tilde{f}$ is the (constant) Coriolis parameter. (not to be confused with the forecast error variance). On this grid, we can use Fourier decomposition to write equation (17) in matrix form, $\mathbf{v} = \mathbf{Dh}$, where $\mathbf{h}$ is a vector of heights and $\mathbf{v}$ is a column vector of length J of geostrophic winds at the gridpoints and the matrix $\mathbf{D}$ is the the discrete derivative operator coresponding to (17).

Equation (4) is the same in the multivariate case, but (5) is,

$$[\mathbf{P}_n^a]^{-1} = [\mathbf{P}_n^f]^{-1} + \mathbf{R}_h^{-1} + \mathbf{D}^T\mathbf{R}_v^{-1}\mathbf{D} \quad , \tag{18}$$

where $\mathbf{R}_h$ and $\mathbf{R}_v$ are observation error covariance matrices for the height and (geostrophic) wind respectively. Assuming that $\mathbf{R}_h$ and $\mathbf{R}_V$ are homogeneous, equations (5 and 18) can be diagonalized as in section 2 to produce for each wavenumber p, equation (7) and a generalized version of (11), viz,

$$\frac{1}{a_n^2(p)} = \frac{1}{f_n^2(p)} + \frac{1}{r_h^2(p)} + \frac{p^2 g^2}{\tilde{f}^2 a^2 r_v^2(p)} \quad , \tag{19}$$

where $r_h^2(p)$ and $r_v^2(p)$ are the spectral variances for wavenumber p in the height and wind observation errors.

We now consider experiments of the same type discussed previously for the multivariate system. We assume that the height and wind observation errors are not spatially (or mutually) correlated

(i.e. $r_h^2(p)$ and $r_v^2(p)$ are independent of p). We will also assume that the model is inviscid ($m^2(p) = 1$) and that $q^2(p)$ is given by equation (15), with $\ell_q = a/6$ as before. We specify a = 2500 km, g = 9.8 ms^{-2}, $\tilde{f} = 10^{-4}$ s^{-1}, $E_h^q = E_h^r = 10$ m and the wind observation error $E_v^r = 2$ ms^{-1}.

This case is plotted in Figure 4, in the same format as Figures 2 and 3. We plot the spectra of r^2 ,$\bar{f}_+^2$ and c_+ as a function of wavenumber p. At p = 0, $r^2(p)$ is the same as in Figure 2a, but as p increases, the wind information reduces $r^2(p)$ with respect to the univariate case. Consequently, $\bar{f}_+^2(p)$ is the same in the univariate and multivariate cases at p = 0, but decreases more rapidly in the multivariate case as p increases. $c_+(p)$ is smaller at high wavenumber in Figure 4 than in Figure 2b, indicating that the addition of wind information speeds up the convergence to the stationary state. Examination of equation (19) shows that as $p \to 0$, $r^2(p) \to r_h^2(p)$ and as $p \to \infty$, $r^2(p) \to \tilde{f}^2 a^2 r_v^2(p) / g^2 p^2$, which is consistent with Figure 4.

It might be noted that if there are no height observations ($r_h^2 \to \infty$), then $r^2(0)$, $\bar{f}_+^2(0)$ and $\bar{a}_+^2(0)$ all become singular in both the viscous and inviscid cases. This is because wind observations alone, do not provide any information about wavenumber zero, and therefore the forecast and analysis errors will increase without bound.

The problem of negative variances

We have assumed that $q^2(p)$ and $r^2(p)$ are positive for all p, which presumes that **R** and **Q** are derived from an underlying stochastic system (i.e. these matrices are positive definite) However, $q^2(p)$ and $r^2(p)$, while positive, are often small. Thus, in numerical practice, *there exists the possibility that* **Q** *and/or* **R** *may have small,negative eigenvalues.* Under these conditions, it is possible that the matrix $\mathbf{H}_n\mathbf{P}^f_n\mathbf{H}^T_n + \mathbf{R}_n$ of equation (3) might not be strictly positive definite (i.e. have one or more very small or negative eigenvalues) causing problems with the analysis equation. It is well known that this can also cause problems in statistical interpolation practice.

We will examine the effect of small,negative eigenvalues on the performance of the Kalman filter using the spectral equations derived earlier. Suppose that $q^2(p)$ and/or $r^2(p)$ is less than zero for some p. In that case, there exists the possibility that the expression under the square root root sign in (11) is less than or equal to zero and there is no stationary solution. This can be examined by finding the roots of $[q^2 + (m^2-1)r^2]^2 + 4r^2q^2$. The roots of this expression respond to the straight lines,

$$q^2 = -r^2\left[|m| \pm 1\right]^2 \quad , \qquad (20)$$

on the q^2/r^2 plane. Along these lines, the two stationary solutions coalesce into a single neutral solution.

In Figure 5, we use (20) to show the regions where there is no

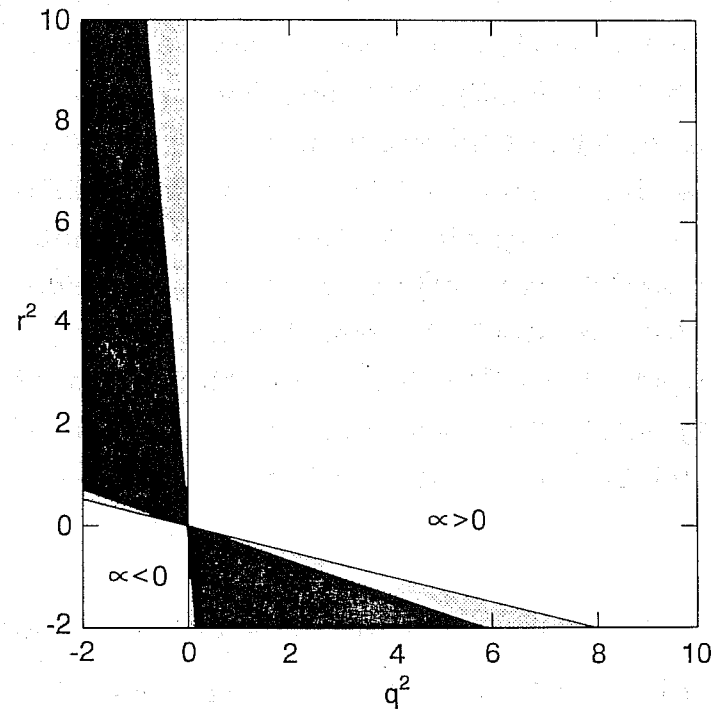

Fig. 5 Graph of eq. (20). Areas shown in grey and black are areas where there is no stationary solution.

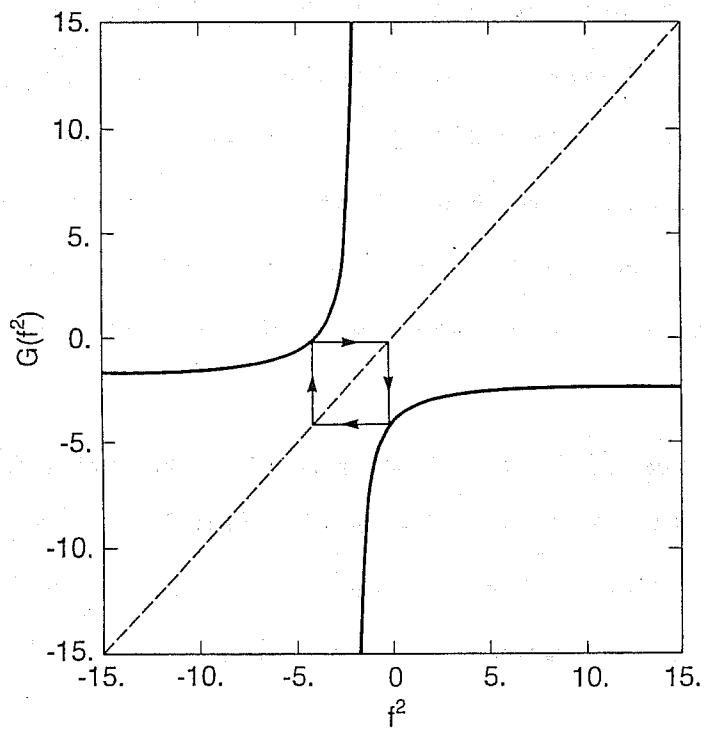

Fig. 6 Map of $G(f^2)$ for the inviscid case ($m^2 = 1$) with $r^2 = 2$ and $q^2 = -4$ in the same format as Fig. 1.

stationary solution as a function of q^2 (abscissa) and r^2 (ordinate). r^2 and q^2 both run from -2 to 10 in arbitrary units. In the case $m^2 = 1$ (inviscid), the two roots are $q^2 = 0$ and $q^2 = -4r^2$. In this case, the area where there is no stationary solution includes the areas shown in gray *and* black. It can be seen that while small negative values of r^2 do not prevent convergence to a stationary solution; there is *generally no stationary solution for small negative values of* q^2. When $m^2 = 1$, $q^2 > 0$ and $-q^2[|m|+1]^{-2} = -q^2/4 < r^2 < 0$, then both stationary solutions (11-12) yield positive values of f^2 and are thus physically realizable. Only (11) is stable, however.

In the case $m^2 = 1/2$, the region of no stationary solution is shown in black, and is clearly smaller than the corresponding region for the inviscid case. Viscosity aids the convergence to a stationary state for small negative values of q^2.
When the observation error is spatially uncorrelated ($r^2(p)$ independent of p) and the model error spectrum is red (as in Figure 2), there exists the possibility that $q^2(p)$ may be negative at large p. The addition of a small amount of viscosity to the model appears to assist the convergence. It should be noted, however, that when $r^2 > 0$ and $-r^2[|m|-1]^2 \leq q^2 \leq 0$, both (11-12) yield non-positive values, and are non-physical.

We have not attempted a detailed analysis for the situation $\beta < 0$ (no fixed points). However, we will discuss a few special cases. Consider first, the inviscid case ($m^2 = 1$) with $r^2 = 2$ and $q^2 = -4$. In Figure 6, we show the map $G(f^2)$ for this case in the

same format as Figure 1. The singularity at $f^2 = -2$ is evident, but this time there are no intersections with the dashed line and, hence, no fixed points. We can examine this case using (12). Define, $\phi = \tan^{-1}[\sqrt{-\beta}/\alpha]$. Then (16) is written as,

$$f_n^2 = \frac{A_1\cos(n\phi) + B_1\sin(n\phi)}{A_2\cos(n\phi) + B_2\sin(n\phi)}, \qquad (21)$$

where A_1, B_1, A_2 and B_2 are constants. If,

$$\phi = \tan^{-1}[\sqrt{-\beta}/\alpha] = \ell_1\pi/\ell_2, \qquad (22)$$

where ℓ_1, ℓ_2 are integers with $\ell_2 \geq 2|\ell_1| \geq 0$, then equation (21) is periodic with period $\ell_2\Delta t$. Thus, if ϕ/π is rational, the solution is periodic.

$m^2 = 1$, $r^2 = 2$ and $q^2 = -4$ implies this case must have a periodicity of $\ell_2 = 2$. This can be confirmed by noting that $G(G(f^2)) = f^2$. The period of (22) depends only on m^2, q^2 and r^2 and does not depend on the initial value f_0^2. A sample trajectory for this case is also illustrated in Figure 6.

Further examination of (22) suggests that when $\beta < 0$ (black and gray regions of the q^2/r^2 diagram of Figure 5), the line $\alpha = 0$ or $q^2 = -r^2(m^2+1)$ corresponds to periodic solutions of periodicity 2. We note that as $\beta \to 0$ from the negative side, there will be increasingly closely packed lines of periodic solutions with increasingly long periods.

An example is shown in Figure 7, which plots f_n^2 as a function of n for 100 iterations commencing with $f_0^2 = 1$. Here $m^2 = 1$, $r^2 = 2$

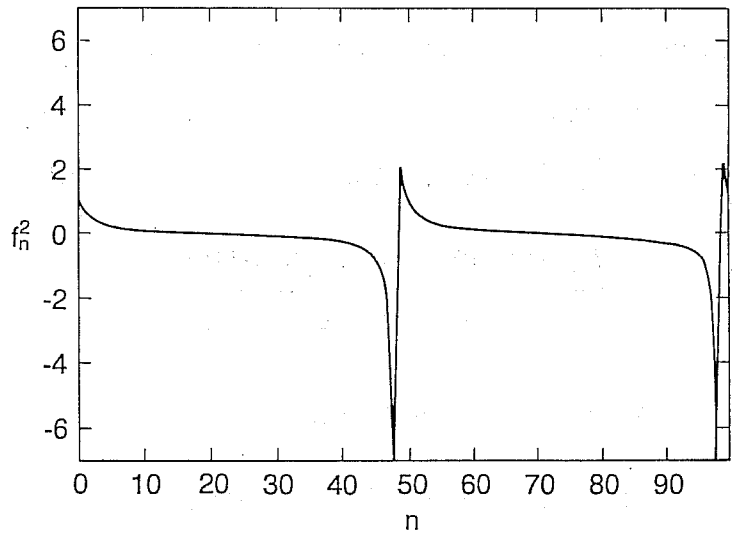

Fig. 7 Graph of f_n^2 as a function of n for 100 iterations commencing with $f_0^2 = 1$. Here $m^2 = 1$, $r^2 = 2$ and $q^2 \approx 0.0079$ has been chosen to make $\phi = \pi/50$

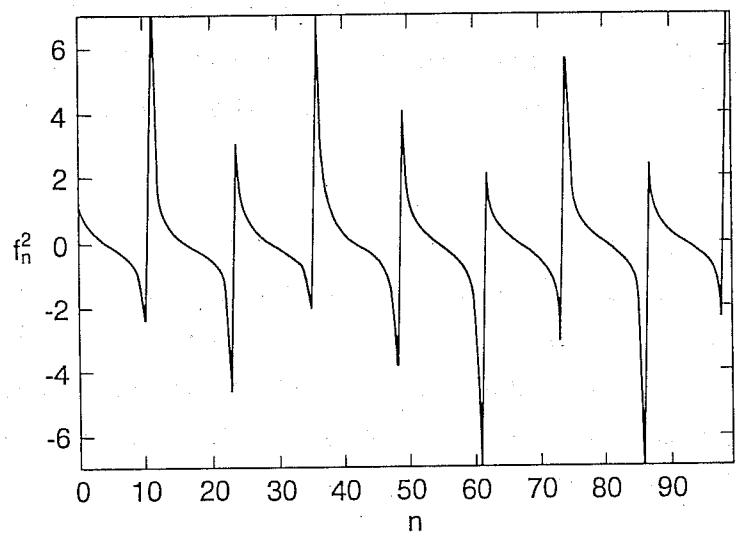

Fig. 8 Graph of f_n^2 for $m^2 = 1$, $r^2 = 2$ and $q^2 \approx -0.1244$ has been chosen to make $\phi = \pi/4\pi = 0.25$

and $q^2 \approx -0.0079$ has been chosen to make $\phi = \pi/50$. From (22), we would expect a periodic solution of period 50 . In this case, the solution hovers near the fixed point for $\beta = 0$ (i.e. $f_+^2 = 0$) and then dramatically shifts away and returns. It should be noted, that even though solutions of the form (21) are periodic, they may contain singularities for certain choices of f_0^2.

When ϕ/π is irrational, then the trajectories are quasi-periodic (and may become singular for certain choices of f_0^2). An example is plotted in Figure 8. In this case $m^2 = 1$, $r^2 = 2$ and $q^2 \approx -0.1244$ has been chosen to make $\phi = \pi/4\pi = 0.25$. It can be seen that the oscillation is almost periodic with a "period" of approximately $12\Delta t$.

Negative values of $r^2(p)$ or $q^2(p)$ for some p, will have an effect on the real space Kalman filter system (4-5). It was verified experimentally that the type of oscillations discussed in this section occurred in equations (4-5) as well.

3. Data assimilation on the equatorial beta plane.

In operational global atmospheric data assimilation systems, analyses are routinely produced for the extratropics of both hemispheres as well as the tropics. However, almost all data assimilation theory has been developed for the extratropics. At this time, there have been no tropical atmospheric applications of the Kalman filter theory.

In this section we describe the application of Kalman filter theory to the shallow water equations on an equatorial beta plane. The theory is two dimensional, but we consider special observation networks which permit the diagonalization of the system. In this way, the complete time and space behaviour of the second moment error statistics can be determined following the ideas of Daley and Ménard(1993). We first derive the low frequency second moment statistics for the equatorial beta plane and then show results for the optimal and sub-optimal filters. See also Daley (1993).

Derivation of the model error for the equatorial beta plane

The theory is based on the normal modes of the linearized shallow water equations on an equatorial beta plane. The longitude and latitude are denoted λ and ϕ respectively and earth's radius a. We define an east/west coordinate $x = a\lambda$ and a north/south coordinate $y = a\phi$. Then,

$$\frac{\partial u}{\partial t} - \beta y v + g\frac{\partial h}{\partial x} = 0 , \qquad (23)$$

$$\frac{\partial v}{\partial t} + \beta y u + g\frac{\partial h}{\partial y} = 0 , \qquad (24)$$

$$\frac{\partial h}{\partial t} + H\left[\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right] = 0, \qquad (25)$$

where u,v are the eastward and northward velocity components and h is the height of the free surface (with mean removed) and t is time. Here, g is the gravitational constant, $\beta = 2\Omega/a$ is the

(constant) beta parameter and H is the mean height of the free surface (or equivalent depth).

The equatorial beta plane equations (23-25) have solutions corresponding to Kelvin waves, Rossby waves and gravity (Poincaré) waves. Solutions that vanish as $y \to \pm\infty$ are of the form $v_k^n(x,y,\omega)\ \exp(-i\omega t)$, where,

$$v_k^n(x,y,\omega) = N_k^n(\omega)\ D_n(y/a_e)\ \exp(ikx). \qquad (26)$$

Here k is the zonal (x) wavenumber and $i = \sqrt{-1}$. n is a non-negative integer meridional (y) wavenumber and,

$$D_n(\sqrt{2}\ z) = 2^{-n/2}\ H_n(z)\ \exp(-z^2/2)\ , \qquad (29)$$

is the parabolic cylinder function of order n. $H_n(z)$ is a Hermite polynomial of order n (see Daley, 1993). ω is the frequency and is given by the dispersion relation,

$$(\omega/c)^2 - k^2 - \beta k/\omega = (2n+1)\beta/c\ . \qquad (28)$$

arbitrary velocity and height fields v(x,y), u(x,y) and h(x,y) can be expanded in terms of the structure functions $v_k^n(x,y,\omega_m)$, $u_k^n(x,y,\omega_m)$ and $h_k^n(x,y,\omega_m)$. We will assume that the expansion has the same number of degrees of freedom in the east/west and north/south directions. Thus define K even, as the truncation limit and write,

$$v(x,y,t) = \sum_{k=-K/2}^{K/2} \sum_{n=0}^{K} \sum_{m=1}^{3} \alpha_\nu(t)\ v_k^n(x,y,\omega_m)\ , \qquad (29)$$

and similarly for u(x,y,t) and h(x,y,t). Here $\alpha_\nu(t)$ is a complex expansion coefficient which is common to the v, u and h expansions and $\nu(n,m,k)$ is a modal index.

Our aim is to construct model error spatial covariance structures for the equatorial beta plane. The spatial covariance structures will be constructed using the ideas of Phillips. There are three basic assumptions.

$$< \alpha_\nu(t)\alpha_{\nu'}(t)^* > = 0 \quad \text{unless} \quad \nu = \nu'. \tag{30}$$

$$< \alpha_\nu(t)\alpha_\nu(t)^* > = 0 \text{ unless } \omega_\nu \leq \omega_c \text{ (cut-off)}. \tag{31}$$

$< \alpha_\nu(t)\alpha_\nu(t)^* >$ is independent of t and ν, for all modes which satisfy (31). (32)

These assumptions then allow us to construct covariance, correlation and variance functions for the domain (x,y). Then define $\gamma_\nu = <\alpha_\nu\alpha_\nu^*>$ and $v_\nu(x,y) = v_k^n(x,y,\omega_m)$ and similarly for $u_\nu(x,y)$ and $h_\nu(x,y)$, with corresponding eigenfrequency ω_ν. We will use as an illustration, the covariance between v and h. The v/h covariance can be written,

$$c_{vh}(x,x',y,y') = <v(x,y)h(x',y')> = \sum_\nu v_\nu(x,y)\ \gamma_\nu\ h_\nu(x',y'), \tag{33}$$

for all $\omega \leq \omega_c$. We can define variances $C_{vv}(x,x,y,y) = <v^2(x,y)>$ and $C_{hh}(x,x,y,y) = <h^2(x,y)>$ and hence correlations,

$$c_{vh}(x,x',y,y') = <v(x,y)h(x',y')>/[<v^2(x,y)><h^2(x,y)>]^{1/2}, \tag{34}$$

In a similar manner we can construct C_{vv} , C_{vu} , C_{uv} , C_{uu} , C_{uh} , C_{hv} , C_{hu} and C_{hh} and the corresponding variances and correlations.

We will now examine the tropical low frequency (slow mode) covariances, variances and correlations obtained from (33). The use of the present parabolic cylinder function expansion has the advantage that the results can be presented in a way which is valid for all equivalent depths. The disadvantage is that the results are not very reliable far from the equator. In general,the covariances are homogeneous in the x direction, but not in the y direction. This is to be expected, because the coefficients in (23-25) vary in y, but not in x. Thus, we show variances as a function of y only. The variances are symmetric with respect to the equator (y=0). $\langle v^2 \rangle$, $\langle u^2 \rangle$ and $g\langle h^2 \rangle/H$ in units of $m^2 s^{-2}$ as a function of y/a_e are plotted in Figure 9. Two case are shown for the height field - the solid curve includes all Kelvin modes which have frequencies less than or equal to the cut-off frequency and the dashed curve treats all Kelvin modes as fast modes. The omission or inclusion of these Kelvin modes had a negligible effect on the wind variance.

Correlations can be calculated from (34). Although these correlations can be determined for any set of points, we will display only correlations with equatorial points. In Figure 10 is shown c_{vv} (panel a), c_{uu} (panel b) and c_{vu} (panel c).

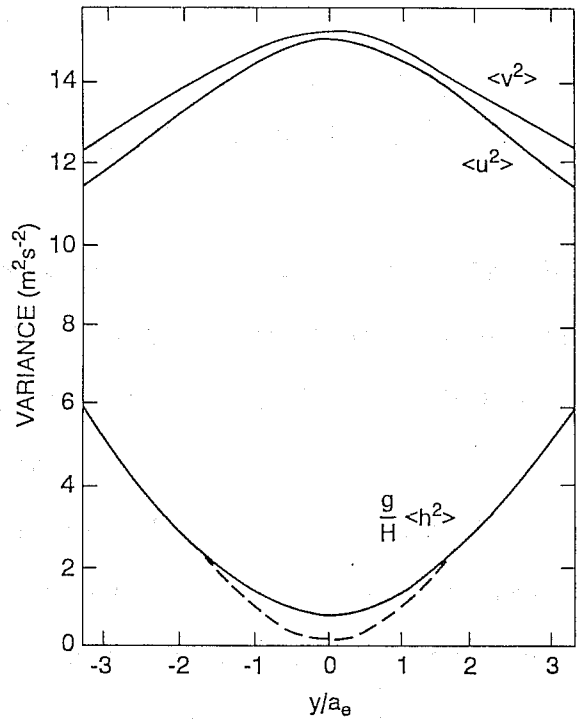

Fig. 9 Graph of $<v^2>$, $<u^2>$ and $g<h^2>/H$ as a function of y/a_e

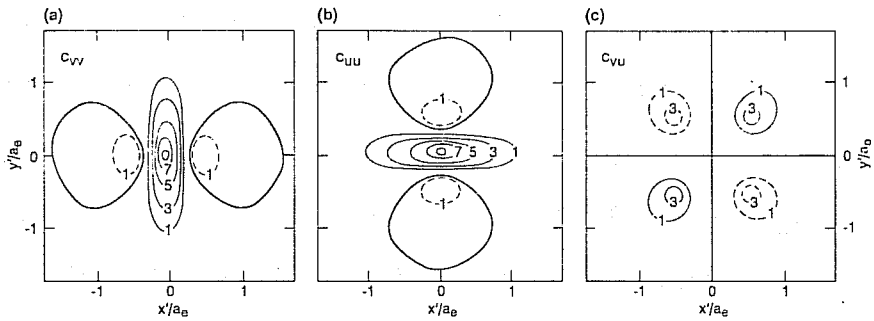

Fig. 10 Graph of c_{vv} (panel a), c_{uu} (panel b) and c_{vu} (panel c)

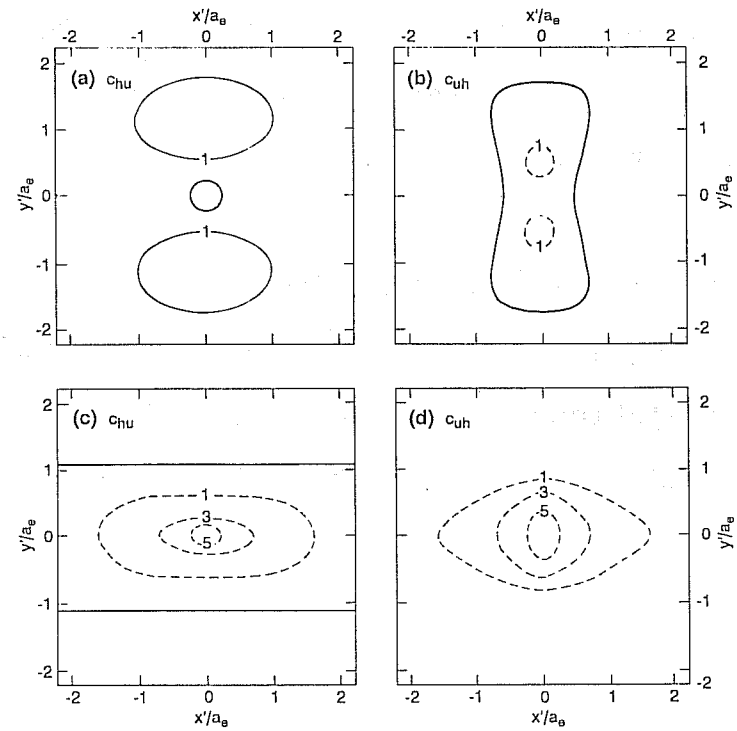

Fig. 11 Graph of c_{hu} and c_{uh} (panels a and b respectively) where all modes have frequencies less than the cut-off including some Kelvin modes are slow. Also graph of c_{hu} and c_{uh} (panels c and d respectively) for the second case where no Kelvin modes are counted as slow modes

h and u have the same symmetry and we cannot rely on symmetry arguments to suggest the form of the h/u and u/h correlations. In Figure 11, we show two cases. In the first case, all modes with frequencies less than the cut-off, including some Kelvin modes are slow. For this case, c_{hu} and c_{uh} are plotted in Figure 11 (panels a and b) respectively. In the second case, no Kelvin modes are counted as slow modes. c_{hu} and c_{uh} for this case are plotted in panels (c and d) respectively. This second case corresponds to the dashed line in Figure 9. Consider first panels (c and d). These correlations are composed entirely of Rossby modes, in which u and h are negatively correlated at the equator. This implies a substantial negative correlation. For Kelvin modes, h and u are positively correlated at the equator. Thus, the addition of the slow Kelvin modes effectively reduces u/h and h/u correlations.

The correlation C_{hh} is plotted in Figure 12a. It is very large scale and somewhat elongated along the x axis. In Figure 12b is plotted the corresponding covariance C_{hh}. The covariance is essentially isotropic, so that the apparent non-isotropy of the correlation is simply due to latitudinal variation of the height $< h^2 >$ depicted in Figure 9. We also examined the characteristic scale of the h/h covariance as a function of latitude. We defined a characteristic scale L_h in the following simple way. Since the h/h covariances were essentially isotropic, the characteristic scale was defined to be the distance at which covariance decreased to 0.3 of its central value. In Figure 13

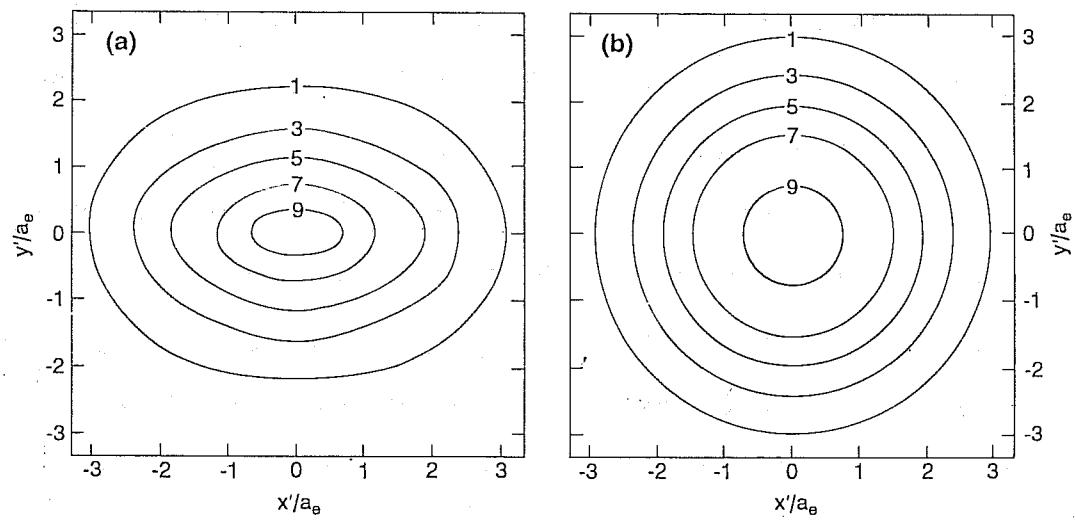

Fig. 12 (a) Graph of correlation c_{hh} (b) and correlation c_{hh}

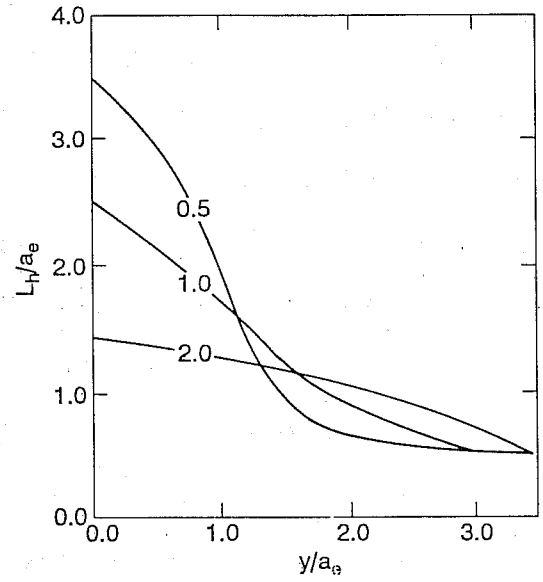

Fig. 13 Graph of L_h/a_e as a functions of y/a_c

we plot L_h/a_e as a function of y/a_e. We consider three cases, $\omega_c/(2\beta c)^{1/2} = 0.5$, 1.0 and 2.0. When the normal frequency cut-off $\omega_c/(2\beta c)^{1/2} = 1$ is applied there is a substantial increase in scale for the h/h covariance as the latitude decreases. This equatorial increase in scale is enhanced/diminished with a more/less stringent cut-off frequency for the slow modes. The effect of varying the cut-off frequency on the wind covariances is much smaller.

Optimal Kalman filter on the equatorial beta plane

On an observation network where y_j are the zeroes of the Hermite polynomials and the the x_j are equally spaced, the Kalman filter can be diagonalized leading to equations (7-9) for each normal mode. This is an enormous simplification of the second momement equations and allows all the results of section 2 to be applied.

In Figure 14, we show the observation and model error variances and the stationary forecast and analysis error variances as a function of latitude for the wind (panel a) and height (panel b). In Figure 15, we plot the wind (m^2s^{-2}) and height (m^2) forecast error variances for the slow and fast modes as a function of time (ordinate) and latitude (abscissa). Panel (a) shows the slow mode wind, panel (b) the slow mode height, panel (c) the fast mode wind and panel (d) the fast mode height. It can be seen that for the slow modes, both wind (panel a) and height (panel b) error variances reach stationarity very quickly. Note that section 2

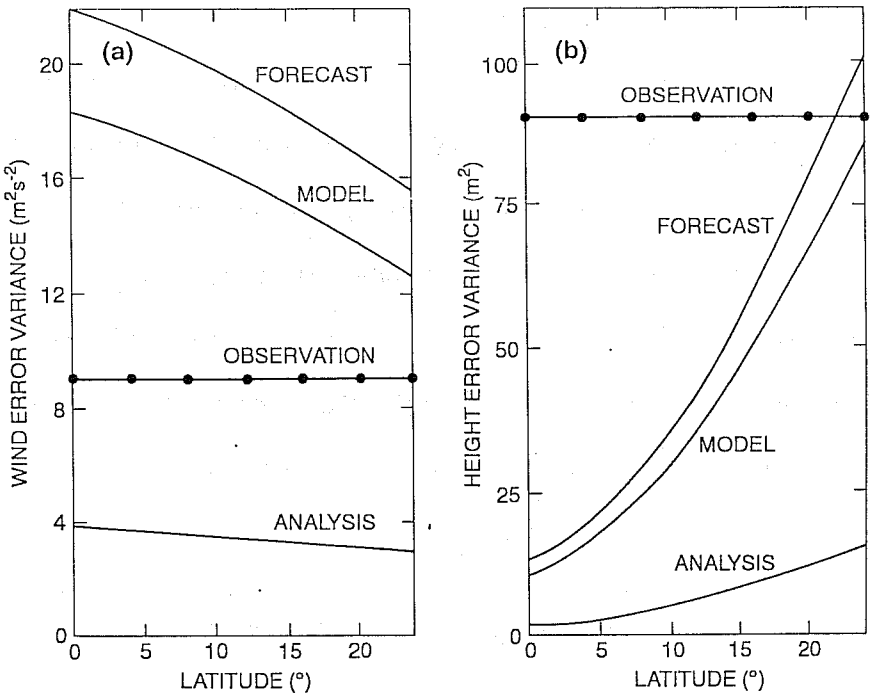

Fig. 14 Graph of the observation and model error variances and the stationary forecast and analysis error variances as a function of latitude for the wind (panel a) and height (panel b)

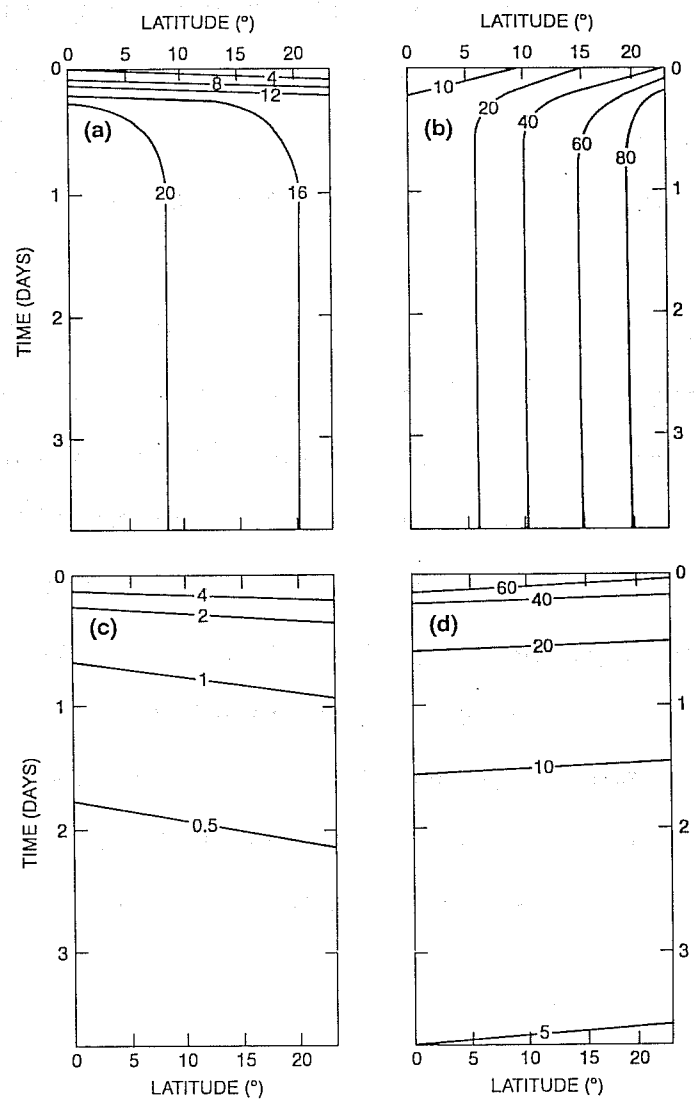

Fig. 15 Graph of the wind and height forecast error variances for the slow and fast modes as a function of time (ordinate) and latitude (abscissa). Panel (a): slow mode wind, panel (b): slow mode height, panel (c): fast mode wind and panel (d): fast mode height

suggests for $m^2(\nu) = 1$, $q^2(\nu) \neq 0$, $c_+(\nu) << 1$ and convergence to stationarity is rapid. For the fast mode wind (panel c) and height (panel d) forecast error variances, convergence to the stationary solution ($f_+^2(\nu) = 0$), is much slower. Again, this is what would be expected when $c_+(\nu) = 1$. The convergence rate for both slow and fast modes is almost independent of latitude.

The sub-optimal filter

The optimal gain matrix (3) is replaced by,

$$\mathbf{K}_s = \mathbf{P}_s^f \mathbf{H}^T [\mathbf{H}\mathbf{P}_s^f \mathbf{H}^T + \mathbf{R}]^{-1} \quad , \tag{35}$$

where $\mathbf{P}_s^f$ is the (incorrectly) specified forecast error covariance. Diagionalization of the equations leads to a general form of (8) and the stationary solutions,

$$f_+^2 = [q^2(f_s^2 + r^2)^2 + m^2 r^2 f_s^4] \; / \; [(f_s^2 + r^2)^2 - m^2 r^4 \;] \; . \tag{36}$$

These equations are analyzed in Daley (1993).

References

Daley, R., 1992: The lagged innovation covariance: a performance diagnostic for atmospheric data assimilation. *Mon. Wea. Rev.* **120**, 178-196.

Daley, R., 1993: Atmospheric data assimilation on the equatorial beta plane. Submitted to *Atmosphere-Ocean.*

Daley, R. and R. Ménard, 1993: Spectral characteristics of Kalman filter systems for atmospheric data assimilation. *Mon. Wea. Rev.* In press.