Rapid update cycling with overlapping windows

Tim Payne, Met Office

Workshop on Meteorological Sensitivity Analysis and Data Assimilation
Roanoke, Tuesday 2nd June 2015
Main benefits of frequent analysis updates

(i) Hourly update for global analysis enables timely BCs for possible/likely future hourly running of UKV and other LAMs

(ii) Error in linear approximation to smooth function declines quadratically with increment size

\[ \phi(x + \delta x) = \phi(x) + \phi'(x) \delta x + \frac{1}{2} \phi''(x)(\delta x, \delta x) + ... \]

hourly update → smaller increment → (hopefully) improved PF Model performance

(iii) Potentially, improved affordability

More on these later.
Scheduling of global main and update runs

Observation availability time

Observation validity time

© Crown copyright. Met Office
HPC resources used for global 4D-Var

© Crown copyright. Met Office
Need optimal (Bayesian) estimation theory for overlapping windows.

Suppose that we at time \( k \) the following vectors of observations have just become available:

\[
\begin{align*}
\mathbf{y}_k^{(k)} & \text{ valid at time } k \\
\mathbf{y}_{k-1}^{(k)} & \text{ valid at time } k - 1 \\
& \cdots \\
\mathbf{y}_{k-L}^{(k)} & \text{ valid at time } k - L
\end{align*}
\]

Superscripts denote when the obs are available and subscripts their validity time, the longest time interval to availability being \( L \) time steps.
<table>
<thead>
<tr>
<th></th>
<th>Observation validity time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$k - 2$</td>
</tr>
<tr>
<td>$(k - 2)$ Obs</td>
<td>$y^{(k-2)}_{k-2}$</td>
</tr>
<tr>
<td>$(k - 1)$</td>
<td>$y^{(k-1)}_{k-2}$</td>
</tr>
<tr>
<td>availability</td>
<td>$y^{(k)}_{k-2}$</td>
</tr>
<tr>
<td>time</td>
<td>$y^{(k+1)}_{k-1}$</td>
</tr>
<tr>
<td>$(k + 1)$</td>
<td>$y^{(k+2)}_{k}$</td>
</tr>
<tr>
<td>$(k + 2)$</td>
<td></td>
</tr>
</tbody>
</table>

**Table:** Notation for observation validity and availability times
Define $y_k$ to be the obs just available and therefore used at time $k$, so

$$y_k = \begin{pmatrix} y^{(k)}_k \\ y^{(k)}_{k-1} \\ \vdots \\ y^{(k)}_{k-L} \end{pmatrix}$$

(1)

Supposing

$$x_{i+1} = f(x_i) + \omega_i \text{ where } \omega_i \sim N(0, Q_i)$$

set

$$x_k = \begin{pmatrix} x_k \\ x_{k-1} \\ \vdots \\ x_{k-L} \end{pmatrix}, \quad f(x_k) = \begin{pmatrix} f(x_k) \\ x_k \\ \vdots \\ x_{k-L+1} \end{pmatrix}, \quad \omega_k = \begin{pmatrix} \omega_k \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

(2)

so

$$x_{k+1} = f(x_k) + \omega_k$$
Supposing also that

\[ y_i^{(k)} = H_i^{(k)} x_i + \nu_i^{(k)} \text{ where } \nu_i^{(k)} \sim N(0, R_i^{(k)}) \]

and setting

\[ \nu_k = \begin{pmatrix} \nu_k^{(k)} \\ \nu_{k-1}^{(k)} \\ \vdots \\ \nu_{k-L}^{(k)} \end{pmatrix} \]

\[ H_k = \begin{pmatrix} H_k^{(k)} \\ H_{k-1}^{(k)} \\ \vdots \\ H_{k-L}^{(k)} \end{pmatrix}, \quad Q_k = \begin{pmatrix} Q_k \\ \vdots \\ Q_k \end{pmatrix} \]

etc
Thus our system can be written

\[ x_{k+1} = f(x_k) + \omega_k \]
\[ y_k = H_k x_k + \nu_k \]

for

\[ \nu_k \sim N(0, R_k) \]
\[ \omega_k \sim N(0, Q_k) \]

and our problem is to find

\[ E[x_k | y_0, y_1, \ldots, y_k] \]

This is now in a form to which we can apply standard filtering theory.
Eg, consider case where $f$ is linearisable. We have

$$f'(x_k) = \begin{pmatrix} f'(x_k) & 0 & 0 & 0 & 0 \\ I & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

and overlapping window problem is solved by applying Extended Kalman Filter to extended states/operators:
Given first guesses $\hat{x}_{L|L-1}, P_{L|L-1}$

For $k = L, L + 1, L + 2, ..$

Assimilate:

$$K = P_{k|k-1}H_k^T(R_k + H_kP_{k|k-1}H_k^T)^{-1}$$

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K(y_k - H_k\hat{x}_{k|k-1})$$

$$P_{k|k} = (I - KH_k)P_{k|k-1}$$

Predict:

$$\hat{x}_{k+1|k} = f(\hat{x}_{k|k})$$

$$P_{k+1|k} = f'(\hat{x}_{k|k})P_{k|k}f'(\hat{x}_{k|k})^T + Q_k$$

End for $k = L, L + 1, L + 2, ..$
Remarks

- Above is EKF solution, but can equally well use unscented KF, ensemble SRF, etc
- It is a sequential solution which is equivalent to assimilating all obs up to that time simultaneously
- Full solution only involves model (or PF model) integration over last time slot, however
- the solution involves carrying $nL \times 1$ state vectors and $nL \times nL$ covariance matrices (more on this later)
- If all new obs occur only in last time slot simplifies to EKS
- Solution is equivalent to a modified version of 4D-Var, as shown on next slide:
Suppose \( L = 2 \) and the analysis error covariance at the end of previous window \( \{i - 2, i - 1, i\} \) is

\[
A_i = \begin{pmatrix}
A_{i,i} & A_{i,i-1} & A_{i,i-2} \\
A_{i-1,i} & A_{i-1,i-1} & A_{i-1,i-2} \\
A_{i-2,i} & A_{i-2,i-1} & A_{i-2,i-2}
\end{pmatrix}
\]

The optimal solution derived above for window \( \{i - 1, i, i + 1\} \) is equivalent to the extended 4D-Var problem: minimise

\[
\frac{1}{2} \left( \begin{array}{c}
\delta_i \\
\delta_{i-1}
\end{array} \right)^T \left( \begin{array}{cc}
A_{i,i} & A_{i,i-1} \\
A_{i-1,i} & A_{i-1,i-1}
\end{array} \right)^{-1} \left( \begin{array}{c}
\delta_i \\
\delta_{i-1}
\end{array} \right) + \\
\frac{1}{2} \sum_{i-1}^{i+1} (y_j^{(i+1)} - (x_j^g + \delta_j))^T R^{-1} (y_j^{(i+1)} - (x_j^g + \delta_j)) + \\
\frac{1}{2} (\delta_{i+1} - M_i^{i+1} \delta_i)^T Q^{-1} (\delta_{i+1} - M_i^{i+1} \delta_i)
\]

\( \odot \) Crown copyright. Met Office
First issue: size of states and covariances.

The state vectors are \( nL \times 1 \) and covariance matrices \( nL \times nL \), ie, enormous!

Several approximations possible, eg

Retain from the previous analysis stage merely \( x_{i-L+1}^a \) and its \( n \times n \) error covariance \( A_{i-L+1,i-L+1} \), which will be used as the background state and its covariance at the beginning of the new sliding window, to be performed by weak constraint 4D-Var.
Optimal solution:

\[
\begin{pmatrix}
{x^g_{i+1}} \\
x^g_i \\
x^g_{i-1}
\end{pmatrix} = \begin{pmatrix}
f^{i+1}_i(x^a_i) \\
x^a_i \\
x^a_{i-1}
\end{pmatrix}
\]

Then minimise

\[
\frac{1}{2} \left( \begin{array}{c}
\delta_i \\
\delta_{i-1}
\end{array} \right)^T \left( \begin{array}{cc}
A_{i,i} & A_{i,i-1} \\
A_{i-1,i} & A_{i-1,i-1}
\end{array} \right)^{-1} \left( \begin{array}{c}
\delta_i \\
\delta_{i-1}
\end{array} \right) + \\
\frac{1}{2} \sum_{i=1}^{i+1} (y_{j}^{(i+1)} - (x^g_j + \delta_j))^T R^{-1}(y_{j}^{(i+1)} - (x^g_j + \delta_j)) + \\
\frac{1}{2} (\delta_{i+1} - M^{i+1}_i \delta_i)^T Q^{-1}(\delta_{i+1} - M^{i+1}_i \delta_i)
\]

Then

\[
\begin{pmatrix}
x^a_{i+1} \\
x^a_i \\
x^a_{i-1}
\end{pmatrix} = \begin{pmatrix}
x^g_{i+1} \\
x^g_i \\
x^g_{i-1}
\end{pmatrix} + \begin{pmatrix}
\delta_{i+1} \\
\delta_i \\
\delta_{i-1}
\end{pmatrix}
\]
Approximation:

\[
\begin{pmatrix}
    x_{i+1}^g \\
    x_i^g \\
    x_{i-1}^g
\end{pmatrix} =
\begin{pmatrix}
    f_{i-1}^{i+1}(x_{i-1}^a) \\
    f_i^i(x_i^a) \\
    x_{i-1}^a
\end{pmatrix}
\]

Then minimise

\[
\frac{1}{2} \delta_{i-1}^T (A_{i-1,i-1})^{-1} \delta_{i-1} + \\
\frac{1}{2} \sum_{i-1}^{i+1} (y_j^{(i+1)} - (x_j^g + \delta_j))^T R^{-1} (y_j^{(i+1)} - (x_j^g + \delta_j)) + \\
\frac{1}{2} (\delta_i - M_{i-1}^i \delta_{i-1})^T Q^{-1} (\delta_i - M_{i-1}^i \delta_{i-1}) + \\
\frac{1}{2} (\delta_{i+1} - M_i^{i+1} \delta_i)^T Q^{-1} (\delta_{i+1} - M_i^{i+1} \delta_i)
\]

Then

\[
\begin{pmatrix}
    x_{i+1}^a \\
    x_i^a \\
    x_{i-1}^a
\end{pmatrix} =
\begin{pmatrix}
    x_{i+1}^g \\
    x_i^g \\
    x_{i-1}^g
\end{pmatrix} +
\begin{pmatrix}
    \delta_{i+1} \\
    \delta_i \\
    \delta_{i-1}
\end{pmatrix}
\]
Several other approximations possible, all sharing property that they converge to optimal solution as model error $Q \rightarrow 0$.

Practically, Issue 1 is likely to be tractable.

Maybe more significant is

**Second Issue, cycling of error covariances**

Currently, the background error covariances used for large scale DA are largely climatological. Overcoming this by use of longer window is still one option, though costly if performing analyses hourly.
Use of longer windows in RUC as one way to overcome shortcomings of fixed $B$.

Tiny example for illustrative purposes:

Suppose for each validity time $i$ two obs eventually become available, $y_{i}^{(i)}$ at time $i$ and $y_{i}^{(i+1)}$ at time $i + 1$.

Model is

$$x_{i+1} = r_{i}x_{i}$$

where $r_{i}$ is randomly drawn from $U[1, 2]$.

At time $i$ we assimilate (with fixed $B$) either

(1) $y_{i-1}^{(i)}$ and $y_{i}^{(i)}$;

(2) $y_{i-2}^{(i-1)}$, $y_{i-1}^{(i-1)}$, $y_{i-1}^{(i)}$ and $y_{i}^{(i)}$;

(3) $y_{i-4}^{(i-3)}$, $y_{i-3}^{(i-3)}$, $y_{i-3}^{(i-2)}$, $y_{i-2}^{(i-2)}$, $y_{i-2}^{(i-1)}$, $y_{i-1}^{(i-1)}$, $y_{i-1}^{(i)}$ and $y_{i}^{(i)}$;

© Crown copyright. Met Office
Observation validity time

Observation availability time
Mean square analysis error at end of window (so all methods have seen same obs) can be calculated in closed form.
So for hourly analyses we could do something like this ...

Observation validity time

Observation availability time

© Crown copyright. Met Office
Alternatively we can go back to the optimal solution and attempt to cycle covariances (approximately) correctly.

Several possibilities:

(a) Estimate posterior covariance from Var Hessian using limited memory quasi-Newton (‘LM Q-N’) machinery, and evolve to next time level by solving auxiliary (also LM Q-N) minimisation problem (cf ‘VKF’);

(b) As (a) but use CG/Lanczos (cf ‘EVIL’)

(c) Use ensemble square root filter (cf current hybrid)
Cycling of covariances by LM Q-N

Exploit following facts:

• Analysis error covariance $A$ in principle equals inverse of Hessian $H$ of $J$

• With LM Q-N methods such as BFGS, at each iteration the estimate of $H^{-1}$

\[
H_{est}^{-1} \leftarrow H_{est}^{-1} + \text{rank 2 matrix}
\]

• Given (a low rank estimate of) $A$ the inverse Hessian of

\[
x^T(MAM^T + Q)x
\]

is $B^{-1}$, the inverse of the background error covariance required at the next cycle.
Remarks

To get good approx to \( H_{est}^{-1} \) need to reformulate problem so that \( H \) is pert to identity

Can achieve this by preconditioning by \( U \) where

\[
UU^T = \text{diag}(B, Q, Q, ..)
\]

Then in limit of zero-sized subspace, method collapses to 4D-Var with non-cycled climatological \( B \).

The estimation of the analysis covariance is effectively zero cost.

Can increase rank of covariance approximations by

- Accumulating vector pairs over many cycles (LM Q-N well suited to this)
  (also, may be scope for parallelism in computation of vector pairs)
- Exploiting structure of problem (‘partial separability’)
More than one way to use ensemble SRF for covariances

One way is based directly on optimal strategy, where basic objects are $nL \times 1$ vectors and $nL \times nL$ covariance matrices

Estimate climatological $B_{clim}$ by bootstrapping, then

Run optimal strategy in its variational form (for the analysis) and ensemble SRF (for the adaptive part of the covariances) in tandem, using for the prior covariance $B$ a convex combination of $B_{clim}$ and the ensemble SRF estimate $B_{ens}$

In this case the ensemble is only run for one time slot, from $i - 1$ to $i$.  

© Crown copyright. Met Office
RMS analysis error, various ensemble sizes

RMS Analysis Error

© Crown copyright. Met Office
Concluding Remarks

There are several potential benefits from frequent update cycling, including timely boundary conditions for LAMs, and potentially better performance and affordability.

For frequent analyses the assimilation windows will normally overlap. An optimal Bayesian solution to this problem is readily obtained.

In the presence of model error this solution involves cycling large state vectors and covariance matrices. One can readily construct approximate approaches involving same-sized quantities to present.

To overcome the limitations of a largely climatological prior covariance matrix a long window approach is still possible, though now costly.

Alternatively, several options exist for covariance cycling, including the use of appropriately constructed ensembles.