

Variational assimilation evolving individual observations and their error estimates

Ocean Applications Technical Note No. 32

M.J. Bell, A. Hines and M.J. Martin

September 2003

Produced by the Met Office
Met Office FitzRoy Road Exeter Devon EX1 3PB United Kingdom
Tel +44 (0)1392 886434 Fax: +44 (0)1392 885681
E-mail: mike.bell@metoffice.com www.metoffice.com



**Variational Assimilation Evolving Individual Observations
and their Error Estimates**

Ocean Applications Technical Note No. 32

M. J. Bell, A. Hines, M. J. Martin

September 2003

Summary

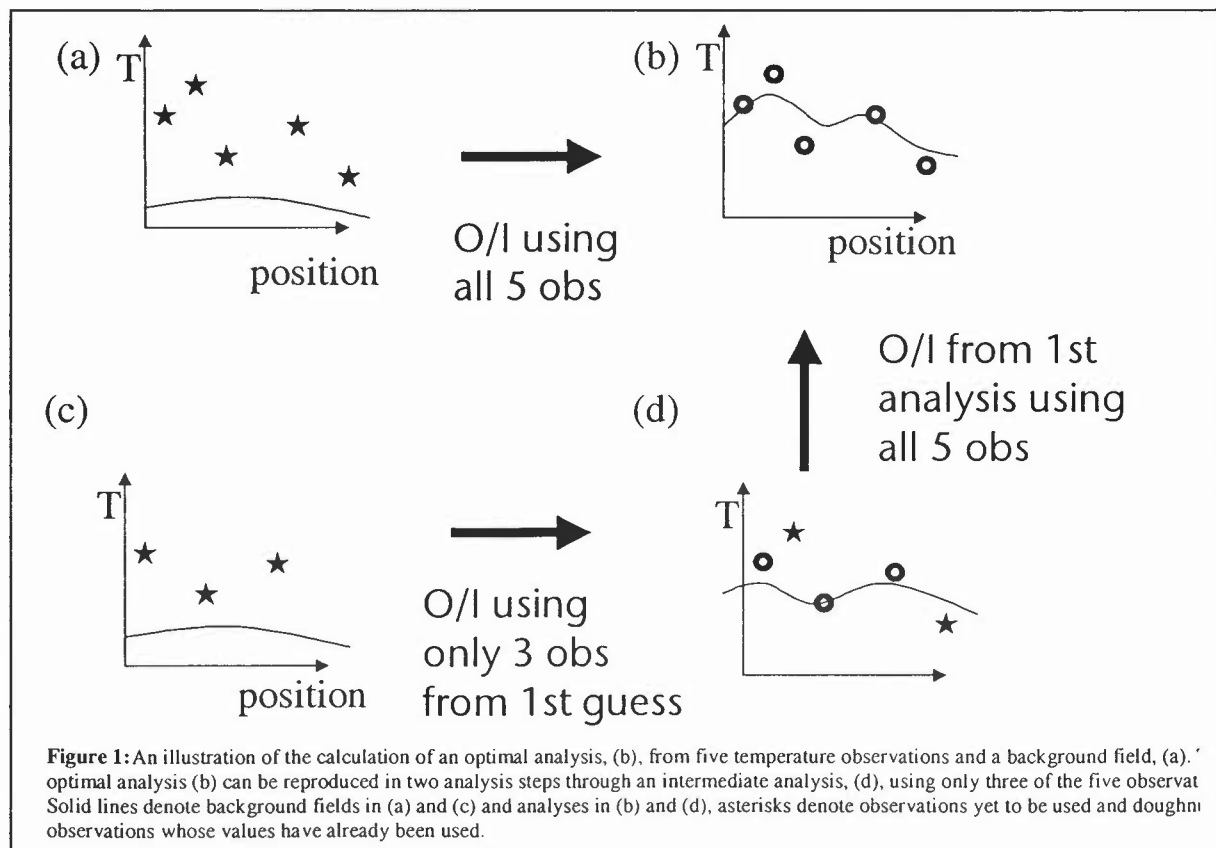
A method is presented for calculating the statistically "optimal" analysis from a set of observations and a background field through a series of "optimal" analyses generated using expanding subsets of the observations. A variant of four-dimensional variational assimilation suited to fluids with Lagrangian conservation properties is then suggested in which observations made over a time period are evolved to a common analysis time and the expected error at the analysis time is minimised. Building on these ideas, a computationally inexpensive Timely Optimal Interpolation (TOI) filter is proposed. This filter uses the full model to evolve the best estimate of the state forward in time and requires estimates of the information retained from used observations to be evolved with time. The advantages and limitations of the TOI filter are discussed, simplifications and extensions to it are suggested, and it is compared and contrasted with other schemes. A suggestion is made for improving the representation of information from the preceding time window in "standard" 4D variational assimilation schemes.

1. Introduction

Data assimilation is of central importance to numerical weather prediction (NWP) systems and is becoming increasingly used for ocean forecasting and a wide range of geophysical and ecosystem monitoring. Over the years methods of increasing sophistication have been explored (Daley 1991), starting with successive correction methods, moving on to statistically "optimal" interpolation (Lorenc 1981), then 3-dimensional (3D) and 4D variational methods (Lorenc 1986), ensemble Kalman filters (Evensen 1994), SEEK filters (Pham et al. 1998), representer and dual formulations (Bennett 1992 and Courtier 1997) and combinations of methods (Hamill & Snyder 2000, Fisher & Andersson 2001). The standard notation for the subject, established by Ide et al. (1997), is followed here.

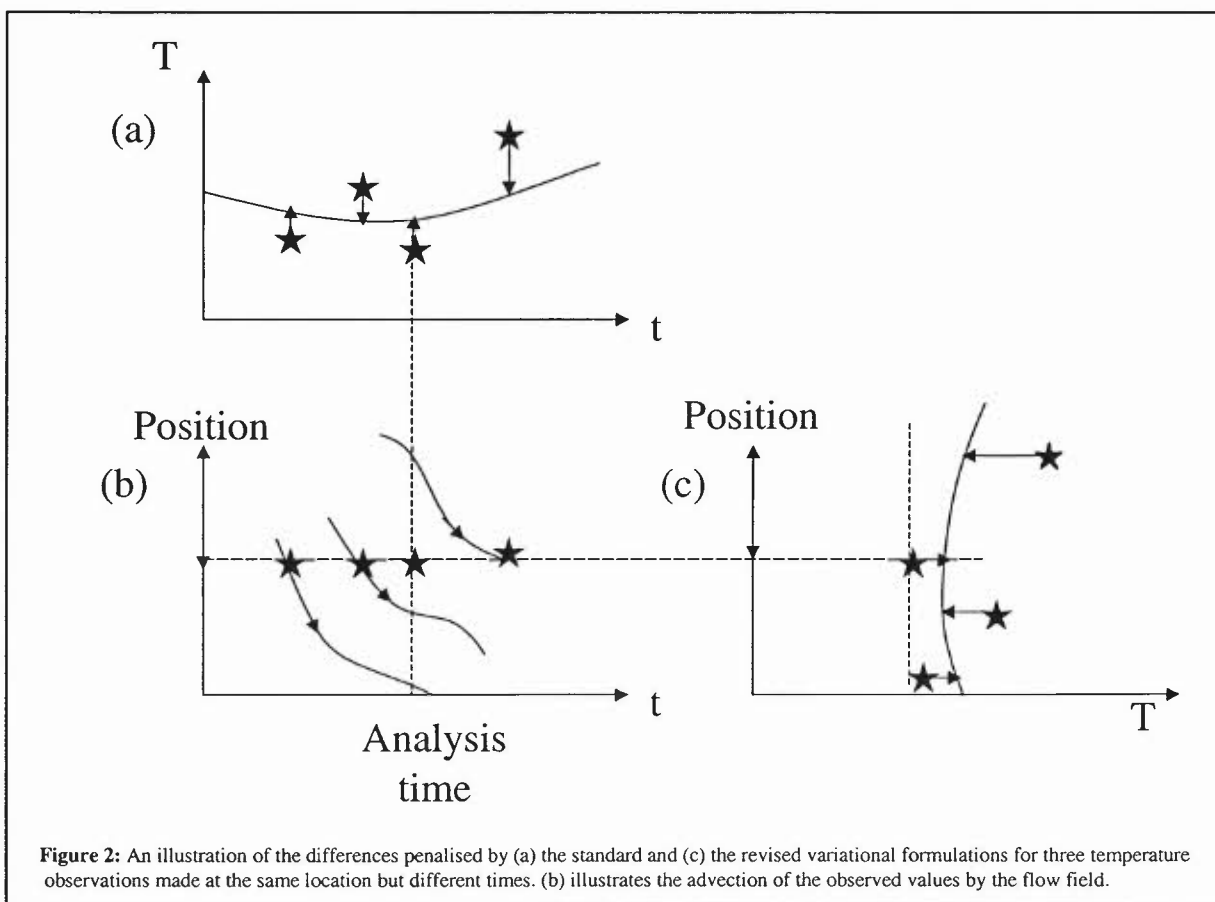
This paper is built on an algebraic identity presented in section 2, and a variant of the 4D variational assimilation penalty function presented in section 3.

The algebraic identity shows how to update an analysis field with information from new observations whilst taking into account the information supplied from previous observations. It is illustrated schematically by Figure 1 for a state vector, \mathbf{x} , consisting simply of temperature as a function of a single ordinate. Figure 1(a) shows a set of five observations, \mathbf{y}^o , and the initial background state, \mathbf{x}^b , and figure 1(b) the optimal analysis, \mathbf{x}^a , obtained by combining these inputs. Figure 1(c) illustrates a sub-set of three of the five observations, \mathbf{y}_1^o , and the same background state as in figure 1(a) and figure 1(d) the analysis, \mathbf{x}_1 , obtained using just those elements. It is shown in section 2 (for a much more general case) that the optimal analysis (figure



1(b)) may be obtained from x_1 using the full set of five observations with the values of each of the three observations already used set equal to the background value at their location.

The variant of the 4D variational assimilation (4DVar) penalty functional is illustrated by figure 2. In the standard formulation of 4DVar, depicted in figure 2(a), a model trajectory is sought which minimises the change to the initial state from its first guess and the misfit to the observations over the selected time window. The formulation proposed here supposes that the observations are evolved to the analysis time (figure 2b) and an analysis is sought which minimises suitably weighted misfits all calculated at the analysis time (figure 2c). This formulation implicitly assumes that the observed quantities can be related to tracer quantities which can be evolved in time.



The ideas outlined above are combined in the Timely Optimal Interpolation (TOI) filter, proposed in section 4, which steps forward in time using the full model to evolve the best estimate of the state and estimates the information retained from observations previously assimilated by evolving their locations and error covariances (forward in time). There are three strands of motivation for the TOI filter. Firstly that it should enable analyses to be updated frequently compared with the period over which some aspects of the information content from observations should be retained. This is particularly relevant to systems that represent phenomena with a wide range of time and space scales that are not uniformly resolved by the observational network. Secondly that the dynamics should be used to retain information in a relatively transparent manner so the assumptions and implied information flow within the

system can be understood and assessed. Thirdly that its implementation should be relatively cheap, i.e. not necessitate large numbers of model integrations.

Section 4 begins by stating an algorithm for implementation of the TOI filter and clarifying the range of model dynamics to which it is applicable and the appropriate definition of the background error covariance. Some simplifications to the filter and extensions to it are then outlined and the advantages and limitations of the filter summarised. Section 5 compares and contrasts the filter with other assimilation schemes and suggests a method for improving the specification of the background error covariance in standard 4DVar. Section 6 summarises the main points of the paper.

2. Optimal interpolation using expanding sub-sets of observations

Lorenc (1986) showed that the minimisation of the now standard variational penalty function is equivalent to a generalised form of optimal interpolation. The analysis state, \mathbf{x}^a , obtained from a background state \mathbf{x}^b , a set of observations \mathbf{y}^o and estimates for the error covariances of the background state, \mathbf{P} , and observations, \mathbf{R} , can be written in the form:

$$\mathbf{x}^a - \mathbf{x}^b = \mathbf{K}(\mathbf{y}^o - \mathbf{H}(\mathbf{x}^b)) \quad ; \quad \mathbf{K} \equiv \mathbf{P}\mathbf{H}^T (\mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R})^{-1}. \quad (1)$$

In the above $\mathbf{H}(\mathbf{x})$ "interpolates" the model state to the observation vector, \mathbf{H} is its tangent linear derivative with respect to the elements of the model state and \mathbf{K} is the Kalman gain matrix.

Suppose that the observations are divided into two sub-sets:

$$\mathbf{y}^o = \begin{pmatrix} \mathbf{y}_1^o \\ \mathbf{y}_2^o \end{pmatrix} \quad ; \quad \mathbf{H}(\mathbf{x}) = \begin{pmatrix} \mathbf{H}_1(\mathbf{x}) \\ \mathbf{H}_2(\mathbf{x}) \end{pmatrix} \quad . \quad (2)$$

A first analysis, \mathbf{x}_1 , may be calculated from \mathbf{x}^b and the first set of observations \mathbf{y}_1^o :

$$\mathbf{x}_1 - \mathbf{x}^b = \mathbf{K}_1(\mathbf{y}_1^o - \mathbf{H}_1(\mathbf{x}^b)) \quad ; \quad \mathbf{K}_1 \equiv \mathbf{P}\mathbf{H}_1^T (\mathbf{H}_1\mathbf{P}\mathbf{H}_1^T + \mathbf{R}_1)^{-1}. \quad (3)$$

A second analysis, \mathbf{x}_2 , may be calculated using \mathbf{x}_1 as the background field and both sets of observations, with the observational increments for the first set of observations set equal to zero:

$$\mathbf{x}_2 - \mathbf{x}_1 = \mathbf{K} \begin{pmatrix} 0 \\ \mathbf{y}_2^o - \mathbf{H}_2(\mathbf{x}_1) \end{pmatrix}. \quad (4)$$

We now prove that \mathbf{x}_2 is identical to \mathbf{x}^a , provided a) that there are no cross-correlations between the errors in the first set of observations and the second set, i.e.

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{21} & \mathbf{R}_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{R}_1 & 0 \\ 0 & \mathbf{R}_2 \end{pmatrix}, \quad (5)$$

and b) that the observation operator is linear.

The last term in (4) will be manipulated using

$$\begin{aligned} \mathbf{y}_2^o - H_2(\mathbf{x}_1) &= \mathbf{y}_2^o - H_2(\mathbf{x}^b + \mathbf{x}_1 - \mathbf{x}^b) \\ &= \mathbf{y}_2^o - H_2(\mathbf{x}^b) - \mathbf{H}_2 \mathbf{K}_1 (\mathbf{y}_1^o - H_1(\mathbf{x}^b)) \end{aligned} \quad (6)$$

where (3) has been used to derive the second equality and the observation operator has been assumed to be linear. Adding (3) and (4) and using (6) we obtain:

$$\mathbf{x}_2 - \mathbf{x}^b = \mathbf{K} \begin{pmatrix} 0 \\ \mathbf{y}_2^o - H_2(\mathbf{x}^b) \end{pmatrix} + \left[\mathbf{K}_1 - \mathbf{K} \begin{pmatrix} 0 \\ \mathbf{H}_2 \end{pmatrix} \mathbf{K}_1 \right] [\mathbf{y}_1^o - H_1(\mathbf{x}^b)]. \quad (7)$$

In order to combine terms we re-organise the first element in the second term on the r.h.s. of (7) as follows:

$$\begin{aligned} \mathbf{K}_1 (\mathbf{y}_1^o - H_1(\mathbf{x}^b)) &= \mathbf{P} \mathbf{H}_1^T (\mathbf{H}_1 \mathbf{P} \mathbf{H}_1^T + \mathbf{R}_1)^{-1} (\mathbf{y}_1^o - H_1(\mathbf{x}^b)) \\ &= \mathbf{P} \begin{pmatrix} \mathbf{H}_1^T & \mathbf{H}_2^T \end{pmatrix} \begin{pmatrix} (\mathbf{H}_1 \mathbf{P} \mathbf{H}_1^T + \mathbf{R}_1)^{-1} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{y}_1^o - H_1(\mathbf{x}^b) \\ 0 \end{pmatrix} \\ &= \mathbf{P} \mathbf{H}^T (\mathbf{H} \mathbf{P} \mathbf{H}^T + \mathbf{R})^{-1} (\mathbf{H} \mathbf{P} \mathbf{H}^T + \mathbf{R}) \begin{pmatrix} (\mathbf{H}_1 \mathbf{P} \mathbf{H}_1^T + \mathbf{R}_1)^{-1} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{y}_1^o - H_1(\mathbf{x}^b) \\ 0 \end{pmatrix} \\ &= \mathbf{P} \mathbf{H}^T (\mathbf{H} \mathbf{P} \mathbf{H}^T + \mathbf{R})^{-1} \begin{pmatrix} I \\ \mathbf{H}_2 \mathbf{P} \mathbf{H}_1^T (\mathbf{H}_1 \mathbf{P} \mathbf{H}_1^T + \mathbf{R}_1)^{-1} \end{pmatrix} (\mathbf{y}_1^o - H_1(\mathbf{x}^b)) \\ &= \mathbf{K} \begin{pmatrix} I \\ \mathbf{H}_2 \mathbf{K}_1 \end{pmatrix} (\mathbf{y}_1^o - H_1(\mathbf{x}^b)). \end{aligned} \quad (8)$$

Note that the penultimate step in (8) uses (5). Using (8), (7) simplifies to

$$\mathbf{x}_2 - \mathbf{x}^b = \mathbf{K} \begin{pmatrix} 0 \\ \mathbf{y}_2^o - H_2(\mathbf{x}^b) \end{pmatrix} + \mathbf{K} \begin{pmatrix} \mathbf{y}_1^o - H_1(\mathbf{x}^b) \\ 0 \end{pmatrix}. \quad (9)$$

Comparing (9) with (1) one sees that \mathbf{x}_2 is identical to \mathbf{x}^a .

By induction this result can be applied to an arbitrary number of expanding sub-sets for any set of observations.

This result can be interpreted using Bratseth's (1986) formulation of the analysis correction scheme in which the observations are incremented as well as the model field during analysis steps and both converge towards the "optimal" interpolation solution. After an analysis step, observations which have previously been used will have the same values as the model analysis - which is the background value for the next analysis step.

The simplest application of this result is to an observation vector formed from observations that are all valid at the same time. In this case the division of observations into subsets is arbitrary

and the difference between “new” and “previous” observations simply relates to this division. Technically the result also applies to observations taken over a period of time when the interpolation operator, H , includes the evolution of the model state to the observation time (though usually the interpolation operator is not then a linear operator). Again one is free to choose the division of observations into subsets, though it would seem natural, and perhaps computationally advantageous, to choose the order of the subsets to correspond to the time order in which the observations have been made.

3. A Variant of Four Dimensional Variational Assimilation

We will write down a penalty functional for calculation of an analysis, \mathbf{x}^a , at a single time t_a using a background field, \mathbf{x}^b , at the same time and observations of Lagrangian tracers, \mathbf{y}_i^o valid at times t_i , from a time period $T_s < t_i < T_f$, which penalises the misfit calculated at the analysis time. $M_{q,p}[\mathbf{x}]$ will (for any p and q) represent the model state obtained at time t_q by the model's evolution from an initial state \mathbf{x} at time t_p .

For the purposes of comparison we first write down the standard formulation of the penalty functional, which is:

$$\mathcal{J}_\alpha(\mathbf{x}) = (\mathbf{x} - \mathbf{x}^b)^T \mathbf{P}^{-1} (\mathbf{x} - \mathbf{x}^b) + \sum_i \left[\mathbf{y}_i^o - H_i(M_{i,\alpha}[\mathbf{x}]) \right]^T \mathbf{R}_i^{-1} \left[\mathbf{y}_i^o - H_i(M_{i,\alpha}[\mathbf{x}]) \right]. \quad (10)$$

This formulation penalises departures between the model trajectory and (a) a background field (usually provided by a previous forecast) at the analysis time and (b) observations at the locations and times for which they are valid.

The alternative penalty functional $\mathcal{J}_\beta(\mathbf{x})$ is suggested for observations which can be related to Lagrangian tracers. For such observations $G_{\beta,i}[\mathbf{y}_i^o]$ will represent the observations made at time t_i whose values and locations have been evolved to the analysis time t_β and $H_{\beta,i}[\mathbf{x}]$ will represent the corresponding observation vector formed by simple “interpolation” of the state \mathbf{x} to these evolved locations. Using this notation the alternative penalty functional is given by

$$\mathcal{J}_\beta(\mathbf{x}) = (\mathbf{x} - \mathbf{x}^b)^T \tilde{\mathbf{P}}^{-1} (\mathbf{x} - \mathbf{x}^b) + \sum_i \left[G_{\beta,i}[\mathbf{y}_i^o] - H_{\beta,i}[\mathbf{x}] \right]^T \tilde{\mathbf{R}}_i^{-1} \left[G_{\beta,i}[\mathbf{y}_i^o] - H_{\beta,i}[\mathbf{x}] \right]. \quad (11)$$

where $\tilde{\mathbf{P}}$ is the background error covariance (at time t_β) and

$$\tilde{\mathbf{R}}_i = \overline{\left[G_{\beta,i}[\mathbf{y}_i^o] - H_{\beta,i}[\mathbf{x}] \right] \left[G_{\beta,i}[\mathbf{y}_i^o] - H_{\beta,i}[\mathbf{x}] \right]^T} \quad (12)$$

is the expected error covariance in the evolved observations at the analysis time, the overbar denoting the expectation operator.

The standard formulation (10) is usually applied with the analysis defined to be at the start of the observation period. At least for passive tracers and accurate velocity fields, it is more natural

with the new formulation to place the analysis in the middle of the observation period (as the best estimate "smoother") or at the end of the observation period as the best analysis for the ensuing forecast. For the latter case the background field in (11) will be taken from a previous analysis valid at the start of the observation time window (T_s) and the analysis will be valid at the end of the observation time window (T_f). So $\tilde{\mathbf{P}}$ should be defined as the error covariance of forecasts of length $T_f - T_s$.

The implementation and interpretation of (11) and (12) depends on whether the tracers are active or passive and on the number of sources which can contribute to the expected growth of error in each of the observations evolved to the analysis time.

For passive tracers where the velocity field can be assumed to be exactly known the 4D problem reduces to a 3D variational analysis problem for each analysis time. The formulation can be trivially extended for a passive tracer τ driven by a source function S with known error characteristics ($D\tau/Dt = S$). The appendix shows that the analysis generated using (11) is identical to that obtained from (10) when

$$\mathbf{M}_{\beta,\alpha} \mathbf{P} \mathbf{M}_{\beta,\alpha}^T = \tilde{\mathbf{P}} \quad ; \quad \mathbf{G}_{\beta,i} \mathbf{R}_i \mathbf{G}_{\beta,i} = \tilde{\mathbf{R}}_i \quad . \quad (13)$$

For passive tracers where the velocity field cannot be assumed to be exactly known the two formulations will in general give different results. For strongly non-linear problems the penalty functional (10) can acquire multiple minima and eventually fractal structure as the time interval covered by the observations is increased (Miller et al. 1994). If $\tilde{\mathbf{P}}$ has a relatively simple structure and the errors in the evolved observations at the analysis time are estimated adequately the alternative penalty functional, (11), should be less badly behaved. Whilst it is probably difficult to estimate the growth in the error in the location or value of an observation, it is an advantage to have this error included in the formulation.

If the velocities advecting the tracers have uncertainties which can be estimated the most appropriate form for the observation penalty is not clear: errors in the observations could be represented as errors in tracer values or as errors in the observation's location. Also errors in previously used neighbouring "evolved" observations will become correlated. Parametrisation of the growth of these correlations would be important for practical applications (see section 4(c)).

For active tracers the position is more complicated because the best estimates of the velocity field for the analysis and background states will differ. Thus it would appear to be beneficial to iterate the analyses, updating the velocity fields used on each iteration. Note that iterations for improvements to the velocity field are still separated from the solution of the 3D variational problem. In the standard formulations of 4DVar the penalty function is minimised by iterative methods such as generalised conjugate gradients. The 3DVar problem by itself requires large numbers of iterations for convergence (unless it is pre-conditioned very well). Thus, even for active tracers, there may still be some computational advantages for the new approach.

The new variant of 4Dvar could be extended in two ways. Firstly as in 3Dvar the observations do not have to be directly related to single model variables. The formulation can be applied to observations which are related to the model state using an observation operator providing the model variables used by the operator can be evolved as Lagrangian tracers. Secondly the dynamics which determine the "evolution" of the observations need not be as simple as pure

adiabatic advection. For example the evolution of vertical profile data could include a representation of vertical mixing.

4. The Timely Optimal Interpolation (TOI) Filter

It is generally thought to be very important to calculate the observational increments $\mathbf{y}^o - H(\mathbf{x}^b)$ as accurately as possible. In view of this point, it is likely to be difficult to evolve observations forwards or backwards to the analysis time, as required by (11), sufficiently accurately. An alternative, that is particularly suited to the application of (11) as a filter rather than a smoother and makes use of the idea for sequentially updating analyses with new observations discussed in section 2, is to use the model to evolve forward the best estimate of the state whilst evolving the expected error covariances of previously used observations.

a. Basic Algorithm for Filter

The filter can be implemented by the following sequence of steps which starts from the background field, \mathbf{x}^b , for the next analysis:

- i. identify the observations that have previously been used, to set their observational increments to zero, and to update their estimated locations
- ii. revise the estimated observational error covariances $\tilde{\mathbf{R}}_i$
- iii. calculate a new analysis, \mathbf{x}^a using both new and previously used observations
- iv. integrate the model forward to the next analysis time incorporating the analysis increments during that period using an appropriate method (e.g. incremental analysis updates (IAU) or a digital filter (Lynch et al. 1997)).

b. Clarification of important points

The period between analyses could be chosen to be short compared with the period for which observations retain their relevance; for NWP the period of the cycle might be 1-3 hours whilst in oceanography it might be 6 hours to 1 day.

As mentioned at the end of section 3 the dynamics which determine the "evolution" of the information assimilated from the observations need not be as simple as pure adiabatic advection. When other processes (such as vertical mixing) are important in the evolution of the model variables related to the observation by the observation operator these should be taken into account in the evolution of the observation error covariance $\tilde{\mathbf{R}}_i$.

The background error covariances are not evolved in this scheme. If the observation error covariances are handled accurately the background error covariance should be defined in terms of forecasts similar in length to the period over which observations are retained.

c. Simplifications

Accurate estimation of the growth of the covariance of the errors in the observations as they are evolved forward in time is likely to be both difficult and important. A simple approach to parametrisation of the error growth is to suppose that a covariance with the same spatial

structure as the background error covariance will develop with a characteristic time scale, τ_c , and functional form $T[t/\tau_c]$ with $T[0]=0$ and $T[(t_a-t_b)/\tau_c]=1$. Then the covariance of the errors at the analysis time t_a of "evolved" observations i and j made at times t_i and t_j with $t_i < t_j$ can be written in the form

$$\tilde{\mathbf{R}}_{ij}(t_a) = \mathbf{R}_{ij} + \mathbf{H}_i^T \mathbf{P} \mathbf{H}_j T[(t_a - t_j)/\tau_c] (1 + \gamma T[(t_j - t_i)/\tau_c]) \quad (14)$$

where \mathbf{R}_{ij} is the covariance of the two original observations and $0 < \gamma < 1$. The higher (lower) values of γ obtain when the growth in error during the period $t_j < t < t_a$ is strongly (weakly) correlated with the growth during the period $t_i < t < t_j$. If the background error covariance contains contributions from processes with significantly different timescales it may be advantageous for (14) to be written as the sum of these contributions each developing with their own timescales.

A second simplification which is convenient and may be appropriate in some circumstances is to neglect the advection of the observation locations. This is likely to produce acceptable results if the distance the observations are advected over the period for which they are retained is small compared with the typical separation of observations and the scales of the background errors.

We have implemented a simplified version of the TOI filter in a system which assimilates oceanographic data to produce a new analysis once a day. So far the advection of observation locations has been neglected and the observation error covariances $\tilde{\mathbf{R}}_{ij}$ are taken to be diagonal rather than parametrised as suggested by (14). One of the problems which our implementation is designed to address is the relatively long period which one must wait for an adequate distribution of vertical profile observations to accumulate. Even when the Argo program (Argo, 1998) is fully implemented it will be necessary to combine data from at least a 5-10 day period to obtain adequate analyses below the surface mixed layer. Similarly all data in a single altimeter track are valid at approximately the same time but analyses should take into account at least the nearest neighbour track which is usually valid at a time offset by 3 days or more. The new scheme enables observations to be fully assimilated into the model as soon as they become available, and does not require long period assimilations to be carried out each day.

d. Extensions

It may not be possible to introduce analysis increments without generating excessive gravity waves when very short update cycles are employed. We describe two methods which could then be used to enable observational increments to be calculated closer to their time of validity in the TOI filter scheme.

Firstly, forecasts from the background field could be interpolated in time to the observations and observational increments calculated accordingly. In our oceanographic system (see section (c) above) the use of this first guess at the appropriate time (FGAT) might be of value for surface temperature data

Secondly, for vertical profiles of oceanographic data in order to avoid aliasing details of the temporal evolution of the near surface layer into spatial variations, it may be appropriate to

evolve the observation profiles, valid on a single day, forward to a common time, using a one-dimensional mixed layer model (MLM). Ideally the analysis increments would be inserted over a short time period compared with that characteristic of the evolution of the near surface layer. If that is not possible the analysis increments might be calculated from a time series of tracer analyses evolved using the MLM.

It may be possible to capture some of the temporal evolution of velocity information using Lagrangian methods similar to those used above for tracers. In semi-geostrophic theory (Hoskins 1975) the momentum equations can be written in the form

$$DX/Dt = -f^{-1}\partial\phi/\partial y \quad ; \quad DY/Dt = f^{-1}\partial\phi/\partial x \quad (15)$$

where f is the Coriolis parameter, ϕ is the geopotential height and X and Y are the geostrophic coordinates defined in terms of the geostrophic velocities, u_g and v_g by

$$X = x + f^{-1}v_g \quad ; \quad Y = y - f^{-1}u_g \quad . \quad (16)$$

This suggests that velocity observations might be interpreted as observations of (X, Y) and advected as active tracers with suitable source terms. Thus it may be possible to apply the filter to the assimilation of velocity data as well as tracer data.

e. Advantages

It may be helpful to summarise the advantages of the TOI filter:

- (i) the observation increments $y^o - H(x^b)$ are calculated and the analysis increments are made to the background x^b using fields at, or very close to, the time of validity of each observation
- (ii) the analysis update cycle is independent of the period for which information from individual observations can be retained
- (iii) there are no artificial windows related to the grouping/selection of observations.
- (iv) recent previous observations add local detail to the implied error covariance of the background field
- (v) the information which is explicitly retained by the system is relatively transparent (i.e. the information retained by tracers is advected in a relatively simple way)
- (vi) no adjoint model is needed by the filter: the CPU expense, coding expense, and difficulties formulating the forward and adjoint models for some physical processes are avoided
- (vii) it is not essential to evolve an ensemble of model states.

Taken together points (i)-(iii) suggest that the scheme may be particularly useful when different spatial scales or regions evolve on different timescales. For example in the ocean, the near-surface mixed layer varies considerably over the course of a day, whilst the deep ocean and mesoscale evolve relatively slowly. It is desirable for some applications to produce forecasts on at least a daily basis as the surface forcing is updated and new observations are received each day.

For comparison we note that 4Dvar calculates increments at obs time but applies them at another time. Its analysis cycle is tied to the observation time window and it has artificial windows (points ii-iv are closely related). When applied with the model as a strong constraint the information retained is transparent but of doubtful validity when long analysis cycles are used. Of course 4DVar needs the model's adjoint but does not require evolution of an ensemble of states

In principle advantages i-iv are shared by the ensemble Kalman filter (EnKF) but in practice analyses are expensive which discourages short analysis cycles and the small number of members in a typical ensemble would limit the amount of information retained in the background error covariance. The EnKF is more ambitious in the flow of information retained (point v) and doesn't need an adjoint but does of course need an ensemble of states to be evolved.

f. Limitations

The TOI filter scheme is not proposed as a panacea and has clear limitations. Important limitations include the following points.

- (a) In common with Kalman filters the scheme is sequential in time. By contrast the 4Dvar approach fits a trajectory through all observations in a given time window. It may be easier to use time tendency information with the trajectory fitting approach. In particular, 4Dvar schemes can counteract systematic model errors by adjustment of control variables (such as the surface wind stress field)
- (b) The scheme only applies to observations for which the information retained by the model can be represented using fairly simple ideas such as Lagrangian advection or one-dimensional mixing
- (c) As in 4Dvar the scheme needs the background error covariances to be well specified.

5. Discussion

a. Comparison with previous work

In some respects the TOI filter scheme is similar to the very elegant representer formulation developed over many years by Bennett and co-workers (Chua & Bennett 2001). In that formulation the model dynamics are used to determine the influence of each observation over the time period both preceding and following it. In the TOI filter formulation we have retained only simplified dynamics (e.g. Lagrangian advection) and observations are only allowed to affect analyses at or after their time of validity. In the representer formulation a background error covariance is not used. Instead the growth rate of model errors is parametrised (these errors having both spatial and temporal correlations).

The TOI filter scheme also has something in common with Cohn (1993) in that both attempt to use Lagrangian dynamics to improve the utilisation of data assimilation schemes. However Cohn uses these ideas to evolve the background error covariance rather than the location of individual observations. Fisher & Lary (1995) also applied Lagrangian ideas to the assimilation of ozone data in a stratospheric model.

b. Application to 4Dvar schemes

The ideas presented in this paper might also be used to augment existing 4Dvar assimilation schemes used in NWP systems. A weakness of those schemes is that the error covariance used for the initial state usually takes no explicit account of the observations received in the period immediately preceding the initial state. It would appear to be a relatively simple matter to advect the locations of at least some of those observations forward to the time of the initial state and to include them in the observation term of the penalty functional with the observation values set equal to those derived from the background field. This should improve the implied spatial structure of the background error covariance matrix.

c. Propagation of signals or knowledge by wave motions

Signals are propagated by Rossby, Kelvin and gravity waves in the atmosphere and ocean and disturbances to the depth of the ocean's thermocline propagate along the equator. Thus it is perhaps a weakness of our filter scheme that it does not capture explicitly the propagation of such information. However, whilst wave motions will certainly propagate displacements to the thermocline along the equator, without accurate knowledge of the local initial state of the thermocline it is hard to see how accurate knowledge of the final local state of the thermocline can propagate in from elsewhere. This issue may be worth further study.

6. Concluding Summary

The main ideas that have been presented in this paper are:

1) Analyses can be built up in stages adding increments from sub-sets of observations. When the observational errors in the different sub-sets are uncorrelated and the observation operator is linear the resulting analyses are the same as those obtained by a standard 3DVar analysis. The estimated errors in the analyses at each stage can be calculated from the error covariance of the initial background state and the observation error covariances.

2) Mindful of the Lagrangian conservation of tracers, of semi-geostrophic theory, and of the growth of errors due to non-linear advection, an alternative formulation of the penalty functional for variational assimilation has been defined in which the information content from individual observations is advected to a single analysis time and the misfit to the evolved observations and the background field is minimised.

3) A computationally inexpensive adaptation of these ideas has been outlined in which the full model is used to evolve the analysis for one cycle forward to provide the background field for the next cycle, the locations of previously used observations are evolved, and the estimate of their error covariance is evolved and used to supplement the background error covariance. This TOI filter scheme enables observations to increment the model very close to their time of validity and is not restricted to observations of Lagrangian tracers. It only requires estimates to be made of the retention of the information assimilated.

4) Local detail based on the evolved locations of previously used data could be introduced into the model error covariances employed in the "standard" 4DVar penalty function.

Acknowledgements

Stuart Bell suggested that the full model be used to evolve the observation values when some of these ideas were first discussed in 1991. Andy White provided a very helpful review.

Appendix: Equivalence of formulations for passive tracers advected by exact velocities

The analysis \mathbf{x}_α^a which minimises the original penalty functional (10) satisfies

$$\mathbf{P}^{-1}(\mathbf{x}_\alpha^a - \mathbf{x}_\alpha^b) = \sum_i \mathbf{M}_{i,\alpha}^T \mathbf{H}_i^T \mathbf{R}_i^{-1} [\mathbf{y}_i^o - H_i(M_{i,\alpha}[\mathbf{x}_\alpha^a])] \quad (\text{A.1})$$

The analysis \mathbf{x}_β^a which minimises the new penalty functional (11) satisfies

$$\tilde{\mathbf{P}}^{-1}(\mathbf{x}_\beta^a - \mathbf{x}_\beta^b) = \sum_i \mathbf{H}_{\beta,i}^T \tilde{\mathbf{R}}_i^{-1} [G_{\beta,i}[\mathbf{y}_i^o] - H_{\beta,i}[\mathbf{x}_\beta^a]] \quad (\text{A.2})$$

We will start from the solution \mathbf{x}_β^a for the new formulation given by (A.2) and manipulate it into a form which can be compared with the solution \mathbf{x}_α^a for the old formulation given by (A.1).

First note that

$$M_{\beta,\alpha}(\mathbf{x}) = M_{\beta,i}[M_{i,\alpha}(\mathbf{x})] \quad ; \quad \mathbf{x}_\beta^b = M_{\beta,\alpha}(\mathbf{x}_\alpha^b) \quad (\text{A.3})$$

and let \mathbf{x}_u be defined by

$$M_{\beta,\alpha}(\mathbf{x}_u) = \mathbf{x}_\alpha^a \quad (\text{A.4})$$

which is acceptable as $M_{\beta,\alpha}$ has an inverse. Recall that \mathbf{x}_u is valid at the analysis time for the first formulation, $M_{i,\alpha}$ evolves a field from the analysis time t_α in the original formulation to the time of the i th observation and $M_{\beta,i}$ evolves a field from the i th observation time to the time t_β of the analysis for the new formulation.

Substituting (A.3) - (A.4) into (A.2) we obtain:

$$\tilde{\mathbf{P}}^{-1} \mathbf{M}_{\beta,\alpha}(\mathbf{x}_u - \mathbf{x}_\alpha^b) = \sum_i \mathbf{H}_{\beta,i}^T \tilde{\mathbf{R}}_i^{-1} [G_{\beta,i}[\mathbf{y}_i^o] - H_{\beta,i}[M_{\beta,i}M_{i,\alpha}\mathbf{x}_u]]. \quad (\text{A.5})$$

Now

$$H_{\beta,i}[M_{\beta,i}(\mathbf{x})] = G_{\beta,i}[H_i(\mathbf{x})] \quad . \quad (\text{A.6})$$

Using this in (A.5) and pre-multiplying through by $\mathbf{M}_{\beta,\alpha}^T$ gives

$$\mathbf{M}_{\beta,\alpha}^T \tilde{\mathbf{P}}^{-1} \mathbf{M}_{\beta,\alpha} (\mathbf{x}_u - \mathbf{x}_\alpha^b) = \sum_i \mathbf{M}_{i,\alpha}^T \mathbf{M}_{\beta,i}^T \mathbf{H}_{\beta,i}^T \tilde{\mathbf{R}}_i^{-1} \mathbf{G}_{\beta,i} [\mathbf{y}_i^o - H_i(M_{i,\alpha} \mathbf{x}_u)] \quad . \quad (\text{A.7})$$

Linearising (A.6) we have

$$\mathbf{M}_{\beta,i}^T \mathbf{H}_{\beta,i}^T = (\mathbf{H}_{\beta,i} \mathbf{M}_{\beta,i})^T = (\mathbf{G}_{\beta,i} \mathbf{H}_i)^T = \mathbf{H}_i^T \mathbf{G}_{\beta,i}^T \quad (\text{A.8})$$

Substituting (A.8) into (A.7) we obtain

$$\mathbf{M}_{\beta,\alpha}^T \tilde{\mathbf{P}}^{-1} \mathbf{M}_{\beta,\alpha} (\mathbf{x}_u - \mathbf{x}_\alpha^b) = \sum_i \mathbf{M}_{i,\alpha}^T \mathbf{H}_i^T \mathbf{G}_{\beta,i}^T \tilde{\mathbf{R}}_i^{-1} \mathbf{G}_{\beta,i} [\mathbf{y}_i^o - H_i(M_{i,\alpha} \mathbf{x}_u)] \quad . \quad (\text{A.9})$$

Then if

$$\mathbf{M}_{\alpha,\beta}^T \tilde{\mathbf{P}}^{-1} \mathbf{M}_{\alpha,\beta} = \mathbf{P}^{-1} \quad ; \quad \mathbf{G}_{\beta,i}^T \tilde{\mathbf{R}}_i^{-1} \mathbf{G}_{\beta,i} = \mathbf{R}_i^{-1} \quad (\text{A.10})$$

(A.9) gives

$$\mathbf{P}^{-1} (\mathbf{x}_u - \mathbf{x}_\alpha^b) = \sum_i \mathbf{M}_{i,\alpha}^T \mathbf{H}_i^T \mathbf{R}_i^{-1} [\mathbf{y}_i^o - H_i(M_{i,\alpha} [\mathbf{x}_u])] \quad (\text{A.11})$$

Comparing (A.11) and (A.1) one sees that $\mathbf{x}_u = \mathbf{x}_\alpha^a$. (A.10) can be used to derive (13) because $M_{\beta,\alpha}$ has an inverse.

References

Argo Science Team 1998: On the design and implementation of Argo - An initial plan for a global array of profiling floats. ICPO Report No. 21. GODAE Report No. 5. GODAE International Project Office, Melbourne, Australia, 32 pp.

Bennett, A. F., 1992: *Inverse Methods in Physical Oceanography*. Cambridge University Press, Cambridge, U. K., 346 pp.

Bratseth, A. M., 1986 Statistical interpolation by means of successive corrections. *Tellus*, 38A, 439-447.

Chua, B. S., and A. F. Bennett, 2001: An inverse ocean modelling system. *Ocean Modelling*, 3, 137-165.



Cohn, S. E., 1993: Dynamics of short-term univariate forecast error covariances. *Mon. Weath. Rev.*, 121, 3123-3149.

Courtier, P., 1997: Dual formulation of four-dimensional variational assimilation. *Quart. J. Roy. Meteor. Soc.* 123, 2449-2461.

Daley, R., 1991: *Atmospheric Data Analysis*, Cambridge University Press, Cambridge, U. K., 457 pp.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, 99 (C5), 10143-10162.

Fisher, M., and D. J. Lary, 1995: Lagrangian four-dimensional variational data assimilation of chemical species. *Quart. J. Roy. Meteor. Soc.*, 121, 1681-1704.

Fisher, M., and E. Andersson, 2001: Developments in 4D-Var and Kalman filtering. ECMWF Tech. Memo. 347.

Hamill, T. M., and C. Snyder, 2000: A hybrid ensemble Kalman filter-3D variational analysis scheme. *Mon. Weath. Rev.*, 128, 2905-2919.

Hoskins, B. J., 1975: The geostrophic momentum approximation and the semi-geostrophic equations. *J. Atmos. Sci.*, 32, 233-242.

Ide, K., P. Courtier, M. Ghil, and A. C. Lorenc, 1997: Unified notation for data assimilation: operational, sequential and variational. *J. Met. Soc. Japan*, 75, 181-189.

Lorenc, A. C., 1981: A global three-dimensional multivariate statistical interpolation scheme. *Quart. J. Roy. Meteor. Soc.*, 109, 701-721.

Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, 112, 1177-1194.

Lynch, P., D. Giard, and V. Ivanovici, 1997: Improving the efficiency of a digital filtering scheme for diabatic initialisation. *Mon. Weath. Rev.*, 125, 1976-1982.

Miller, R., M. Ghil, and F. Gauthiez, 1994: Advanced data assimilation in strongly nonlinear systems. *J. Atmos. Sci.*, 51, 1037-1056.

Pham, D.T., J. Verron, and M. C. Roubaud, 1998: A singular evolutive extended Kalman filter for data assimilation in oceanography. *J. Mar. Sys.*, 16, 323-340.

