Chapter 9

Atmospheric Data Assimilation: Advanced Methods

9.1 Developments toward Advanced Methods

The tendency in operational data assimilation centers these days such as NCEP and ECMWF, is to evolve in the direction of eliminating certain hypothesis about approximating the evolution of forecast error covariances. As discussed previously, one of the most fundamental ingredients of the methods based on estimation theory is the propagation of error covariances by means of the dynamics of the system. If it were possible to calculate this evolution completely, the separability hypothesis between the vertical and horizontal correlations, geostrophic assumption, homogeneity, and isotropy hypothesis would not be necessary, because the dynamical properties would be present in the corresponding error covariance matrix; no artificial properties would have to be imposed by ad hoc constraints. Moreover, the error covariances would evolve instead of being stationary. However, we know that the calculation of the forecast error covariance is impractical due to the large computational burden. The computing progress of the past few years has been very promising. As a consequence, it is becoming possible to develop methods that allow for a slow relaxation of many of these conventional hypothesis. Some data assimilation systems today are designed with the goal of allowing easy progress and implementation of improvements on the error covariance structure. An example of such a versatile system is the Physical–Space Statistical Analysis System (PSAS; da Silva et al. [37] e Guo & da Silva [68], of the Goddard Space Flight Center. The procedure to relax the hypothesis in the construction of error covariances is referred to as construction of *error covariance models*, as each premise that gets eliminated, or substituted by less restrictive premises, generates a new error covariance model.

In this section we describe two ideas for relaxing some of the conventional constraints imposed on modeling error covariances. Initially, we consider a way of eliminating the *geostrophic* dynamical constraint, by presenting a way of building error covariance with coherent balances from the governing dynamics, for simple dynamics. Later, we consider a model that eliminates the *separability* hypothesis between vertical and horizontal corre-

lations. Also in this Lecture, we present the basic ideas of what is called *parameterization* of error covariance and how to calibrate the quantities involved in these parameterizations. Finally, we discuss an application of the Kalman filter for a simple dynamical system, and some approximations to the Kalman filter with possible practical potential.

9.1.1 Generation of Balanced Cross–Covariances

The imposition of the geostrophic balance dynamic constraint used in OI is certainly problematic since the equations of motion described in Lecture 7 do not satisfy this constraint exactly, but only approximately. Consequently, the forecast error covariance matrix which corresponds to the complete system of governing equations is only approximately geostrophic. When the geostrophic approximation is used to construct a forecast error covariance $\mathbf{S}_k^f$, the corresponding gains obtained through (8.48) have components in the rapid modes of the system, consequently generating a analyses $\mathbf{w}_k^a$ initialized incorrectly, that is, containing rapid waves, that degrade the quality of the forecasts. One of the ways to avoid this type of inconsistency is to develop a procedure that combines the real intrinsic balances of the governing dynamics with balances imposed in the forecast error covariance matrix by means of the slow modes of the system, without them necessarily being geostrophic modes. The generation of balanced covariances in this form suggests the possibility of eliminating the initialization stage of the analysis produced at a given instant of time. This fact was initially observed in the implementation of the analysis system of NCEP, the so called "Spectral Statistical–Interpolation" (SSI) developed by Parrish & Derber [112], using the linear balance relation, instead of the geostrophic relation. Similarly, the European system at ECMWF substitutes the geostrophic relation by a more general balance obtained through the Hough modes (see Heckley et al. [74]). What we describe below is a simplified procedure, analogous to this latter one.

To describe this type of procedure we will follow the treatment of Todling & Cohn [129], remembering that this procedure can be extended to the nonlinear case in which the dynamics is that of a general circulation model. In the simple cases to be considered here, the governing dynamics that we have in mind is linear with n degrees of freedom and consequently n normal modes; n_S modes classified as slow. Moreover, we take a simple system, such as that from the discretized shallow water equations on the plane. Therefore, the system variables are the zonal and meridional winds u and v, respectively, and the heights h, at each grid point.

In this case, as in OI, the idea is to base the construction the error covariance matrix $\mathbf{S}^f$, omitting the index k referring to the time t_k, by specifying only the error forecast error covariance matrix for the height fields, designated by $\mathbf{S}^{f|hh}$. Here we use the same notation as in the previous lecture, with the following difference: now the height–height forecast error covariance is a matrix of dimension $n/3 \times n/3$, instead of a function. This is simply due to the fact that we are now assuming that the governing equations have been discretized in some way. Therefore, not only the height–height forecast error covariance is a matrix, but also all the other covariances and cross–covariances in $\mathbf{S}^f$ are matrices. That is, the

decomposition (8.50) is redefined as

$$\mathbf{S}^f = \begin{pmatrix} \mathbf{S}^{f|uu} & \mathbf{S}^{f|uv} & \mathbf{S}^{f|uh} \\ \mathbf{S}^{f|vu} & \mathbf{S}^{f|vv} & \mathbf{S}^{f|vh} \\ \mathbf{S}^{f|hu} & \mathbf{S}^{f|hv} & \mathbf{S}^{f|hh} \end{pmatrix} \tag{9.1}$$

where now all the elements of $\mathbf{S}^f$ are written in bold face because they are matrices of dimension $n/3 \times n/3$. Notice here we use height instead of the geopotential function, without loss of generality.

We should recognize the fact that not every error covariance matrix hh (referring to height–height), $\mathbf{S}^{f|hh}$, can be a block corresponding to a "slow" multivariate error covariance matrix $\mathbf{S}^f$. This is easily understood if we notice that in general $\mathbf{S}^{f|hh}$ has dimension $n/3 \times n/3$, but n_S can be less than $n/3$. If $\mathbf{S}^{f|hh}$ should be part of a slow matrix $\mathbf{S}^f$, then, it can be at most rank n_S. Thus, we have to establish from the start the "slowness" of $\mathbf{S}^{f|hh}$, and later build the rest of the error covariance matrix, that is , the cross covariance matrices so that the resulting $\mathbf{S}^f$ is slow. This leads to an expression similar to the one introduced in (8.61), but in which the operators that act on $\mathbf{S}^{f|hh}$ carry the fact that $\mathbf{S}^f$ is "slow" instead of geostrophic.

As in the shallow water problem considered in Lectures 6 and 7, we assume that the normal modes of the system in question can be collected as the columns of a matrix $\mathbf{V}$, of dimension $n \times n$. The matrix formed by n_S slow vectors is represented by $\mathbf{V}_S$ and has dimension $\times n_S$. It is necessary to partition this last matrix as

$$\mathbf{V}_S = \begin{pmatrix} \mathbf{V}_u \\ \mathbf{V}_v \\ \mathbf{V}_h \end{pmatrix} \tag{9.2}$$

where each block has dimension $n/3 \times n_S$, and where for example, $\mathbf{V}_h$ represents the matrix of the height components of slow normal modes. Furthermore, we designate by $\mathcal{S}_h$ the subspace that spanned by the n_S columns of the matrix $\mathbf{V}_h$.

First, we want to modify any covariance matrix $\mathbf{S}^{f|hh}$ to make it "slow". This problem can be arranged as follows. Find a matrix $\mathbf{X}$ that satisfies the following conditions:

$$Range\, \mathbf{X} \in \mathcal{S}_h\,, \tag{9.3}$$

and

$$\mathbf{X}^T = \mathbf{X}\,, \tag{9.4}$$

minimizing at the same time the scalar functional

$$\eta \equiv ||\mathbf{X} - \mathbf{S}^{f|hh}||_F^2 \tag{9.5}$$

The condition (9.3) imposes that the columns of $\mathbf{X}$ belong to the subspace $\mathcal{S}_h$, i.e., that the height components are a linear combination of slow modes. The second condition imposes that $\mathbf{X}$ be symmetric, since it has to represent a covariance matrix. We do not demand above that this matrix be positive semi–definite, but we will see that the solution to the problem (9.3)–(9.5) is positive semi–definite, and therefore it can represent a covariance

matrix. The last condition above demands that $\mathbf{X}$ be as close as possible to $\mathbf{S}^{f|hh}$ in the Frobenius norm $||.||_F$, which is defined as the sum of the squares of the elements of a matrix (See Golub & Van Loan [67]). This is a norm that penalizes equally each element of the difference between $\mathbf{S}^{f|hh}$ and $\mathbf{X}$. Another norm could have been chosen, as for example a norm with weights, but the Frobenius norm leads to a very simple solution.

As we seen below, the solution of the problem (9.3)–(9.5) is unique and given by

$$\mathbf{X} \;=\; \mathbf{\Pi}_h \mathbf{S}^{f|hh} \mathbf{\Pi}_h^T \,, \tag{9.6}$$

where $\mathbf{\Pi}_h$ is the orthogonal projector for the subspace $\mathcal{S}_h$, that is,

$$\mathbf{\Pi}_h \;=\; \mathbf{V}_h \, (\mathbf{V}_h^T \mathbf{V}_h)^{-1} \, \mathbf{V}_h^T \,. \tag{9.7}$$

The solution, $\mathbf{X}$ in (9.6) is positive semi–definite, since it corresponds to a congruence transformation of a positive semi–definite matrix $\mathbf{S}_k^{f|hh}$. Also, notice that it is not possible for the solution to be positive definite, because $\mathbf{\Pi}_h$ has rank n_S, since the matrix is invertible $\mathbf{V}_h^T \mathbf{V}_h$ in (9.7) has dimension $n_S \times n_S$; consequently $\mathbf{X}$ is at most rank n_S.

We want to show that $\mathbf{X}$ given in (9.6) is the unique solution of (9.5). For this, notice that as $\mathbf{\Pi}_h$ is the orthogonal projector on $\mathcal{S}_h$, it satisfies the following condition:

$$\begin{aligned} Range\, \mathbf{\Pi}_h &= \mathcal{S}_h \,, && \text{(9.8a)} \\ \mathbf{\Pi}_h^2 &= \mathbf{\Pi}_h \,, && \text{(9.8b)} \\ \mathbf{\Pi}_h^T &= \mathbf{\Pi}_h \,. && \text{(9.8c)} \end{aligned}$$

Equation (9.3) is satisfied by $\mathbf{X}$, given the condition (9.8a); moreover (9.4) it is obviously satisfied. Now we show that $\mathbf{X}$ given in (9.6) minimizes (9.5), in a unique way.

For the moment, let us denote by $\mathbf{X}_*$ the solution given in (9.6), i.e.,

$$\mathbf{X}_* \;\equiv\; \mathbf{\Pi}_h \mathbf{S}^{f|hh} \mathbf{\Pi}_h^T \,. \tag{9.9}$$

Then, any $\mathbf{X}$ which satisfies (9.3) and (9.4) should be of form

$$\mathbf{X} \;=\; \mathbf{X}_* \,+\, \tilde{\mathbf{X}} \,, \tag{9.10}$$

where

$$Range\, \tilde{\mathbf{X}} \in \mathcal{S}_h \,, \tag{9.11}$$

and

$$\tilde{\mathbf{X}}^T = \tilde{\mathbf{X}} \,. \tag{9.12}$$

To show that $\mathbf{X}_*$ is the only minimizer, we need to show that $\tilde{\mathbf{X}} = \mathbf{0}$ at the minimum.

Substituting (9.10) in (9.5) we have

$$\begin{aligned} \eta \;=\; \eta(\mathbf{X}) \;&\equiv\; ||\mathbf{X} - \mathbf{S}^{f|hh}||_F^2 \\ &=\; \mathrm{Tr}\, (\mathbf{X} - \mathbf{S}^{f|hh})^T (\mathbf{X} - \mathbf{S}^{f|hh}) \\ &=\; \mathrm{Tr}\, (\mathbf{X}_* - \mathbf{S}^{f|hh} + \tilde{\mathbf{X}})^T (\mathbf{X}_* - \mathbf{S}^{f|hh} + \tilde{\mathbf{X}}) \\ &=\; \eta(\mathbf{X}_*) \,+\, ||\tilde{\mathbf{X}}||_F^2 \,+\, 2\delta \,, \end{aligned} \tag{9.13}$$

where we use the definition of the Frobenius norm, as well as the definition of the *trace* operator (see Golub & Van Loan [67], pp. 56 and 332), and where we write

$$\delta \;=\; \mathrm{Tr}\,\tilde{\mathbf{X}}^T(\mathbf{X}_* - \mathbf{S}^{f|hh})\,. \tag{9.14}$$

From expressions (9.11) and (9.12) it follows that $\tilde{\mathbf{X}}$ should be of form

$$\tilde{\mathbf{X}} \;=\; \mathbf{\Pi}_h\mathbf{Y}\mathbf{\Pi}_h^T\,, \tag{9.15}$$

for any symmetric matrix $\mathbf{Y}$. Substituting the expression above and (9.9) in (9.14) we obtain

$$\begin{aligned}\delta &= \mathrm{Tr}\,\mathbf{\Pi}_h^T\mathbf{Y}\mathbf{\Pi}_h(\mathbf{\Pi}_h\mathbf{S}^{f|hh}\mathbf{\Pi}_h^T - \mathbf{S}^{f|hh}) \\ &= \mathrm{Tr}\,\mathbf{Y}\mathbf{\Pi}_h(\mathbf{\Pi}_h\mathbf{S}^{f|hh}\mathbf{\Pi}_h^T - \mathbf{S}^{f|hh})\mathbf{\Pi}_h^T\,,\end{aligned} \tag{9.16}$$

where we use (9.8c) and the fact that

$$\mathrm{Tr}\,\mathbf{A}^T\,\mathbf{B} \;=\; \mathrm{Tr}\mathbf{B}\,\mathbf{A}^T\,, \tag{9.17}$$

for any two matrices $\mathbf{A}$ and $\mathbf{B}$ with same dimensions. Using (9.8b) in (9.16) we have

$$\begin{aligned}\delta &= \mathrm{Tr}\,\mathbf{Y}(\mathbf{\Pi}_h\mathbf{S}^{f|hh}\mathbf{\Pi}_h^T - \mathbf{\Pi}_h\mathbf{S}^{f|hh}\mathbf{\Pi}_h^T) \\ &= 0\,,\end{aligned} \tag{9.18}$$

and therefore (9.13) can be written as:

$$\begin{aligned}\eta(\mathbf{X}) &= \eta(\mathbf{X}_*) + ||\tilde{\mathbf{X}}||_F^2 \\ &\geq \eta(\mathbf{X}_*)\,,\end{aligned} \tag{9.19}$$

where the sign of equality prevails if, and only if, $\tilde{\mathbf{X}} = \mathbf{0}$, since $||\tilde{\mathbf{X}}||_F$ is canceled, and if and only if, $\tilde{\mathbf{X}} = \mathbf{0}$. Then, $\mathbf{X}_*$ minimizes η, in a unique way, completing the demonstration.

Combining the expressions (9.6) and (9.7) we have that

$$\mathbf{X} \;=\; \mathbf{V}_h\,\hat{\mathbf{X}}\,\mathbf{V}_h^T\,, \tag{9.20}$$

where

$$\hat{\mathbf{X}} \;\equiv\; (\mathbf{V}_h^T\mathbf{V}_h)^{-1}\mathbf{V}_h^T\mathbf{S}^{f|hh}\mathbf{V}_h(\mathbf{V}_h^T\mathbf{V}_h)^{-1} \tag{9.21}$$

is a symmetric matrix of dimension $n_S \times n_S$. It is not difficult to observe that any slow covariance matrix $\mathbf{S}$ should be of form

$$\mathbf{S} \;=\; \mathbf{V}_S\hat{\mathbf{S}}\mathbf{V}_S^T\,, \tag{9.22}$$

for any symmetric matrix $\hat{\mathbf{S}}$, of dimension $n_S \times n_S$. In this case, $\hat{\mathbf{S}}$ is the representation of $\mathbf{S}$ in the space of normal modes. Comparing (9.2), (9.20) and (9.22), it follows that

$$\mathbf{S}^f \;\equiv\; \mathbf{V}_S\hat{\mathbf{X}}\mathbf{V}_S^T \tag{9.23}$$

is a unique slow covariance matrix for which the covariance block hh coincides with the matrix of slow hh covariance $\mathbf{X}$. In this way, the formulas (9.21) and (9.23) provide a way of building a dynamically balanced slow error covariance matrix $\mathbf{S}^f$, given general height error covariance matrix $\mathbf{S}^{f|hh}$. The matrix $\hat{\mathbf{X}}$ is a representation in the space of normal modes of $\mathbf{S}^f$. Equations (9.21) and (9.23) replace the construction of geostrophically balanced covariances through (8.61).

9.1.2 A Non–Separable Covariance Model

Let us consider now the case of abandoning the vertical separability hypothesis that has been mentioned in the previous lecture. This hypothesis is known to be responsible for the failure of data assimilation system in producing analysis capable of forecast regions of rapid vertical atmospheric motion. The baroclinic instability involves such motions and it is one of the main atmospheric instabilities. Assimilation systems currently in operations still largely underestimate these atmospheric instabilities due to their poorly prescribed forecast error covariance models. Since these models account for no correlation among fields at different vertical levels, what happens in those cases is that the information provided by observations at a certain levels of the atmosphere is not correctly transferred to other levels because of the vertical separability assumptions. The treatment of this section is due to Bartello & Mitchell [7], and it has the goal of building error covariance functions for which the vertical and horizontal relations are entirely determined by the dynamics, and therefore being non–separable. What these authors proposed is based on a system of simplified equations, analogously to what its done in OI, where the covariance structure is build on the basis of the geostrophic balance relation. As simple as it might be, Bartello & Mitchell's model is an extremely promising one.

Following Bartello & Mitchell's description, we consider the system of primitive equations linearized about a basic state with buoyancy fluctuation $\nu = \sqrt{g^2/(c_p T)}$, independent of height, where c_p is the the constant of specific heat to constant pressure for the dry air. Moreover, the vertical coordinate is taken as the pressure, so that $Z = -H \ln(p/p_s)$, where $H = RT/g$ is the height scale, $p_s = 1000$ mb is the pressure at the surface, and R, T, and g has the same meanings as in Lecture 7. The basic state is that of rest, and the Coriolis parameter $f = f_0$ is taken to be constant.

The equations of motion in this case (see Holton [82], Section 11.3) are given by

$$\frac{\partial u}{\partial t} - f_0 v + \frac{\partial \phi}{\partial x} = 0 \tag{9.24a}$$

$$\frac{\partial v}{\partial t} + f_0 u + \frac{\partial \phi}{\partial y} = 0 \tag{9.24b}$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial W}{\partial Z} - \frac{W}{H} = 0 \tag{9.24c}$$

$$\frac{\partial}{\partial t}\left(\frac{\partial \phi}{\partial Z}\right) + \nu^2 W = 0 \tag{9.24d}$$

The boundary conditions on top and at the surface of the atmosphere can be written as:

$$W = \frac{dZ}{dt} = -\frac{\delta}{g}\frac{\partial \phi}{\partial t} \tag{9.25}$$

for $Z = 0$ e $Z = \bar{Z}$, where $\bar{Z}$ represents the top of the atmosphere. Here, $\delta = 0$ corresponds to a zero vertical speed.

We can eliminate W, partially, from the equations above by substituting (9.24d) in (9.24c), that is,

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = \frac{1}{\nu^2}\left(\frac{\partial}{\partial Z} - \frac{1}{H}\right)\frac{\partial}{\partial t}\left(\frac{\partial \phi}{\partial Z}\right)$$

$$= \frac{1}{\nu^2}\frac{\partial}{\partial t}\left(\frac{\partial^2}{\partial Z^2} - \frac{1}{H}\frac{\partial}{\partial Z}\right)\phi \tag{9.26}$$

where we say partially since the boundary conditions are still given as a function of W.

Let us look for solutions of the form

$$u = U(x,y,t)A(Z) \tag{9.27a}$$
$$v = V(x,y,t)A(Z) \tag{9.27b}$$
$$\phi = \Phi(x,y,t)A(Z) \tag{9.27c}$$
$$W = \Omega(x,y,t)B(Z) \tag{9.27d}$$

Substituting these functions in the equations for u and v (9.24a) and (9.24b), respectively, it is simple to see that due to the fact that these equations do not involve the coordinate Z, they involve U, V and Φ only, that is,

$$\frac{\partial U}{\partial t} - f_0 V + \frac{\partial \Phi}{\partial x} = 0 \tag{9.28a}$$
$$\frac{\partial V}{\partial t} + f_0 U + \frac{\partial \Phi}{\partial y} = 0 \tag{9.28b}$$

Now substituting (9.27) in (9.26) we have that

$$A\left(\frac{\partial U}{\partial x} + \frac{\partial V}{\partial y}\right) = \frac{1}{\nu^2}\frac{\partial \Phi}{\partial t}\left(\frac{d^2A}{dZ^2} - \frac{1}{H}\frac{dA}{dZ}\right) \tag{9.29}$$

or yet,

$$\left(\frac{\partial \Phi}{\partial t}\right)^{-1}\left(\frac{\partial U}{\partial x} + \frac{\partial V}{\partial y}\right) = \frac{1}{\nu^2 A}\left(\frac{d^2A}{dZ^2} - \frac{1}{H}\frac{dA}{dZ}\right) \tag{9.30}$$

Noticing that the left hand side of this equality is independent of Z, while the right hand side is independent of (x,y,t), we can separate this equation in two:

$$\frac{\partial \Phi}{\partial t} + c^2\left(\frac{\partial U}{\partial x} + \frac{\partial V}{\partial y}\right) = 0 \tag{9.31}$$
$$\left(\frac{d^2A}{dZ^2} - \frac{1}{H}\frac{dA}{dZ}\right) + \frac{\nu^2}{c^2}A = 0 \tag{9.32}$$

where c^2 is the separability constant. The expression (9.32) is the vertical structure equation.

To re–write the boundary conditions (9.25) with solutions given by (9.27) we substitute (9.25) in (9.24d)

$$\frac{\partial}{\partial t}\left(\frac{\partial \phi}{\partial Z}\right) = \nu^2\frac{\delta}{g}\frac{\partial \phi}{\partial t} \tag{9.33}$$

for $Z = 0$ e $Z = \bar{Z}$. And therefore, using (9.27) we have that

$$\frac{dA}{dZ} = \frac{\delta\nu^2}{g}A \tag{9.34}$$

for $Z = 0$ e $Z = \bar{Z}$.

The vertical structure equation (9.32), with boundary conditions (9.34), is a Sturm–Liouville problem (e.g., Arfken [5]) which can be solved without difficulty. Due to the boundary conditions, the constant c^2 can assume only certain discrete values c_n^2, with corresponding solutions $A(Z) = A_n(Z)$. Through the orthogonality properties of the solutions of the Sturm–Liouville problem we can write

$$\int_0^{\bar{Z}} A_n(Z) A_m(Z) e^{-Z/H}\, dZ \;=\; \delta_{nm}, \tag{9.35}$$

where δ_{nm} is a Kronecker delta.

Furthermore, the equations (9.28a), (9.28b), (9.31) for U, V, and Φ, respectively, representing the horizontal structure can be solved by the applying Fourier transform, in analogy to what we did in Lecture 7. Assuming that we are treating the case for the infinite plane, U, V and Φ can be written as

$$\begin{pmatrix} U \\ V \\ \Phi \end{pmatrix} (\mathbf{r}, t) \;=\; \int_{R^2} \begin{pmatrix} \hat{U} \\ \hat{V} \\ \hat{\Phi} \end{pmatrix} (\mathbf{s}, t)\, e^{-i\mathbf{s}^T\mathbf{r}}\, d\mathbf{s} \tag{9.36}$$

where $\mathbf{r} = (x, y)$ and $\mathbf{s} = (k, \ell)$, and the Fourier transform is given as in (2.54). The coefficients $\hat{U}$, $\hat{V}$, $\hat{\Phi}$ can be determined by substituting the expression above in equations (9.28a), (9.28b), (9.31). So that the solution for u, v and ϕ can be put in form

$$\mathbf{w}(\mathbf{r}, Z, t) \;=\; \sum_{n=0}^{\infty} A_n(Z) \int_{R^2} \hat{\mathbf{w}}(\mathbf{s}, t)\, e^{-i\mathbf{s}^T\mathbf{r}}\, d\mathbf{s} \tag{9.37}$$

where $\mathbf{w} \equiv (u, v, \phi)^T$ and $\hat{\mathbf{w}} \equiv (\hat{u}, \hat{v}, \hat{\phi})^T$.

To build a covariance model we can follow a similar path to that in Section 8.4. Assume that the real state $\mathbf{w}^t(\mathbf{r}, Z, t)$, as well as the forecast state $\mathbf{w}^f(\mathbf{r}, Z, t)$, obey the same equations of motion (9.24a)–(9.24d). Therefore, due to the linearity we have that

$$\mathbf{e}^f(\mathbf{r}, Z, t) \;=\; \sum_{n=0}^{\infty} A_n(Z) \int_{R^2} \hat{\mathbf{e}}^f(\mathbf{s}, t)\, e^{-i\mathbf{s}^T\mathbf{r}}\, d\mathbf{s} \tag{9.38}$$

where $\mathbf{e}^f = \mathbf{w}^f - \mathbf{w}^t$ is the forecast error. The vertical structure functions are the same, for the errors as well as for the fields, since the errors follow the same separation of variables (9.27).

The error covariance matrix between two spatial points $(\mathbf{r}_i, Z_i)$ and $(\mathbf{r}_j, Z_j)$ can be found by calculating the outer product between error vectors $\mathbf{e}^f$ at two spatial points i and j, and by using the ensemble mean operator. That is,

$$\begin{aligned} \mathbf{S}^f(\mathbf{r}_i, Z_i, \mathbf{r}_j, Z_j, t) \;&\equiv\; \mathcal{E}\{\mathbf{e}^f(\mathbf{r}_i, Z_i, t)\, (\mathbf{e}^f(\mathbf{r}_i, Z_i, t))^*\} \\ &=\; \sum_{n,m=0}^{\infty} A_n(Z_i)\, A_m(Z_j) \\ &\quad \int_{R^2} \int_{R^2} \mathcal{E}\{\hat{\mathbf{e}}^f(\mathbf{s}_i, t)(\hat{\mathbf{e}}^f(\mathbf{s}_j, t))^*\}\, e^{i(\mathbf{s}_i^T\mathbf{r}_i - \mathbf{s}_j^T\mathbf{r}_j)}\, d\mathbf{s}_i\, d\mathbf{s}_j \end{aligned} \tag{9.39}$$

where $*$ represents the transpose conjugated. This expression can still be written as

$$\mathbf{S}^f(\mathbf{r}_i, Z_i, \mathbf{r}_j, Z_j, t) \;=\; \sum_{n,m=0}^{\infty} A_n(Z_i)\, A_m(Z_j)\, \bar{\mathbf{S}}^f(\mathbf{r}_i, \mathbf{r}_j, t) \tag{9.40}$$

where $\bar{\mathbf{S}}^f(\mathbf{r}_i, \mathbf{r}_j, t)$ is a horizontal covariance

$$\bar{\mathbf{S}}^f(\mathbf{r}_i, \mathbf{r}_j, t) \;=\; \int_{R^2}\int_{R^2} \mathcal{E}\{\hat{\mathbf{e}}^f(\mathbf{s}_i, t)(\hat{\mathbf{e}}^f(\mathbf{s}_j, t))^*\}\, e^{i(\mathbf{s}_i^T\mathbf{r}_i - \mathbf{s}_j^T\mathbf{r}_j)}\, d\mathbf{s}_i\, d\mathbf{s}_j \tag{9.41}$$

depending on time t. From the expression (9.40) we see that the covariance is non–separability for the horizontal and vertical components.

At this point Bartello & Mitchell [7] comment that this covariance model can be used to determine the complete covariance matrix, that is, the covariance functions $S^{f|uu}$, $S^{f|uv}$, etc., where homogeneity and isotropy hypothesis can be employed. Another possibility is to use only the block of the covariance matrix $\mathbf{S}^f$ corresponding to the height–height covariance function $S^{f|hh}$. The remainder of the covariances and cross–covariances can be determined by means of the geostrophic balance relation. This, in fact, simplifies calibration procedures that have to be used so that we obtain error covariances with relevant (physical) meaning for assimilation systems.

Imposition of the homogeneity and isotropy assumptions for $S^{f|hh}$ lead to

$$\bar{S}^{f|hh}(r, t) \;=\; 2\pi \int_0^{\infty} \hat{S}^{f|hh}(\omega)\, J_0(\omega r)\, \omega\, d\omega \tag{9.42}$$

where J_0 is the order–zero Bessel function, and $r = |\mathbf{r}_i - \mathbf{r}_j|$. This is identical to the result that we obtained in (2.66) when we discussed isotropic covariances in R^2.

Finally we can write

$$S^{f|hh}(r, Z_i, Z_j, t) \;=\; 2\pi \sum_{n,m=0}^{\infty} A_n(Z_i)\, A_m(Z_j) \int_0^{\infty} \hat{S}^{f|hh}(\omega)\, J_0(\omega r)\, \omega\, d\omega \tag{9.43}$$

is the complete height–height error covariance function. In practice, this function needs to be transformed into the matrix $\mathbf{S}^{f|hh}$), and adjusted to real data; for example, by means of fitting techniques, such as least squares. Details on practical implementations are discussed in the original work of Bartello & Mitchell [7].

9.1.3 Covariance Tuning

At this point it should be clear that modeling error covariances is fundamental in atmospheric data assimilation. Once a covariance model is constructed, for example for the forecast error covariance $\mathbf{S}^f$, we need to make the analytical model correspond to reality in some way. This is done in general by comparison with the data provided by the observational network and the model forecasts provided by the general circulation models. As mentioned in the previous section, one of the consistent ways of making the adjustment is by means of least squares methods. For schemes such as conventional OI, or models described in the

previous section, the height–height error covariance matrix $\mathbf{S}^{f|hh}$ is modeled in some way and the remainder of the multivariate error covariance is obtained by some type of balance constraint. In the example of OI given in Section 8.4 the correlation distance is in general the parameter to be calibrated, or determined, by means of the comparison against data. For the model of the previous section, the basic state variables, such as the temperature T, and also the order of truncation of the sum and integral in (9.43), are parameters to be estimated in order to calibrate the error covariance statistics.

In this section we describe a way of estimating parameters in an error covariance model based on the ideas of maximum likelihood seen in Lecture 4. This procedure was suggested by Dee [41, 42] with the main intention of calibrating parameters in models for the model error covariance $\mathbf{Q}_k$ and of observation $\mathbf{R}_k$, in the context of advanced assimilation schemes like the Kalman filter, since the statistics of these errors is in general not well known. A particular application of this method is when we need to estimate parameters in the forecast error covariance $\mathbf{S}^f$ (see Dee [41, 42]).

Consider the case in which a error field is represented by the m_k–vector $\mathbf{v}_k$, in time t_k. We want to approximate the covariance matrix of these errors by a matrix $\mathbf{S}_k(\theta)$, where θ is an r–vector of parameters to be determined. We can write

$$\mathcal{E}\{\mathbf{v}_k \mathbf{v}_k^T\} \;\approx\; \mathbf{S}_k(\theta)\,. \tag{9.44}$$

In what follows we refer to the error vectors $\mathbf{v}_k$ as pseudo–innovations.

To calibrate $\mathbf{S}_k(\theta)$ based on samples (or realizations) of the pseudo–innovations vector $\mathbf{v}_k$, we assume that the errors represented by the error covariance matrix are normally distributed, with mean zero and covariance $\mathbf{S}_k(\theta)$,

$$\mathbf{v}_k \;\sim\; \mathcal{N}\left(\mathbf{0}, \mathbf{S}_k(\theta_*)\right)\,, \tag{9.45}$$

at least for some choice of the parameters in (9.44) so that $\theta = \theta_*$.

As we have seen in previous lectures, the assumption made above, together with the assumption that the pseudo–innovations $\{\mathbf{v}_k\}$, for $k = 1, 2, \cdots, K$, is an independent sequence, says that the conditional probability density $p_{\{\mathbf{v}_k\}|\theta}(\{\mathbf{v}_k\}|\theta) = p(\{\mathbf{v}_k\}|\theta)$ is given by the product of Gaussian densities

$$\begin{aligned} p(\{\mathbf{v}_k\}|\theta) &= \prod_{k=1}^{K} p(\mathbf{v}_k|\theta) \\ &= \prod_{k=1}^{K} \frac{1}{(2\pi)^{p_k/2}|\mathbf{S}_k(\theta)|^{1/2}} \\ &\quad \exp\left[-\frac{1}{2}[(\mathbf{v}_k - \boldsymbol{\mu}_k)^T\, \mathbf{S}_k^{-1}(\theta)\, (\mathbf{v}_k - \boldsymbol{\mu}_k)]\right]\,, \end{aligned} \tag{9.46}$$

where $\boldsymbol{\mu}_k = \mathcal{E}\{\mathbf{v}_k\}$.

Following the methodology of maximum likelihood estimation we can obtain an estimate $\theta_{\rm ML}$ for θ_* as being the value that maximizes the conditional probability above, that is,

$$\theta_{\rm ML} \;=\; \arg\max_{\theta}\, p(\{\mathbf{v}_k\}|\theta) \;=\; \arg\min_{\theta}\; f(\theta) \tag{9.47}$$

where

$$f(\theta) \equiv \sum_{k=1}^{K} \left[\ln |\mathbf{S}_k(\theta)| + (\mathbf{v}_k - \boldsymbol{\mu}_k)^T \mathbf{S}_k^{-1}(\theta) (\mathbf{v}_k - \boldsymbol{\mu}_k) \right] \tag{9.48}$$

which is obtained by taking the natural logarithm of (9.46) and ignoring the constant term.

Assuming that the covariance model is stationary, that is,

$$\mathbf{S}_k(\theta) = \mathbf{S}(\theta) \tag{9.49}$$

the likelihood function takes the form

$$\begin{aligned} f(\theta) &= \sum_{k=1}^{K} \left[\ln |\mathbf{S}(\theta)| + (\mathbf{v}_k - \boldsymbol{\mu}_k)^T \mathbf{S}^{-1}(\theta) (\mathbf{v}_k - \boldsymbol{\mu}_k) \right] \\ &= K \ln |\mathbf{S}(\theta)| + \sum_{k=1}^{K} \mathrm{Tr} \left[(\mathbf{v}_k - \boldsymbol{\mu}_k)^T \mathbf{S}^{-1}(\theta) (\mathbf{v}_k - \boldsymbol{\mu}_k) \right] \end{aligned} \tag{9.50}$$

where we introduce the *trace* operator in the last equality, for convenience, after observing that the function f is scalar. Now, notice that

$$\begin{aligned} \sum_{k=1}^{K} \mathrm{Tr} \left[(\mathbf{v}_k - \boldsymbol{\mu}_k)^T \mathbf{S}^{-1}(\theta) (\mathbf{v}_k - \boldsymbol{\mu}_k) \right] &= \sum_{k=1}^{K} \mathrm{Tr} \left[\mathbf{S}^{-1}(\theta) (\mathbf{v}_k - \boldsymbol{\mu}_k) (\mathbf{v}_k - \boldsymbol{\mu}_k)^T \right] \\ &= \mathrm{Tr} \left[\mathbf{S}^{-1}(\theta) \sum_{k=1}^{K} (\mathbf{v}_k - \boldsymbol{\mu}_k) (\mathbf{v}_k - \boldsymbol{\mu}_k)^T \right] \end{aligned} \tag{9.51}$$

where we use the property of the trace that $\mathrm{Tr}(\mathbf{AB}) = \mathrm{Tr}(\mathbf{BA})$, and the last equality is obtained by exchanging the order between the sum under k with the trace operator, since this last one is also a summation operation.

By defining $\bar{f} = f/K$ we have that

$$\bar{f}(\theta) \equiv \ln |\mathbf{S}(\theta)| + \mathrm{Tr}(\mathbf{S}^{-1}(\theta) \bar{\mathbf{S}}) \tag{9.52}$$

where $\bar{\mathbf{S}}$ is the sampling covariance matrix, or sample covariance:

$$\bar{\mathbf{S}} = \frac{1}{K} \sum_{k=1}^{K} (\mathbf{v}_k - \boldsymbol{\mu}_k) (\mathbf{v}_k - \boldsymbol{\mu}_k)^T \tag{9.53}$$

The function, defined in (9.52) is the one to be minimized so that we can determine the parameters θ. This can be done in practice by means of function minimization methods. Many of these methods need the gradient function (9.52), as discussed in Dee [41]. For a relatively small number of parameters θ, that is, when $r \sim o(1)$, and with a reasonable quantity of data $m \sim o(10^2)$, it is possible to obtain a good estimate of parameters, as indicated recently by recent work (Dee 1996, pers. communic.).

9.2 The Kalman Filter for a Simple Model

The Kalman filter was implemented by Cohn & Parrish [33] for a simple model of the atmosphere and we consider this case as an example for a data assimilation system in what follows. We consider the shallow–water equations linearized about a basic state with constant zonal velocity U and zero meridional velocity, and apply it to a β–plane. These equations are discretized with the finite difference scheme discussed in Lecture 7. The distinction from what we saw in Lecture 7, and the application now, is that the boundary conditions here, and in Cohn & Parrish [33], are only periodic in the East–West direction, with "solid walls" in the North–South direction, that is, the perturbations in meridional velocity are zero for all time along the North–South boundaries. This makes the morphology of the finite difference scheme somewhat different from that shown in Fig. 7.1. The extent of the domain of interest in this case is shown here in Fig. 9.1, and encompasses a region with a size equivalent to that of the contiguous United States. The necessary parameters to fully define the system and finite–difference are listed in Table 9.2.

Table 9.1: Shallow–water model parameters as in Cohn & Parrish [33].

Parameters	Values
I grid points in the zonal directions	25
J grid points in the meridional direction	16
East–West extent of channel, L_x	5000 km
North–South extent of channel, L_y	3000 km
Grid size, $\Delta x = \Delta y$	200 km
Time step, Δt	400 s
Coriolis parameter, f_0	6.15×10^{-5} s^{-1}
β–plane parameter, β	1.82×10^{-11} m^{-1}s^{-1}
Basic state geopotential height Φ_0	3×10^4 m^{2}s^{-2}
Basic state zonal speed U	25 ms^{-1}

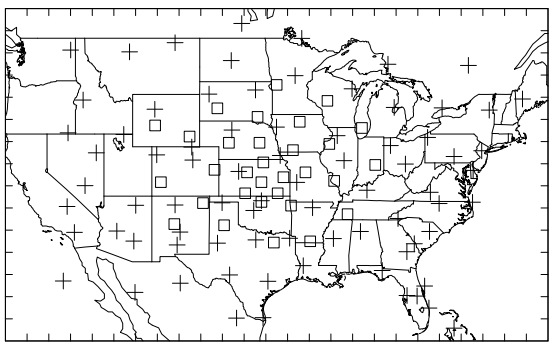

Figure 9.1: Domain of the model of Cohn & Parrish [33] encompassing the contiguous United State. The tick–marks indicate grid points; the "plus" signs indicate radiosonde observations; and the "squares" indicate the wind profilers.

To mimic a real data assimilation system, Cohn & Parrish [33] considered two observational

networks existing in the region of study, just as indicated in Fig. 9.1. The observational network referred as the A–network is composed of 77 radiosondes that observe the winds and the mass field (heights) every 12 hours; and the observational system referred to a the B–network is composed of the radiosondes of the A–network plus 31 wind profilers, which observe the only winds every hour. The error standard deviation for each of these observing systems are shown in Table 9.2.

The equations corresponding to the Kalman filter are implemented for this system with the intention of studying the error evolution in an assimilation period of 2.5 days. Since, in this case, the system of governing equations is linear, the error covariance evolution is decoupled from the state estimate evolution. Therefore, as in Cohn & Parrish [33], we focus discussion only in the behavior of the error evolution, and ignore what happens with the states.

Table 9.2: Observational error standard deviations.

Observations	σ_u (ms^{-1})	σ_v (ms^{-1})	σ_h (m)	No. Obs.
Radiosondes	2.8	2.8	11	77
Wind Profilers	1.5	1.5	——	31

Fig. 9.2 shows the result of assimilation experiments using the observational networks A and B introduced above. The figure shows the time evolution of the error standard deviation, averaged over the domain, for all three variables of the system u, v and h. The plotted quantities correspond to the square root of the sum of the elements of the main diagonal of the forecast and/or analysis error covariance matrices $\mathbf{P}^{f,a}$, divided by the total number of grid points, for each one of the variables. The curves indicated by A refer to the results obtained when only the radiosonde data is assimilated. In this case, we see that at every 12 hours the curves display a jump, resulting in an instantaneous reduction of errors. These jumps correspond to the analysis times, when the radiosonde observations are processed by the filter. Between two consecutive observation intervals, the errors grow due to the presence of the model error, represented by the matrix $\mathbf{Q}_k$ of Lecture 5 (see Cohn & Parrish [33] for more details on this quantity). Notice further in Fig. 9.2 that the errors in the meridional velocity and heights are below the radiosonde observational error levels (indicated by the curves marked OLV; numbers listed in Table 9.2). The errors in the zonal winds do not fall below the observational error level, which is a particular property of the solution of the shallow–water equations.

When the wind profilers are present (curves indicated as B), we see that during two consecutive A–network observation periods, errors decrease every hour due to the assimilation of these wind profilers. The presence of these extra wind observations produce an overall reduction in the errors in all variables, including heights which are not directly observed by the B–network. The contribution of the wind profilers to reducing the height errors is a consequence of the fact that the analysis procedure of the Kalman filter is multivariate, and moreover, that the Kalman filter transfers the information content in the wind profilers observations appropriately to the height fields.

Fig. 9.3 shows the spatial distribution of the forecast error standard deviations after 2.5 days in the assimilation cycle. The contour maps are built from the square root of the

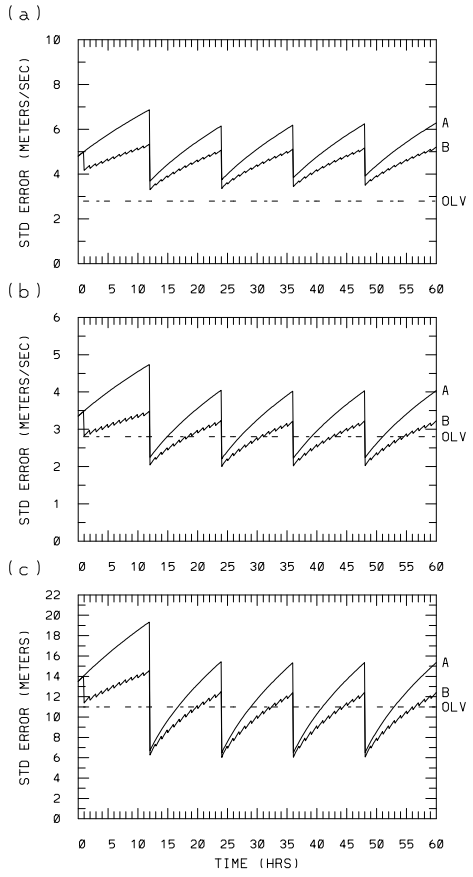

Figure 9.2: Results obtained with the Kalman filter: forecast and analysis expected error standard deviations averaged over the domain, for each of the three variables of the model, as a time function. The panels show the result for the errors in the (a) zonal velocity, (b) meridional velocity, and (c) height. Each curve is indicate by the observational system in question, that is, A for radiosondes and B for the radiosondes and wind profilers. The height errors are given in meters while the wind errors are in meters per second. The dotted lines, indicated by OLV, refer to the radiosondes observational error levels for each of the variables, according to Table 9.2.

elements of the main diagonal part of $\mathbf{P}_k^f$ for the winds and the height fields. The panels (a) refer to the experiment with the A–network, including only the radiosondes, while panels (b) refer to the experiment using both the A– and B–networks. In the case of the radiosonde–only assimilation [panels (a)] we see that the forecast error standard deviation practically uniformly distributed over the domain. This results from the fact that the radiosonde network is relatively uniformly distributed over the domain. The forecast error standard deviation in v, panel (a.2) has pronounced gradients near the North and South boundaries due to the boundary condition $v = 0$ along these boundaries. This boundary condition is equivalent to observing the variable v along of the North and South boundaries, without observation error. The presence of the dynamics allow for the consequent appearance of gradients along the boundaries in the zonal wind and height forecast error fields as seen in panels (a.1) e (a.2), respectively.

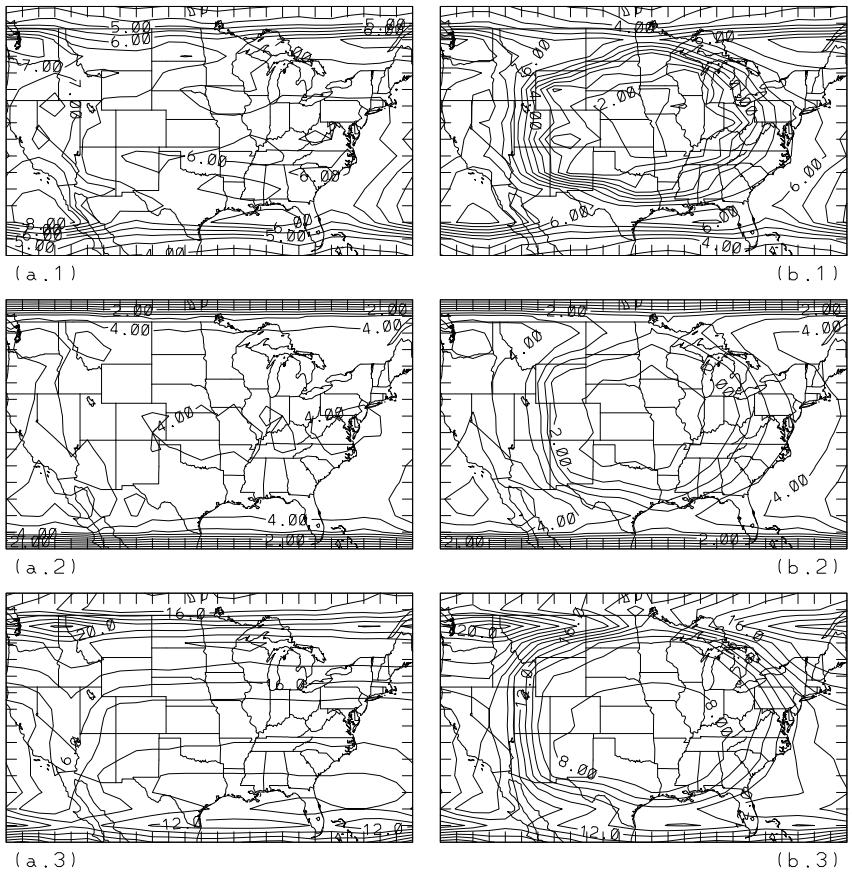

Figure 9.3: Spatial distribution of the forecast error standard deviation. Panels (a) refer to experiment A, while panels (b) refer to experiment (B): (a.1) and (b.1) for σ_u, (a.2) and (b.2) for σ_v, and (a.3) and (b.3) for σ_h. Contour interval is of 1 m for height errors, and of 0.5 m s^{-1} for wind errors.

The introduction of wind profilers, experiment with B–network, produces large gradients in the East–West direction, as seen in panels (b.1)–(b.3) of Fig. 9.3. These gradients result from the fact that the wind profilers are mainly located in the central region of the domain, contrary to the radiosondes, and therefore reflect the differences between regions of dense observation density and those of sparse observation density. We also see that in the panels

referring to the experiment in question, the Western error gradients are more pronounced than the Eastern ones. In other words, the contours close to the Eastern boundary are more separated among themselves than those close to the Western boundary [more markedly noticeable in panels (b.1) and (b.3), for the forecast error standard deviations of u and h, respectively]. This is a sole consequence of the error propagation induced by the Kalman filter equations, which makes the wind profilers information in the central part of the domain be advected in the direction of the flow. Therefore, the forecast error standard deviation is smaller in the East than in the West side of the wind profilers region. Specifically, this error propagation is due to the calculation of the first term in the expression (5.17), for $\mathbf{P}_k^f$.

It is important to mention that current operational data assimilation systems do not possess the ability to propagate information in the pronounced way seen in the results of the experiment using the B–network. The reason is, as discussed previously in this and in the previous lecture, that the forecast error covariance matrix in operational systems is assumed stationary and prescribed in some way to cope with computational feasibility. The incorporation of a dynamic flavor in the forecast error covariance matrix for operational systems has been the object of a great number of basic research. We can imagine, by what we studied so far, that an intensive line of research in atmospheric data assimilation is the search for alternative ways to propagate error covariance that are computationally feasible and do not involve as many computations as those following the Kalman filter, or its nonlinear extensions.

In what follows here we concentrate on alternatives to simplify expression F2 in Table 5.3.1, for the case of linear filter. Once the more viable alternatives for the linear case are determined, it is possible to make extensions to the nonlinear case, which represent in fact the cases of practical interest in meteorology. The majority of existing approximations in the literature belong to one, or more, or the following categories (see Todling & Cohn [129] for more references and explanations):

- covariance error modeling (e.g., OI: Bergman [10], Gandin [57], Jiang & Ghil [85], Lorenc [95], McPherson et al. [102]; SSI: Parrish & Derber [112]; three–dimensional variational analysis (3D–Var): Andersson et al. [2], Heckley et al. [74], Pailleux [110], Vasiljević et al. [133]; PSAS: da Silva et al. [37])
- dynamic simplification (e.g., Dee [43], Todling & Cohn [129])
- reduced resolution (order resolution; e.g., Cohn & Todling [34], Fukumori [54], Fukumori & Malanotte–Rizzoli [55], Le Moyne & Alvarez [92], Verlaan & Heemink [135])
- local representation (e.g., Boggs et al. [16], Cohn [28, 29], Parrish & Cohn [113], Riedel [118])
- limiting filtering (e.g., Fu et al. [53], Fukumori et al. [56], Heemink [75], Heemink & Kloosterhuis [76], Hoang et al. [77])
- Monte Carlo approach (e.g., Leith [93], Evensen [51])

Some of these possibilities have been tested in the context of the the Kalman filter applied to the linear shallow–water equations with the goal of investigating its behaviors in comparison with the exact result provided by the Kalman filter. We describe briefly below the

approximations considered in Todling & Cohn [129] for stable dynamics, and in Cohn & Todling [34] for stable and unstable dynamics. These approximations range from a simplified representation of the assimilation scheme by optimal interpolation [OI; item (a)], a somewhat improved version of OI which allows for advection of the height error covariance field by an advection operator $\mathbf{A}$ [HVA; section item (b)], to an even more sophisticated scheme that allows for the propagation of all height–height error covariance field by means of a simplified dynamics $\mathbf{A}$ [SKF; item (c)]. Since these schemes specify the height–height error covariance, it is necessary to use an algorithm to generate the missing covariances and cross–covariances. At this point we can impose the geostrophic balance constraint, however, as we have mentioned before, this does not generate good results, except in some cases. Alternatively, we can use the cross–variance generation algorithm studied previously in this lecture. Some of the schemes cited below use this procedure.

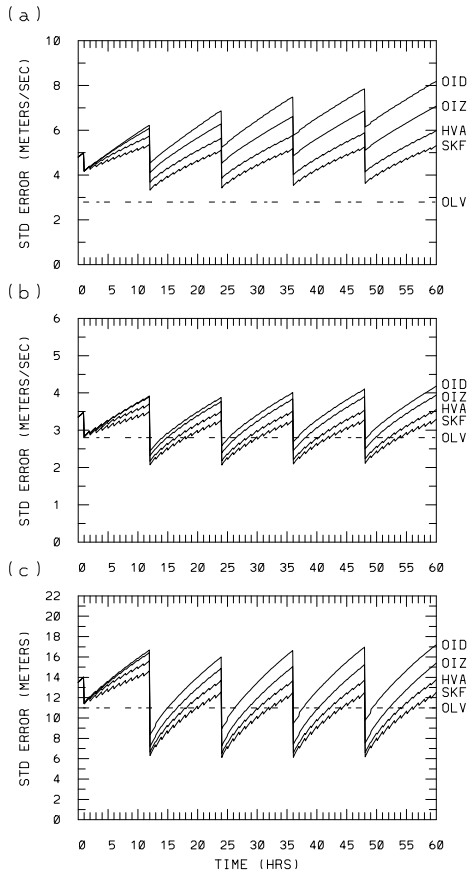

Figure 9.4: Analogous to Fig. 9.2, but only for the case of the B–network. The curves refer to the evaluation result of the performance of different approximations for the system of Cohn & Parrish. The approximations being evaluated are: balanced optimal interpolation with domain averaged error growth (OID); balanced optimal interpolation with latitudinal dependent error growth (OIZ); advection of balanced height error variance field (HVA); and advection of balanced height–height error covariance field (SKF). The Kalman filter results, that serve as the basis of comparison to these approximations are those indicated by the curves B in Fig. 9.2.

A summary of the results of evaluation of these approximations is presented in Fig. 9.4.

Considering the system of Cohn & Parrish [33], for the B–network case, the curves in the figure show the performance of two types of OI: OID, which uses a single constant value for the height error variance growth rate (elements along the diagonal of matrix $\mathbf{D}^h$ mentioned in item (a) below); and OIZ, which uses a height variance growth rate with latitudinal dependency. Also, the performance of the two other schemes, the HVA and the SKF, is concisely described in items (b) and (c) below. The figure shows the gradual improvement occurring when we increase the sophistication of the assimilation scheme, due to the gradual incorporation of dynamic information. In particular, the performance of the SKF scheme is practically indistinguishable from the exact result shown in Fig. 9.2 (curves B).

The way in which the schemes OI, HVA and SKF incorporate some dynamics (or not, in the case of OI), in the generation of forecast error covariances is very simplistic. To test these simplifications further an unstable dynamical system has been considered. Although the Kalman filter for unstable linear systems produces reliable results (see Ghil & Todling [65] and Todling & Ghil [130]), approximate assimilation schemes, based on the ideas described above, do not produce results equally reliable. In this way, alternative data assimilation schemes are necessary; some possibilities are briefly described in the last items (d) and (e) below. Both schemes described in these items are iterative. An Lanczos type algorithm (e.g, Golub & Van Loan [67]) is necessary to implement either schemes. The partial singular value decomposition filter (PSF) [item (d)] proposes to use the L leading singular modes of the propagator (tangent linear model) to propagate the analysis error covariance matrix; the partial eigendecomposition filter (PEF) [item (e)] proposes to generate only L leading eigenvalues/vector of the forecast error covariance matrix, $\mathbf{S}^p$. These two schemes are low–rank and information referring to the trailing part of the error covariances, which we designate by $\mathbf{S}^p_T$, should be provided in some way. That is, in these two cases there is a need to model trailing error covariances. These approximations are adequate also for stable dynamics.

(a) Optimal Interpolation

Category: Error Covariance Modeling

Forecast Error Covariance Matrix Partition $\mathbf{S}^f_k$:

$$\mathbf{S}^f_k \equiv \begin{bmatrix} \mathbf{S}^{f|uu} & \mathbf{S}^{f|uv} & \mathbf{S}^{f|uh} \\ \mathbf{S}^{f|vu} & \mathbf{S}^{f|vv} & \mathbf{S}^{f|vh} \\ \mathbf{S}^{f|hu} & \mathbf{S}^{f|hv} & \mathbf{S}^{f|hh} \end{bmatrix}_k$$

Block corresponding to the forecast error covariance (hh):

$$\mathbf{S}^{f|hh}_k = (\mathbf{D}^{f|h}_k)^{1/2} \mathbf{C}^{hh} (\mathbf{D}^{f|h}_k)^{1/2}$$

$\mathbf{D}^{f|h}_k$ is a diagonal matrix $n/3 \times n/3$ corresponding to the height error variance; $\mathbf{C}^{hh}$ is the height–height error correlation matrix $n/3 \times n/3$, which is prespecified.

$\mathbf{D}^{f|h}_k$ is in general modeled to account for a linear variance error growth in time according to

$$\mathbf{D}^{f|h}_k \;=\; \mathbf{D}^{a|h}_{k-\ell} \;+\; \mathbf{D}^h$$

where ℓ is the number of model time steps between two consecutive analyses; $\mathbf{D}^h$ is a diagonal matrix corresponding to the error growth.

$\mathbf{C}^{hh}$ is prespecified assuming homogeneity and isotropy of the mass error field, and it is usually considered to be Gaussian. The rest of the error covariance matrix is obtained by the balanced covariance generation procedure described previously.

(b) Variance Evolution

Category: Local representation

Height error variance propagation:

$$\mathbf{D}_k^{p|h} = \mathbf{A}_{k,k-\ell}\mathbf{D}_{k-\ell}^{a|h}$$

where $\mathbf{A}_{k,k-\ell}$ represents the operator of an advection scheme.

Construction of the height–height error covariance matrix:

$$\mathbf{S}_k^{p|hh} = (\mathbf{D}_k^{p|h})^{1/2}\mathbf{C}^{hh}(\mathbf{D}_k^{p|h})^{1/2}\,.$$

The covariances remaining are calculated by means of the balanced covariance generation procedure discussed above .

(c) Simplified Kalman Filter

Category: dynamic simplification

Propagation of Height–height error covariance:

$$\mathbf{S}_k^{p|hh} = \mathbf{A}_{k,k-\ell}\,\mathbf{S}_{k-\ell}^{a|hh}\,\mathbf{A}_{k,k-\ell}^T$$

where $\mathbf{A}_{k,k-\ell}$ represents the operator of an advection scheme; the balanced covariances generation procedure is used for the remainder of the covariances.

(d) Partial Singular Value Decomposition Filter

Category: Local representation/reduced resolution

Forecast error covariance:

$$\mathbf{S}_k^f = \mathbf{S}_{k,k-\ell}^p + \tilde{\mathbf{Q}}_{k,k-\ell}$$

where $\mathbf{S}^p_{k,k-\ell}$ is the dynamically propagated part — predictability error covariance.

Consider the following singular value decomposition of the propagator $\mathbf{\Psi}_{k,k-\ell}$:

$$\mathbf{\Psi}_{k,k-\ell} = \left(\mathbf{U}\,\mathbf{D}\,\mathbf{V}^T\right)_{k,k-\ell}$$

and partition the matrices above in leading (L) and trailing (T) parts so that:

$$\mathbf{U}_{k,k-\ell} \;=\; [\mathbf{U}_L\,\mathbf{U}_T]_{k,k-\ell}\;, \qquad \mathbf{V}_{k,k-\ell} \;=\; [\mathbf{V}_L\,\mathbf{V}_T]_{k,k-\ell}$$

$$\mathbf{D}_{k,k-\ell} \;=\; \left[\begin{array}{cc} \mathbf{D}_L & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_T \end{array}\right]_{k,k-\ell}$$

Consider the following model for $\mathbf{S}^p_{k,k-\ell}$:

$$\mathbf{S}^p_{k,k-\ell} = (\mathbf{S}^p_L + \mathbf{S}^p_T)_{k,k-\ell} = \left(\tilde{\mathbf{\Psi}}\,\mathbf{S}^a\,\tilde{\mathbf{\Psi}}^T + \mathbf{S}^p_T\right)_{k,k-\ell}$$

where

$$\tilde{\mathbf{\Psi}}_{k,k-\ell} = \left(\mathbf{U}_L\,\mathbf{D}_L\,\mathbf{V}_L^T\right)_{k,k-\ell}$$

so that,

$$\mathbf{S}^p_{L,k,k-\ell} = \left(\sum_{i=1}^{N_L}\sum_{j=1}^{N_L} d_i d_j \left(\mathbf{v}_i^T\mathbf{S}^a\mathbf{v}_j\right)\,\mathbf{u}_i\mathbf{u}_j^T\right)_{k,k-\ell}$$

where $d_i = diag(\mathbf{D}_L)_i$, $\mathbf{u}_i = col(\mathbf{U}_L)_i$, $\mathbf{v}_i = col(\mathbf{V}_L)_i$,
e $N_L = n\underline{o}\,cols(\mathbf{V}_L)$.

$\mathbf{S}^p_T$ is specified by an adaptively tuned covariance model based on the innovations. Computational cost $\sim o(10L)$ model integrations.

(e) Partial Eigendecomposition Filter

Category: Local representation/reduced resolution

Forecast covariance error:

$$\mathbf{S}^f_k = \mathbf{S}^p_{k,k-\ell} + \tilde{\mathbf{Q}}_{k,k-\ell}$$

where $\mathbf{S}^p_{k,k-\ell}$ is the dynamically propagated part — predictability error covariance.

Consider the following eigendecomposition for the forecast error covariance $\mathbf{S}^p$:

$$\mathbf{S}^p_{k,k-\ell} = \left(\mathbf{\Psi}\,\mathbf{S}^a\,\mathbf{\Psi}^T\right)_{k,k-\ell} = \left(\mathbf{U}\mathbf{D}\mathbf{U}^T\right)_{k,k-\ell}$$

and partition the factors above in leading (L) and trailing (T) parts:

$$\mathbf{U}_{k,k-\ell} \;=\; [\mathbf{U}_L\,\mathbf{U}_T]_{k,k-\ell}\;,$$

$$\mathbf{D}_{k,k-\ell} = \begin{bmatrix} \mathbf{D}_L & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_T \end{bmatrix}_{k,k-\ell}$$

Assume the following approximation for $\mathbf{S}^p_{k,k-\ell}$:

$$\mathbf{S}^p_{k,k-\ell} = (\mathbf{S}^p_L + \mathbf{S}^p_T)_{k,k-\ell} = \left(\mathbf{U}_L\, \mathbf{D}_L\, \mathbf{U}_L^T + \mathbf{S}^p_T\right)_{k,k-\ell}$$

$\mathbf{S}^p_T$ is specified for an adaptively tuned covariance model based on the innovations. Computational cost $\sim o(10L)$ model integrations.

The motivation to search for approximate schemes for covariance propagation, instead of ways of implementing an algorithm for complete error covariance propagation, goes beyond the fact that the latter is computationally infeasible. In fact, even if calculating the complete covariance evolution were (or comes to be) feasible, it would be a wast of computational resources, since:

- The governing equations are nonlinear, consequently only approximate schemes are possible — in general, we cannot calculate moments of all orders.
- Many observational systems involve nonlinear relations among the forecast and observed variables. Therefore, the same problem mentioned in the item above applies, i.e., we cannot calculate moments of all orders.
- Lack of knowledge of model and observation error statistics. This means that, at best we have to approximate, model and parameterize these quantities, forcing the assimilation system in to a sub–optimal situation.
- The number of available observations in a given time is relatively close to the number of degrees of freedom of the system, and it tends to become even greater. This means that the equations for the analysis step, that is, equations F3, F5 of Table 5.3.1, or their nonlinear equivalents, should be approximated in some way, due to the excess of computational effort required to solve them exactly.

Even the promising developments in parallel computation [78, 79, 100], and the possibility of an increase in computational capability, will not be sufficient to solve the covariance propagation equation completely. This because the tendency in meteorology has been to increase the resolution of general circulation models, whenever there is an increase in computer power. Thus, pushing computers to their limit, just to produce a forecast, leaving little room for forecasting the statistics . It will always be necessary to have approximate models for error covariance propagation, to make operational systems for atmospheric data assimilation computational feasible and practical.

9.3 The Fixed–Lag Kalman Smoother

9.3.1 Theory

Let us address the issue of improving an estimate provided by the Kalman filter by making use of observations *past* the analysis time. Let us use the Bayesian approach for this purpose, but to keep the derivation simple we consider only the problem of improving the estimate at time t_{k-1} given the observations at just one time step ahead that time, that is, at time t_k. This constitutes the lag–1 smoother problem (this can be identified with the fixed–point smoother). We could extend the problem to that of improving the filter state estimate at time $t_{k-\ell}$ using observations up to and including time t_k. This is the general, lag–ℓ[1], fixed–lag smoother problem. The solution to the general problem can be found in a variety of texts, as for example, Anderson & Moore [1] and Meditch [103].

In complete analogy to the previous section we can seek for the minimum variance estimate, which written in terms of the conditional mean is

$$\begin{aligned} \mathbf{w}^a_{k-1|k} &= \mathcal{E}\{\mathbf{w}^t_{k-1}|\mathbf{W}^o_k\} \\ &= \int_{-\infty}^{+\infty} \mathbf{w}^t_{k-1} p(\mathbf{w}^t_{k-1}|\mathbf{W}^o_k)\, d\mathbf{w}^t_{k-1} \end{aligned} \tag{9.54}$$

where we use the notation $j|k$ to indicate the estimate at time t_j conditioned on observations up to and including time t_k. Notice that in this notation the analysis estimate provided by the Kalman filter can be indicated as $\mathbf{w}^a_{k|k}$, and the forecast estimate can be indicated as $\mathbf{w}^f_{k|k-1}$. Similarly, the analysis and forecast error covariances can be indicated by $\mathbf{P}^a_{k|k}$ and $\mathbf{P}^f_{k|k-1}$, respectively. Once again, the fundamental quantity to determine is the conditional probability density $p(\mathbf{w}^t_{k-1}|\mathbf{W}^o_k)$ in the expression above.

Repeated use of the definition of conditional probability gives

$$\begin{aligned} p(\mathbf{w}^t_{k-1}|\mathbf{W}^o_k) &= p(\mathbf{w}^t_{k-1}|\mathbf{w}^o_k, \mathbf{W}^o_{k-1}) \\ &= \frac{p(\mathbf{w}^t_{k-1}, \mathbf{w}^o_k, \mathbf{W}^o_{k-1})}{p(\mathbf{w}^o_k, \mathbf{W}^o_{k-1})} \\ &= \frac{p(\mathbf{w}^o_k|\mathbf{w}^t_{k-1}, \mathbf{W}^o_{k-1}) p(\mathbf{w}^t_{k-1}, \mathbf{W}^o_{k-1})}{p(\mathbf{w}^o_k, \mathbf{W}^o_{k-1})} \\ &= \frac{p(\mathbf{w}^o_k|\mathbf{w}^t_{k-1}, \mathbf{W}^o_{k-1}) p(\mathbf{w}^t_{k-1}|\mathbf{W}^o_{k-1}) p(\mathbf{W}^o_{k-1})}{p(\mathbf{w}^o_k|\mathbf{W}^o_{k-1}) p(\mathbf{W}^o_{k-1})} \\ &= \frac{p(\mathbf{w}^o_k|\mathbf{w}^t_{k-1}, \mathbf{W}^o_{k-1}) p(\mathbf{w}^t_{k-1}|\mathbf{W}^o_{k-1})}{p(\mathbf{w}^o_k|\mathbf{W}^o_{k-1})}, \\ &= \frac{p(\mathbf{w}^o_k|\mathbf{w}^t_{k-1}) p(\mathbf{w}^t_{k-1}|\mathbf{W}^o_{k-1})}{p(\mathbf{w}^o_k|\mathbf{W}^o_{k-1})}, \end{aligned} \tag{9.55}$$

[1] This ℓ here is not to be confused with the ℓ used earlier in this notes to denote the number of time steps between two consecutive model time steps. Remember that the number of model time steps between consecutive observations has been fixed to one from Fig. 5.1 on.

where the last equality is obtained after noticing that the observation sequence is white in time. In this equation we recognize the denominator as being the same denominator as that in (5.27) corresponding to the probability density function of the innovation vector. Moreover, changing k into $k-1$ in (5.38) we can identify $p(\underline{\mathbf{w}}_1^t|\mathbf{W}_{k-1}^o)$ as the probability density function of the filter analysis at time t_{k-1}. Thus, the only quantity remaining to be calculated in the expression above is the first term in the numerator.

Because the statistics of all errors and initial state are Gaussian, the probability density $p(\mathbf{w}_k^o|\mathbf{w}_{k-1}^t)$ is also Gaussian and can be written as

$$p(\mathbf{w}_k^o|\mathbf{w}_{k-1}^t) = \frac{1}{(2\pi)^{m_k/2}|\tilde{\mathbf{R}}_k|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{w}_k^o - \mathcal{E}\{\mathbf{w}_k^o|\mathbf{w}_{k-1}^t\})^T(\tilde{\mathbf{R}}_k)^{-1}(\mathbf{w}_k^o - \mathcal{E}\{\mathbf{w}_k^o|\mathbf{w}_{k-1}^t\})\right] \tag{9.56}$$

where $\tilde{\mathbf{R}}_k$ denotes the following conditional error covariance

$$\tilde{\mathbf{R}}_k \equiv \mathcal{E}\{[\mathbf{w}_k^o - \mathcal{E}\{\mathbf{w}_k^o|\mathbf{w}_{k-1}^t\}][\mathbf{w}_k^o - \mathcal{E}\{\mathbf{w}_k^o|\mathbf{w}_{k-1}^t\}]^T|\mathbf{w}_{k-1}^t\}\,. \tag{9.57}$$

Using (5.1) and (5.3) we can calculate the conditional mean above as

$$\begin{aligned}\mathcal{E}\{\mathbf{w}_k^o|\mathbf{w}_{k-1}^t\} &= \mathcal{E}\{(\mathbf{H}_k\mathbf{w}_k^t + \mathbf{v}_k)|\mathbf{w}_{k-1}^t\} \\ &= \mathcal{E}\{(\mathbf{H}_k\boldsymbol{\Psi}_{k-1}\mathbf{w}_{k-1}^t + \mathbf{H}_k\mathbf{b}_{k-1}^t + \mathbf{v}_k)|\mathbf{w}_{k-1}^t\} \\ &= \mathbf{H}_k\boldsymbol{\Psi}_{k-1}\mathbf{w}_{k-1}^t \end{aligned} \tag{9.58}$$

where we noticed that the noise sequences $\{\mathbf{b}_k^t\}$ and $\{\mathbf{v}_k\}$ have mean zero. Similarly, it follows that the conditional error covariance $\tilde{\mathbf{R}}_k$ is given by

$$\tilde{\mathbf{R}}_k = \mathbf{H}_k\mathbf{Q}_{k-1}\mathbf{H}_k^T + \mathbf{R}_k \tag{9.59}$$

This completes the amount of information required to fully determine the probability density (9.56) and consequently the probability density (9.55).

Substituting (5.27) with $k \to k-1$, (5.31) and (9.56) in (9.55) it follows that

$$p_(\mathbf{w}_{k-1}^t|\mathbf{W}_k^o) = \frac{|\mathbf{H}_k\boldsymbol{\Psi}_{k-1}\mathbf{P}_{k-1}^a\boldsymbol{\Psi}_{k-1}^T\mathbf{H}_k^T + \tilde{\mathbf{R}}_k|^{1/2}}{(2\pi)^{n/2}|\mathbf{P}_{k-1}^a|^{1/2}|\tilde{\mathbf{R}}_k|^{1/2}} \exp[-\frac{1}{2}J] \tag{9.60}$$

where J, in this case, is defined as

$$\begin{aligned} J(\mathbf{w}_{k-1}^t) \equiv\ & (\mathbf{w}_k^o - \mathbf{H}_k\boldsymbol{\Psi}_{k-1}\mathbf{w}_{k-1}^t)^T\tilde{\mathbf{R}}_k^{-1}(\mathbf{w}_k^o - \mathbf{H}_k\boldsymbol{\Psi}_{k-1}\mathbf{w}_{k-1}^t) \\ & + (\mathbf{w}_{k-1}^t - \mathbf{w}_{k-1}^a)^T(\mathbf{P}_{k-1}^a)^{-1}(\mathbf{w}_{k-1}^t - \mathbf{w}_{k-1}^a) \\ & - (\mathbf{w}_k^o - \mathbf{H}_k\mathbf{w}_k^f)^T(\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^T + \mathbf{R}_k)^{-1}(\mathbf{w}_k^o - \mathbf{H}_k\mathbf{w}_k^f) \\ =\ & (\mathbf{w}_k^o - \mathbf{H}_k\boldsymbol{\Psi}_{k-1}\mathbf{w}_{k-1}^t)^T\tilde{\mathbf{R}}_k^{-1}(\mathbf{w}_k^o - \mathbf{H}_k\boldsymbol{\Psi}_{k-1}\mathbf{w}_{k-1}^t) \\ & + (\mathbf{w}_{k-1}^t - \mathbf{w}_{k-1}^a)^T(\mathbf{P}_{k-1}^a)^{-1}(\mathbf{w}_{k-1}^t - \mathbf{w}_{k-1}^a) \\ & - (\mathbf{w}_k^o - \mathbf{H}_k\boldsymbol{\Psi}_{k-1}\mathbf{w}_{k-1}^a)^T(\mathbf{H}_k\boldsymbol{\Psi}_{k-1}\mathbf{P}_{k-1}^a\boldsymbol{\Psi}_{k-1}^T\mathbf{H}_k^T + \tilde{\mathbf{R}}_k)^{-1}(\mathbf{w}_k^o - \mathbf{H}_k\boldsymbol{\Psi}_{k-1}\mathbf{w}_{k-1}^a) \end{aligned} \tag{9.61}$$

where the first two terms in J come from the two terms in the numerator of (9.55), respectively, while the last term in J comes from the denominator of (9.55).

Hence, the the probability density function in (9.54) can be written as

$$p(\mathbf{w}_{k-1}^t|\mathbf{W}_k^o) \;=\; \frac{1}{(2\pi)^{n/2}|\mathbf{P}_{k-1|k}^a|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{w}_{k-1}^t - \mathbf{w}_{k-1|k}^a)^T(\mathbf{P}_{k-1|k}^a)^{-1}(\mathbf{w}_{k-1}^t - \mathbf{w}_{k-1|k}^a)\right] \tag{9.62}$$

where its maximum $\mathbf{w}_{k-1|k}^a$ corresponds to the estimate we are seeking and is given by

$$\mathbf{w}_{k-1|k}^a \;=\; \mathbf{w}_{k-1}^a \;+\; \mathbf{P}_{k-1}^a \boldsymbol{\Psi}_{k-1}^T \mathbf{H}_k^T \boldsymbol{\Gamma}_k^{-1}(\mathbf{w}_k^o - \mathbf{H}_k \mathbf{w}_k^f)\,, \tag{9.63}$$

with corresponding error covariance

$$(\mathbf{P}_{k-1|k}^a)^{-1} \;=\; (\mathbf{P}_{k-1}^a)^{-1} \;+\; \boldsymbol{\Psi}_{k-1}^T \mathbf{H}_k^T \tilde{\mathbf{R}}_k^{-1} \mathbf{H}_k \boldsymbol{\Psi}_{k-1}\,. \tag{9.64}$$

These quantities are sometimes referred to as the retrospective analysis and the retrospective analysis error covariance matrix, respectively (see Todling et al. [131]).

In complete analogy to the remark made when introducing the maximum *a posteriori* functional J_{MAP} in (4.65), we could derive the same results in (9.63) and (9.64) by minimizing the cost function J_{MAP} for this case, that is,

$$\begin{aligned} J_{\text{MAP}}(\mathbf{w}_{k-1}^t) \;\equiv\; & (\mathbf{w}_k^o - \mathbf{H}_k \boldsymbol{\Psi}_{k-1} \mathbf{w}_{k-1}^t)^T \tilde{\mathbf{R}}_k^{-1} (\mathbf{w}_k^o - \mathbf{H}_k \boldsymbol{\Psi}_{k-1} \mathbf{w}_{k-1}^t) \\ & + (\mathbf{w}_{k-1}^t - \mathbf{w}_{k-1}^a)^T (\mathbf{P}_{k-1}^a)^{-1} (\mathbf{w}_{k-1}^t - \mathbf{w}_{k-1}^a)\,. \end{aligned} \tag{9.65}$$

As before, this corresponds to considering only the probability distributions in the numerator of (9.60).

An alternative expression to (9.64) can be obtained using the Sherman–Morrison–Woodbury formula (c.f., Golub & Van Loan [67], p. 51), which gives

$$\begin{aligned} \mathbf{P}_{k-1|k}^a \;&=\; \mathbf{P}_{k-1}^a \;-\; \mathbf{P}_{k-1}^a \boldsymbol{\Psi}_{k-1}^T \mathbf{H}_k^T \boldsymbol{\Gamma}_k^{-1} \mathbf{H}_k \boldsymbol{\Psi}_{k-1} \mathbf{P}_{k-1}^a \\ &=\; \mathbf{P}_{k-1}^a \;-\; \mathbf{K}_{k-1|k} \mathbf{H}_k \mathbf{P}_{k,k-1|k-1}^{fa} \end{aligned} \tag{9.66}$$

where we introduced the following definitions, to write the last equality:

$$\mathbf{P}_{k|k-1}^{fa} \;\equiv\; \boldsymbol{\Psi}_{k-1} \mathbf{P}_{k-1}^a \tag{9.67a}$$

$$\mathbf{K}_{k-1|k} \;\equiv\; (\mathbf{P}_{k|k-1}^{fa})^T \boldsymbol{\Gamma}_k^{-1} \tag{9.67b}$$

Observer that the advantage of expression (9.66) over (9.64) for $\mathbf{P}_{k-1|k}^a$ is that (9.66 does not involve the inverse of usually large error covariance matrices. Also, with the definitions in (9.67), the expression for the lag–1 smoother estimate (9.63) becomes

$$\mathbf{w}_{k-1|k}^a \;=\; \mathbf{w}_{k-1}^a \;+\; \mathbf{K}_{k-1|k}(\mathbf{w}_k^o - \mathbf{H}_k \mathbf{w}_k^f) \tag{9.68}$$

Therefore, the smoother analysis at time t_{k-1} using data up to an including t_k corresponds to an update of the filter analysis $\mathbf{w}_{k-1}^a$ at time t_{k-1}, based on the same innovation vector used to calculated the filter analysis at the time of the latest observation, t_k.

The procedure above can be generalized to any number of lags larger than one, but the probabilistic framework above does not provide the simplest method for this generalization. A much simpler way is to use the approach of state augmentation combined with the

minimum variance approach we described earlier in this notes. This procedure is outlined in Anderson & Moore [1], and it is explicitly invoked in Todling & Cohn [128] to derived the nonlinear extended fixed–lag smoother. Here, we only list the equations that needed to be supplemented to the linear Kalman filter so it becomes the linear fixed–lag Kalman smoother:

$$\mathbf{w}^a_{k-\ell|k} = \mathbf{w}^a_{k-\ell|k-1} + \mathbf{K}_{k-\ell|k}(\mathbf{w}^o_k - \mathbf{H}_k\mathbf{w}^f_k) \tag{9.69a}$$

$$\mathbf{P}^{aa}_{k,k-\ell|k} = \left(\mathbf{I} - \mathbf{K}_{k|k}\mathbf{H}_k\right)\mathbf{P}^{fa}_{k,k-\ell|k-1} \tag{9.69b}$$

$$\mathbf{P}^a_{k-\ell|k} = \mathbf{P}^a_{k-\ell|k-1} - \mathbf{K}_{k-\ell|k}\mathbf{H}_k\mathbf{P}^{fa}_{k,k-\ell|k-1} \tag{9.69c}$$

where the generalization of definitions (9.67) are

$$\mathbf{P}^{fa}_{k,k-\ell|k-1} = \mathbf{\Psi}_{k-1}\mathbf{P}^{aa}_{k-1,k-\ell|k-1} \tag{9.70a}$$

$$\mathbf{K}_{k-\ell|k} = \left(\mathbf{P}^{fa}_{k,k-\ell|k-1}\right)^T\mathbf{H}^T_k\mathbf{\Gamma}^{-1}_k \tag{9.70b}$$

respectively. In applying these expressions to the case we have just derived of $\ell = 1$ we must recognize the fact that we used a compact notation for the filtering problem, since in that case it avoided unnecessary complexity. Explicitly, when using $\ell = 1$ in (9.69a) we are confronted with the estimate $\mathbf{w}^a_{k-1|k-1}$, on the right–hand–side of the equation, which ultimately arises from the filtering problem. The filter estimate at time t_{k-1} is conditioned on all observations up to and including time t_{k-1}, which is just (5.24),

$$\mathbf{w}^a_{k-1|k-1} \equiv \mathcal{E}\{\mathbf{w}^t_{k-1}|\mathbf{W}^o_{k-1}\} = \mathbf{w}^a_{k-1} \tag{9.71}$$

which reduces to the simpler notation used in the filtering problem. A similar argument applies to the analysis error covariance $\mathbf{P}^a_{k-1|k-1}$ we encounter when substituting $\ell = 1$ in (9.69b) and (9.69c), i.e., we must realize that $\mathbf{P}^a_{k-1|k-1} = \mathbf{P}^a_{k-1}$. In this case, however, not using the notation with the conditioning explicitly written is more than just notational simplification for the filtering problem. It represents the fact that the filter error covariances are not conditioned on the data, as we observed in (5.41) and (5.42).

9.3.2 Application to a Linear Shallow-Water Model

Let us now examine the results of the fixed–lag Kalman smoother applied to the linear shallow–water model of the previous section. The interest here is to improve up on previously calculated filter analysis by using data past the analysis time. Following Cohn et al. [35], we consider the case of the A–network introduced above, but to show the more stringent results from that work we consider the case in which only the western half of the radiosondes are used in the assimilation experiments.

Figure 9.5 displays the time evolution of the domain–averaged root–mean–square errors for a period of 10 days of assimilation. The figure depicts only to the analysis errors, for all three variables of the model, in contrast to Fig. 9.2 where both the analysis and forecast errors are displayed, for these variables. This means that whenever comparing both figures, we should only care about the lower envelope of the curves in Fig. 9.2, corresponding to the analysis errors. The results in Fig. 9.5 are for both the Kalman filter and the Kalman

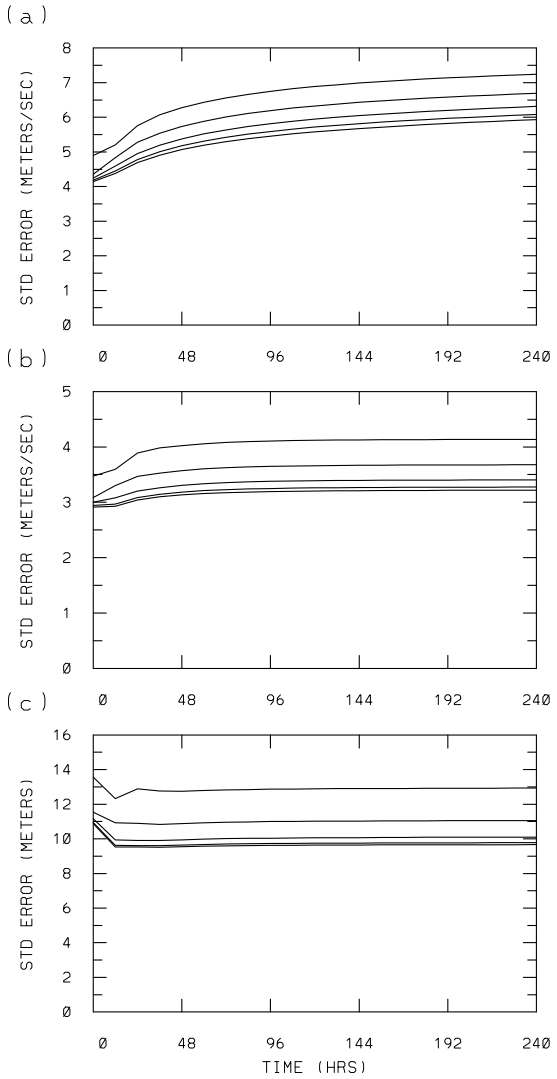

Figure 9.5: Expected analysis error standard deviations averaged over the domain, for each variable of the model, as a function of time. Panels are ordered as in Fig. 9.2, and results are for the fixed–lag Kalman smoother using only the western half of the radiosonde observations displayed in Fig. 9.1. The uppermost curve in each panel corresponds to the Kalman filter analysis (analogous to the lower envelope of the curves in Fig. 9.2), and successively lower curves are for the retrospective analysis results for 12–hour lags $\ell = 1, 2, 3$ and 4.

smoother for up to lag $\ell = 4$. The highest curve on each panel correspond to the filter errors. When compared against the corresponding curves (labeled A) in Fig. 9.2, we see that the errors are now larger, for all variables, than they were. This is a consequence of the fact that the number of observations is roughly one–half of what it is in the experiment with the A–network of Fig. 9.2. Successive lower curves, than the filter curve, in each of the panels in Fig. 9.5 refer to successive retrospective analyses obtained using data $12, 24, 36$, and 48 hours ahead of the analysis time. These correspond to the fixed–lag Kalman smoother results for lags $\ell = 1, 2, 3,$ and 4, respectively. The improvement achieved by the consecutive retrospective analyses is clearly seen by the decrease in the errors with the increase of the lags.

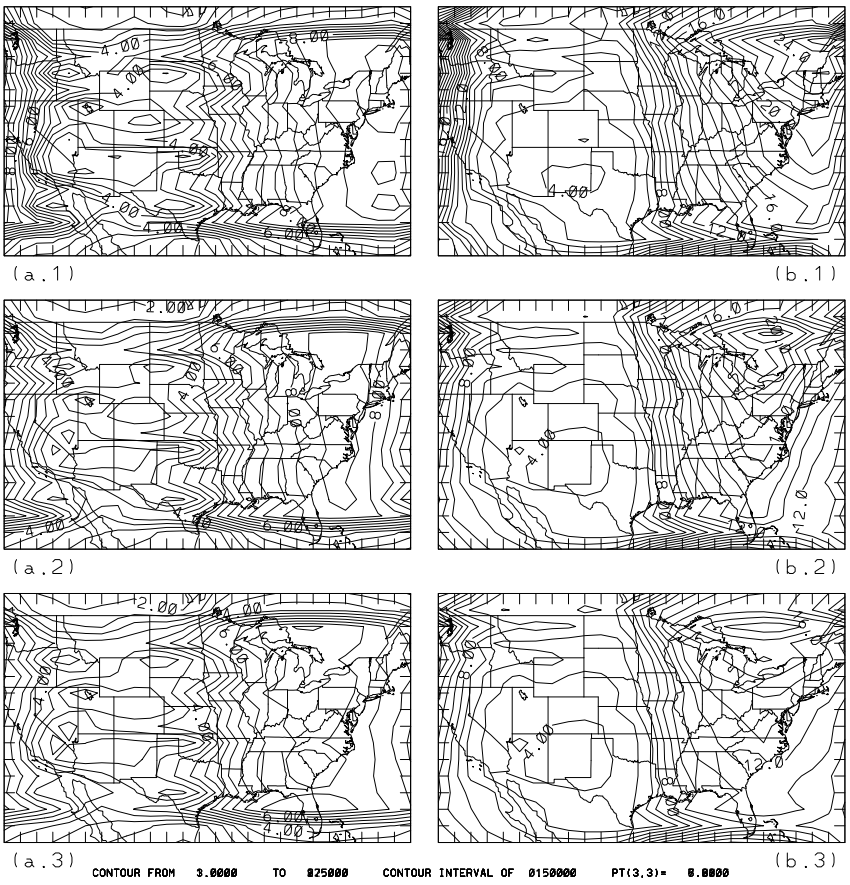

Figure 9.6: Maps of analysis error standard deviations at day 2 of the assimilation period for the fixed–lag Kalman smoother. Panels (a.1)–(a.3) refer to errors in u for the Kalman filter and Kalman smoother for lags $\ell = 1$ and 4, respectively; panels (b.1)–(b.3) refer to errors in h for the Kalman filter and Kalman smoother for lags $\ell = 1$ and 4, respectively.

To further illustrate the analysis improvement due to smoothing we show in Fig. 9.6 maps of the analysis errors in the u (left panels) and h (right panels) fields, at day 2. The top–two panels are for the filter analysis, the central–two panels for lag $\ell = 1$, i.e., when the smoother uses data at day 2.5 to further correct the filter analysis at day 2, and the bottom–two panels are for lag $\ell = 4$, that is, when data between days 2.5 and 4 have been used to correct the filter analysis at day 2. Looking at the filter results [panels (a.1) and (b.1)], we see that the large errors are located over the eastern part of the do-

main, where there are no radiosondes in this case. This is more dramatically seen from the height error fields, but the same is true for all fields, including the errors in v (not shown). Just as we encountered before for the case of Fig. 9.3, the contours in the top–two maps reflect the change in the observational data density. Moreover, the larger separation among the contours as we look from West to East reflect the advection of errors in the direction of the flow. Looking at the retrospective analyses results in the central–two and bottom–two panels, we observe an overall reduction of errors, especially for the lag $\ell = 4$ case, when compared against the top–two panels in the figure. However, the most striking feature of all in the figure is the propagation of the "region" of maximum analysis errors over the data void from East to West, as we look down from the top panels in the figure to the bottom panels, in each column. The magnitude of maximum errors not only gets reduced but also propagates against the flow. This illustrates the ability of the fixed–lag smoother to propagate information upstream (see also Ménard & Daley [105]).

Exercise

The Kalman filter applied to a linear advection equation. Consider the one–dimensional advection equation

$$\frac{\partial u}{\partial t} + U\frac{\partial u}{\partial x} = 0$$

where $U = const.$ represents the advection speed, applied to a periodic domain defined by the interval $[-2, 2]$ over the line. Take for initial condition a "rectangular" wave of the form

$$u(x, t = 0) = \begin{cases} 1, & \text{for } -1 \leq x \leq 0 \\ 0, & \text{otherwise} \end{cases}$$

Using an up–wind finite difference scheme we can write an approximate solution to the advection equation as

$$v_j = Cv_{j-1} + (1 - C)v_j$$

where v_j represents the numeric solution for $u(x = j\Delta x)$ with Δx as the spatial interval, and where $C = U\Delta t/\Delta x$ is the Courant number, with time step Δt.

Simulation experiments: Using the parameters in the table below, obtain plots of the state evolution at the initial and final times for an integration taken from $T_0 = 0$ to $t_{final} = 1$ using the following Courant numbers: (i) $C = 1$, $C = 0.95$, and $C = 0.95$. Explain the difference in the results.

Table 9.3: Parameters for the finite–difference.

Domain	for $-2 \leq x \leq 2$
Mesh size	$\Delta x = 0.2$
Time step	$\Delta t 0.05$

Let us now slowly build the components to have a Kalman filter assimilation experiment constructed. We assume that all error statistics necessary for the filter are Gaussian and white. What we have to do next is to construct error covariance matrices for all stochastic

processes involved in the problem. For simplicity, we take the perfect model assumption, so that we do not have to worry about the model error $\mathbf{b}_k^t$ and its error covariance $\mathbf{Q}_k$, i.e., $\mathbf{b}_k^t = 0$ at all times. Moreover, we assume there are no correlations among observation errors, so that the observation error covariance matrix $\mathbf{R}_k$ is diagonal. Furthermore, this matrix is assumed to be time independent with elements alone the diagonal to be specified depending of the experiments to be performed below under these conditions. The only error covariance left to specify it that of the initial estimate, $\mathbf{P}_0^a$. Constructing spatial error covariances that really satisfy the requirements of being an error covariance can be a pretty delicate issue. Instead of getting into these problems, we use a Matlab function gcorr provided with this exercise to construct an appropriate error correlation field that can be used to generate the require covariance matrix. Use the help of this function to see its usage and perform the following tasks for three different choices of the (de)correlation length parameter $L_d = 0.5, 1$ and 1.5 the following plots:

1. Is the matrix you constructed an acceptable correlation matrix?
2. Plot the two–point correlation function at two distinct arbitrary locations. Comment on what you see.
3. Make a contour plot of the correlation matrix. What you will see corresponds to the "shape" of a homogeneous and isotropic correlation matrix.

Now write a Matlab program with the Kalman filter equations for the state estimates $\mathbf{w}_k^f$ and $\mathbf{w}_k^a$, and their corresponding covariances $\mathbf{P}_k^f$ and $\mathbf{P}_k^a$, respectively.

In this problem we use what is called *simulated observations*, where we take the solution of the advection equation at specific spatial locations and time intervals Δt_{obs} and add a Gaussian distributed error to it. This can be done in Matlab using the random number generator randn for normally distributed variables.

The following experiments fall in the category of what is referred in the literature as *wave generation*, where we take the initial guess (initial analysis) to be zero, i.e., $\mathbf{w}_0^a = \mathbf{0}$ and we try to reconstruct the true state processing the observations with the Kalman filter.

In what follows, you are asked to make plots for the true state and its estimate at the final time of the assimilation as well as plots of the time evolution of the domain–averaged forecast and analysis error standard deviation.

Half observation coverage case: The first case we consider is one for which the observation are all located over the left–half of the domain.

1. Following the choice of parameters in the table below, obtain the output for the true state and its estimate at the final time of the assimilation experiment.
2. What happens if the observation error level is increased to 0.1?
3. What happens if in the previous case, the assimilation period is 5 time units?
4. Comment on the results you just obtained.

Table 9.4: Observational error standard deviations.

Assimilation time period	1 time unit
Courant number	0.95
Obs frequency	$4 \times \Delta t$
Obs error std dev	0.02
Obs sparsity	left–half of grid points

You can try lots of other combinations and possibilities with this little program. There is a lot you can learn from just a small example such as this. Changing at least one of the parameters in the table above, here are a couple of other possible scenarious to investigate:

- Take observations at every grid point.
- Change the Courant number to make the dynamics more (numerically) dissipative.

What happens to the filter results in these cases?

Acknowledgments

The present notes are part of a larger set of notes initially prepared for a two–week course presented at the Instituto de Matemática Pura e Aplicada, Rio de Janeiro, in January 1996. I express my gratitude to Professor Dan Marchesin for his invitation and encouragement. I am also thankful to Beata Gundelach for her careful editing of the original Portuguese version of these notes. The Portuguese version became obsolete when these notes were used as the basis for a course thaught at the Meteorology Department of the University of Maryland in the Winter semester of 1997, after invitation by Peter Lyster, who I thank for the support and interest. I also thank Robert Rosenberg and the management of General Sciences Corporation for their understanding during that semester. Since then, shorter versions of these notes have been used at different occasions for short introductory courses in the subject of atmospheric data assimilation: Laboratório Nacional Computation Científica, Rio de Janeiro; George Mason Univerity, Virgina; and at the Workshop on Inverse Methods in Global Biogeochemical Cycles, Crete. Finally, I thank all the participants of all these classes who in many ways contributed to my constant motivation to improve the quality of these notes.