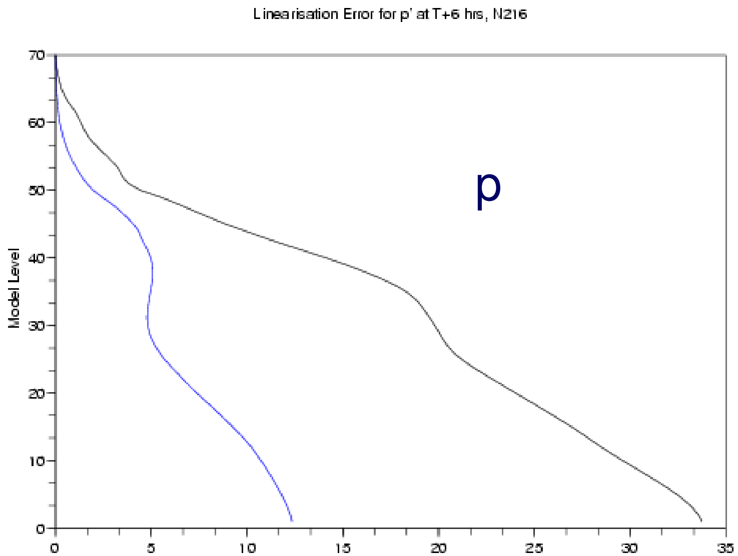
Accounting for Linearisation Error in the Extended Kalman Filter and 4D-Var

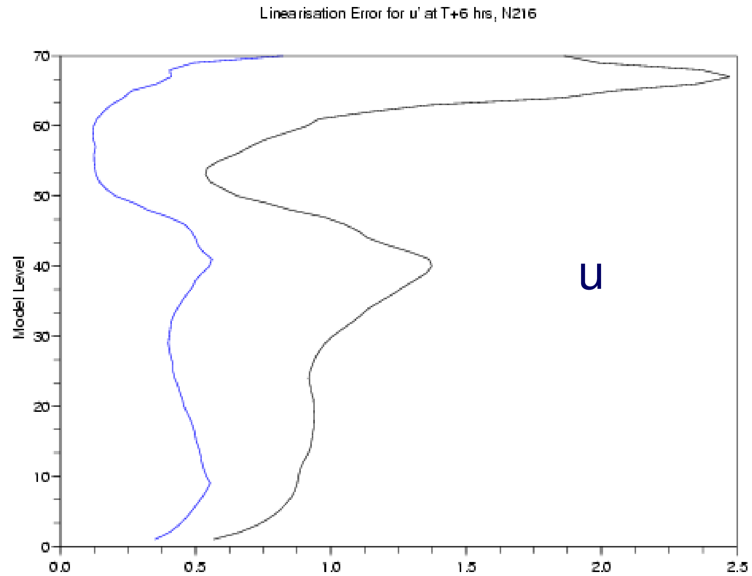
Tim Payne

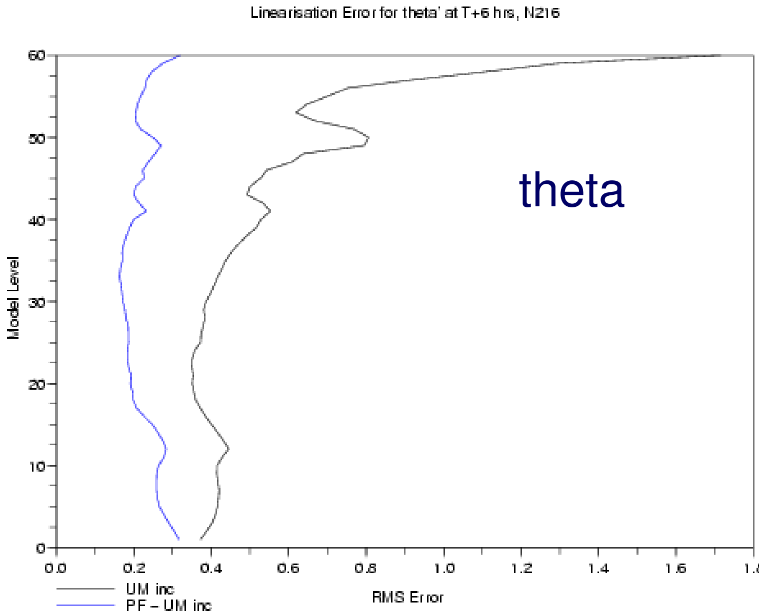
Ninth Workshop on Adjoint Model Applications in Dynamic Meteorology

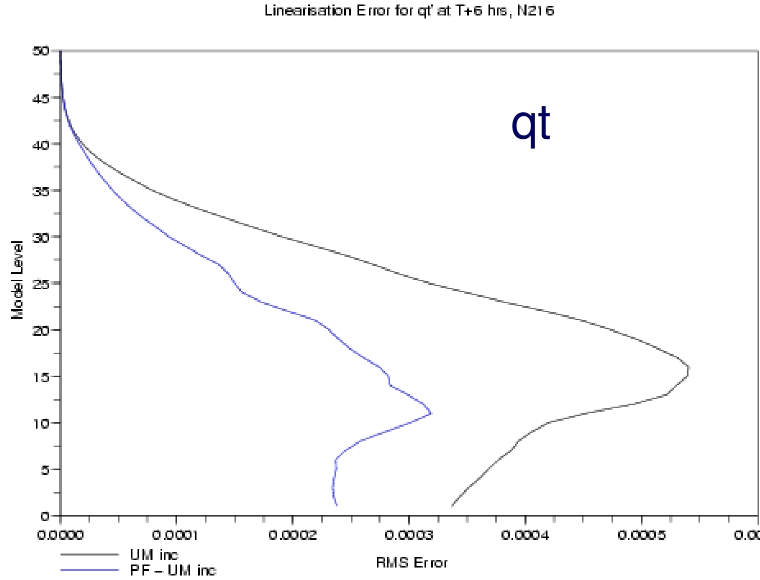
October 10th 2011

432X325X70 analysis grid, RMS Linearisation Error

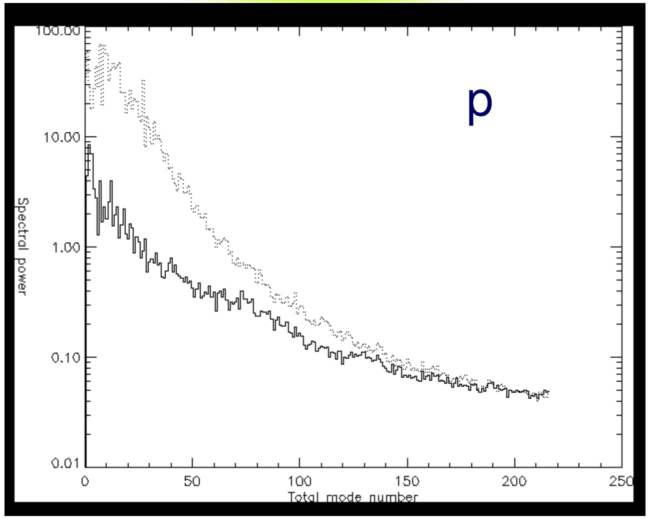


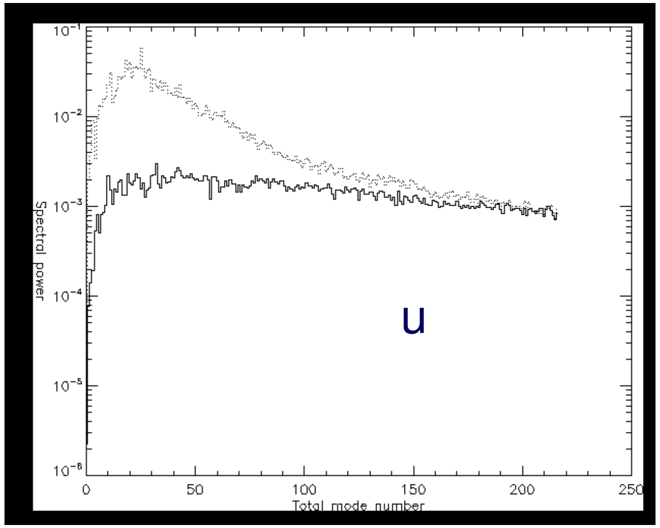


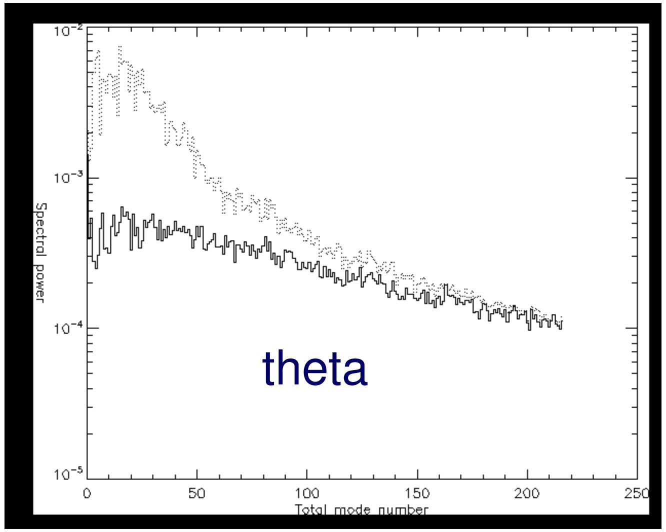


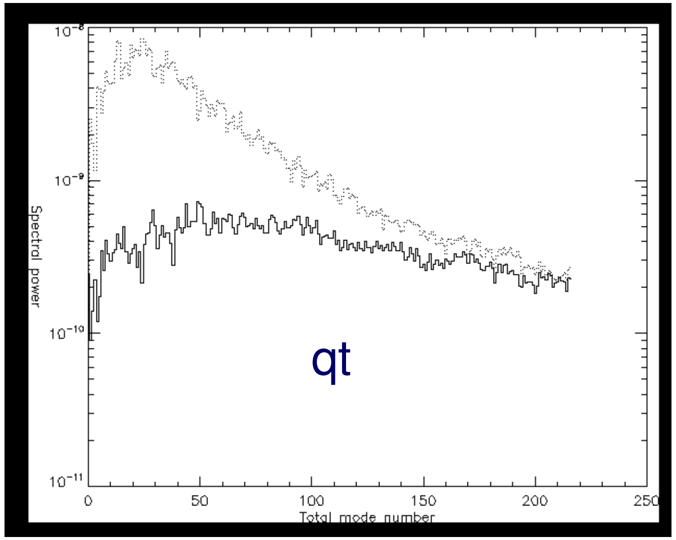


432X325X70 analysis grid, level 20 (~3km)

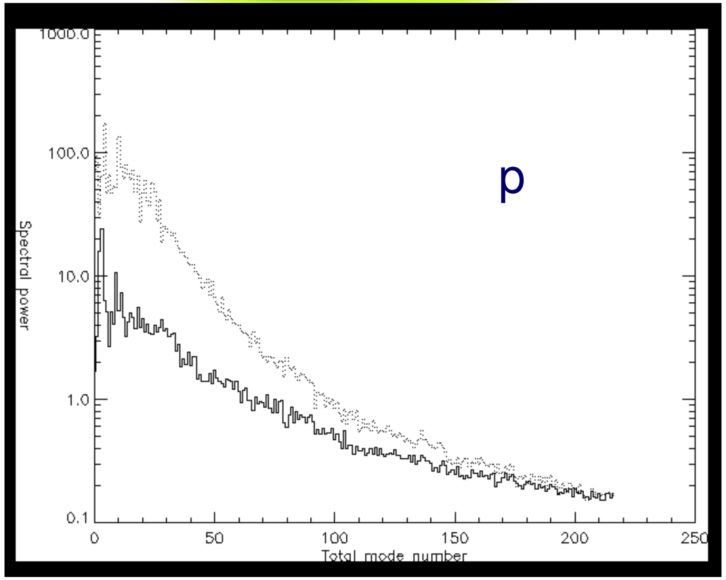


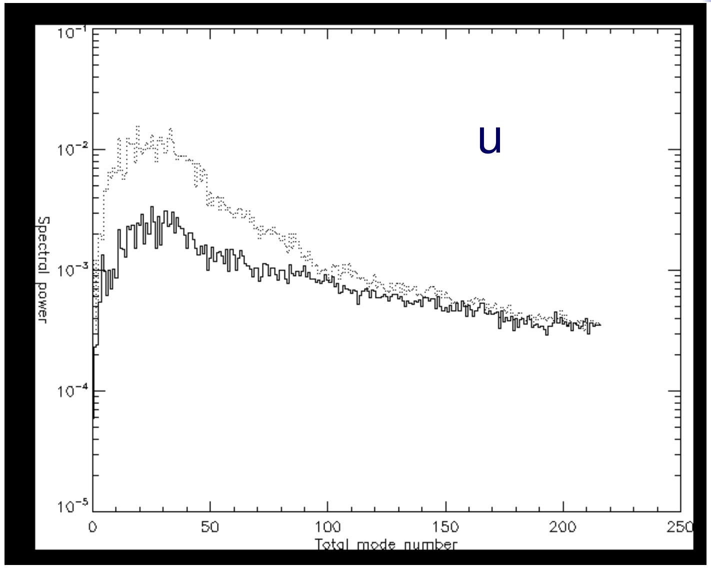


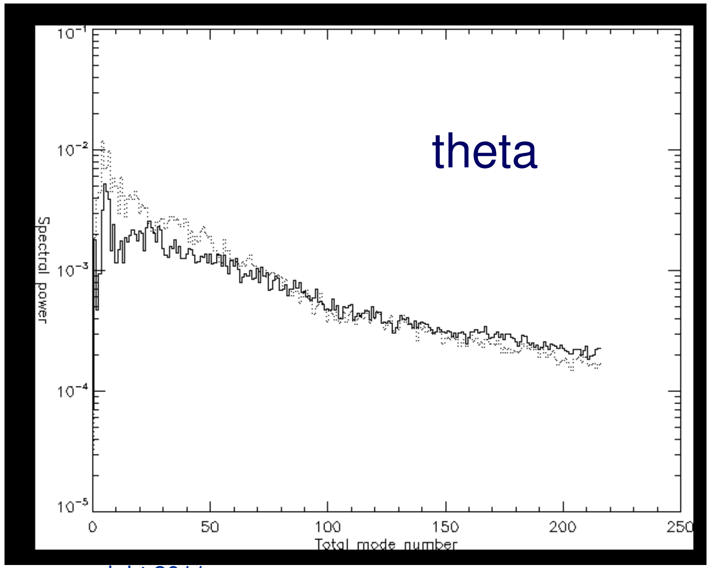


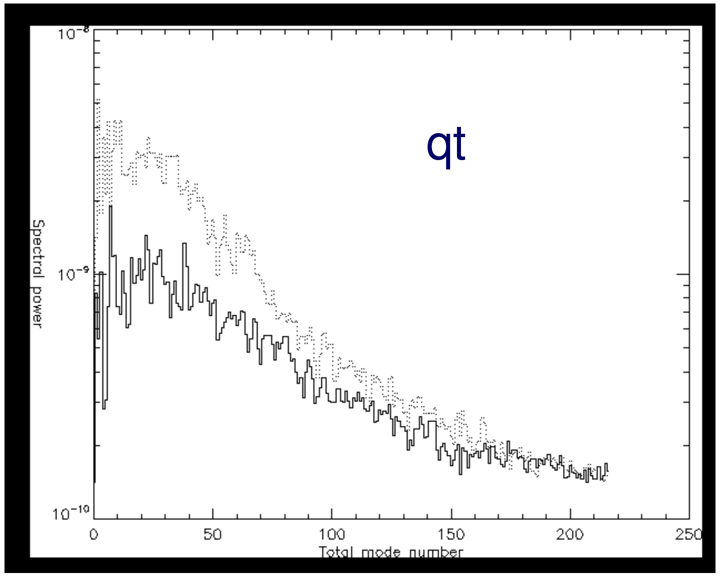


432X325X70 analysis grid, level 1

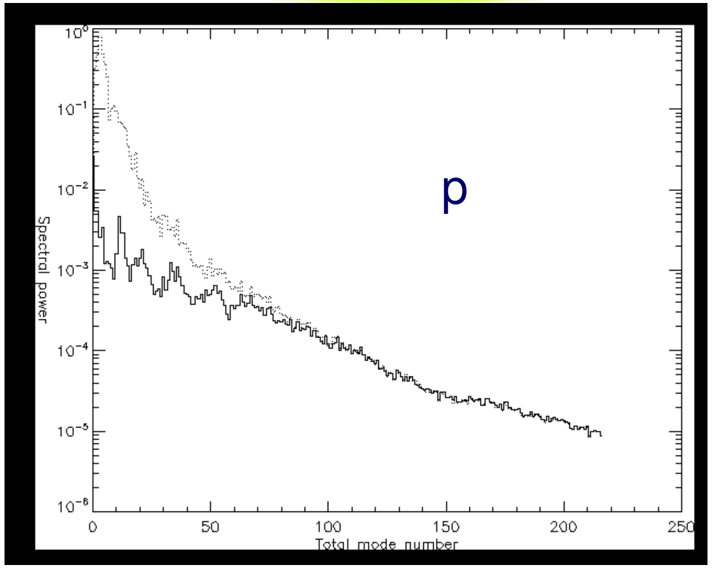


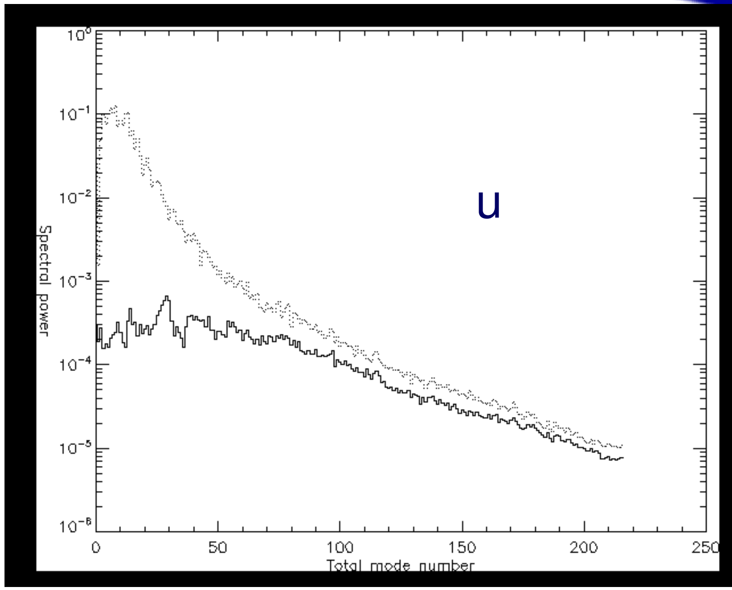


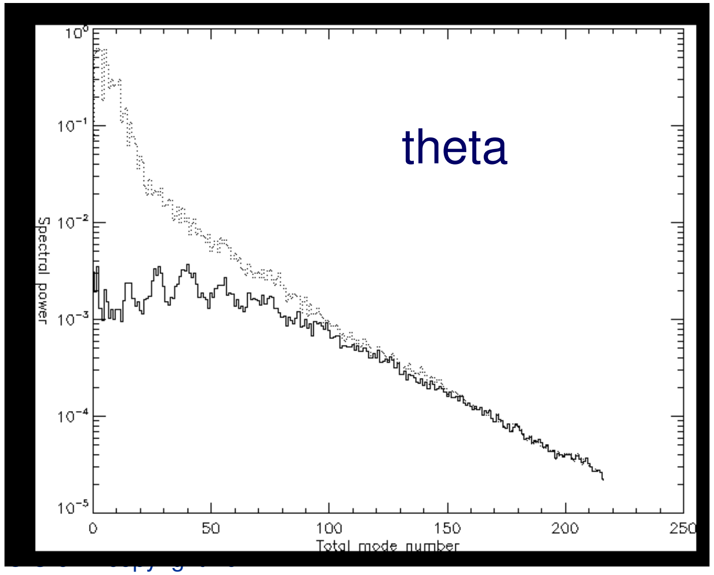




432X325X70 analysis grid, level 60







Making use of knowledge of errors in the linear model

- Lots of diagnostic information about linearisation errors
- Fairly consistent between different cases
- Plausible that in incremental 4D-Var the error from the linear ("PF") model is as large or larger than full ("UM") model error
- PF model error arises through
  Processes missing or approximated
  Lower resolution (as well as linearisation)

  But this is one case where we know what the errors are (cf background errors, full model error, even observation errors)

  Therefore unlimited scope to model them compactly

- How can we use this information to improve 4D-Var?

Optimal estimation where one has two models - linear case

We suppose we have a good model ('UM') M_i with small model error

$$\mathbf{x}_{i+1} = M_i \mathbf{x}_i + \boldsymbol{\epsilon}_i^M \tag{1}$$

and a second linear model M_i^P ('PF Model') which approximates the forecast of increments

$$M_i \mathbf{x}_i - M_i \mathbf{x}_i^g = M_i^p (\mathbf{x}_i - \mathbf{x}_i^g) + \boldsymbol{\epsilon}_i^{Mp} \tag{2}$$

where $\mathbf{x}_i^g$ are guess states.

As usual the observation model is

$$\mathbf{y}_i = H_i \mathbf{x}_i + \boldsymbol{\epsilon}_i^o$$

We will suppose $\boldsymbol{\epsilon}_i^M$, $\boldsymbol{\epsilon}_i^{Mp}$ and $\boldsymbol{\epsilon}_i^o$ are uncorrelated with

$$\boldsymbol{\epsilon}_i^M \sim \mathcal{N}(\mathbf{0}, Q^M), \ \boldsymbol{\epsilon}_i^{Mp} \sim \mathcal{N}(\mathbf{0}, Q^P), \ \boldsymbol{\epsilon}_i^o \sim \mathcal{N}(\mathbf{0}, R).$$

Linear case, cont'd

If we combine (1,2) we obtain

$$\mathbf{x}_{i+1} \;=\; M_i^p\mathbf{x}_i + [M_i\mathbf{x}_i^g - M_i^p\mathbf{x}_i^g] + \mathbf{w}_i$$

where $\mathbf{w}_i = \boldsymbol{\epsilon}_i^M + \boldsymbol{\epsilon}_i^{Mp}$ which is in the form of signal model for a KF with forcing $M_i\mathbf{x}_i^g - M_i^p\mathbf{x}_i^g$ and

$$\mathbf{w}_i \sim \mathcal{N}(\mathbf{0}, Q^M + Q^P)$$

There is also a variational equivalent: let

$$\boldsymbol{\delta}_i = \mathbf{x}_i - \mathbf{x}_i^g$$

Then if $\boldsymbol{\delta}_m$ is obtained by minimising

$$J = \boldsymbol{\delta}_0^T B^{-1}\boldsymbol{\delta}_0 + \frac{1}{2}\sum_{i=0}^{m}[\mathbf{y}_i - H(\mathbf{x}_i^g + \boldsymbol{\delta}_i)]^T R^{-1}[\mathbf{y}_i - H(\mathbf{x}_i^g + \boldsymbol{\delta}_i)]$$

$$+\frac{1}{2}\sum_{i=1}^{m}[\boldsymbol{\delta}_i - M_{i-1}^p\boldsymbol{\delta}_{i-1} + \mathbf{x}_i^g - M_{i-1}\mathbf{x}_{i-1}^g]^T Q^{-1}[\boldsymbol{\delta}_i - M_{i-1}^p\boldsymbol{\delta}_{i-1} + \mathbf{x}_i^g - M_{i-1}\mathbf{x}_{i-1}^g]$$

where $Q = Q^M + Q^P$, then $\mathbf{x}_m = \mathbf{x}_m^g + \boldsymbol{\delta}_m$ is identical to $\hat{\mathbf{x}}_{m|m}$ output from the Kalman Filter.

Nonlinear full model

Suppose we have a (fairly accurate) nonlinear full model f at high resolution

$$\mathbf{x}_{k+1} = f_k(\mathbf{x}_k) + \mathbf{w}_k \tag{3}$$

Let P be the projection from full to low resolution

$$\mathbf{z} = P\mathbf{x}$$

and P^+ is a pseudo-inverse of P which attempts to reconstruct the intermediate values by some form of interpolation.

$$PP^+ = I_{low\ res}$$

We wish to estimate the states $\mathbf{x}_1, \mathbf{x}_2, ...$ from observations $\mathbf{y}_1, \mathbf{y}_2, ...$

Linear model for evolution of increments

We have a *linear* map G_k which approximates the forecast of low resolution increments:

$$G_k(P\mathbf{x}_k - P\hat{\mathbf{x}}_{k|k}) \approx Pf_k(\mathbf{x}_k) - Pf_k(\hat{\mathbf{x}}_{k|k}) \tag{4}$$

If f_k was differentiable and easy to differentiate (and the increments were small) we would naturally take G_k to be the tangent-linear

$$G_k P = \frac{\partial}{\partial \mathbf{x}_k} P f_k(\mathbf{x}_k)\bigg|_{\mathbf{x}_k = \hat{\mathbf{x}}_{k|k}}$$

Linearisation error as a stochastic error

Decompose the right hand side of (3):

$$\begin{aligned}\mathbf{x}_{k+1} &= f_k(\mathbf{x}_k) + \mathbf{w}_k \\ &= f_k(\mathbf{x}_k) - f_k(\hat{\mathbf{x}}_{k|k}) + f_k(\hat{\mathbf{x}}_{k|k}) + \\ P^+G_k(P\mathbf{x}_k - P\hat{\mathbf{x}}_{k|k}) &- P^+G_k(P\mathbf{x}_k - P\hat{\mathbf{x}}_{k|k}) + \mathbf{w}_k\end{aligned}$$

we will consider the error in (4)

$$\zeta_k = f_k(\mathbf{x}_k) - f_k(\hat{\mathbf{x}}_{k|k}) - P^+G_k(P\mathbf{x}_k - P\hat{\mathbf{x}}_{k|k}) \tag{5}$$

as a stochastic error (akin to the model error $\mathbf{w}_k$), and

$$\mathbf{u}_k = f_k(\hat{\mathbf{x}}_{k|k}) - P^+G_kP\hat{\mathbf{x}}_{k|k}$$

as a forcing, leaving us with

$$\mathbf{x}_{k+1} = P^+G_kP\mathbf{x}_k + \mathbf{u}_k + \zeta_k + \mathbf{w}_k \tag{6}$$

Issues in forming EKF

If ζ_k and $\mathbf{w}_k$ were white (ie uncorrelated in time) and uncorrelated with each other then we would simply form

$$Q = E[\zeta_k \zeta_k^T] + E[\mathbf{w}_k \mathbf{w}_k^T]$$

and obtain a fairly standard looking (extended) Kalman Filter, in this case for our system with error in the linear model.

The main complications in forming an EKF are

(i) that the linearisation error ζ_k as defined in (5) is a function of the analysis (which is a function of the linear model), so estimates of covariance matrices will need to be iterated, and

(ii) that in practice the error ζ_k in the linear model is often strongly correlated in time.

Need for compact representation of error correlations

$$\mathbf{x}_{k+1} = P^{+} G_k P \mathbf{x}_k + \mathbf{u}_k + \zeta_k + \mathbf{w}_k$$

where $\mathbf{w}_k$ is white but linearisation error ζ_k is correlated in time.

In principle we get vast non-sparse matrix of error correlations of size (no of variables)$\times$ (number of time steps).

We get a much more compact representation if we approximate the correlations by supposing

$$E[\zeta_{i+j}\zeta_i^T] = A^j \tilde{Q}$$

some A, $\tilde{Q}$

Time correlated errors

If

$$\boldsymbol{\zeta}_{i+1} = A\boldsymbol{\zeta}_i + \boldsymbol{\eta}_i \tag{7}$$

where

$$\boldsymbol{\eta}_i \sim \mathcal{N}(0, E)$$

for some symmetric positive definite E, where

$$E[\boldsymbol{\zeta}_0 \boldsymbol{\eta}_i^T] = 0 \; for \; all \; i$$

and

$$E[\boldsymbol{\eta}_i \boldsymbol{\eta}_j^T] = 0 \; for \; all \; i \neq j$$

then let $\tilde{Q}$ satisfy

$$E = \tilde{Q} - A\tilde{Q}A^T$$

then as $i \to \infty$ we have

$$E[\boldsymbol{\zeta}_i \boldsymbol{\zeta}_i^T] \to \tilde{Q}$$
$$E[\boldsymbol{\zeta}_{i+j} \boldsymbol{\zeta}_i^T] \to A^j \tilde{Q}$$

Signal model for system with time correlated linearisation error

So take the signal model to be

$$\begin{pmatrix} \mathbf{x}_{k+1} \\ \zeta_{k+1} \end{pmatrix} = \begin{pmatrix} G_k & I \\ 0 & A_k \end{pmatrix} \begin{pmatrix} \mathbf{x}_k \\ \zeta_k \end{pmatrix} + \begin{pmatrix} \mathbf{u}_k \\ 0 \end{pmatrix} + \begin{pmatrix} \mathbf{w}_k \\ \boldsymbol{\eta}_k \end{pmatrix} \tag{8}$$

The first line of (8) is (6).

The second line is (7), our model for the evolution of linearisation errors.

We will denote the double-sized vectors by underlines, so (8) is written as

$$\underline{\mathbf{x}}_{k+1} = \underline{G}_k \underline{\mathbf{x}}_k + \underline{\mathbf{u}}_k + \underline{\mathbf{w}}_k$$

Similarly writing

$$\underline{H}_k = (H_k \;\; 0)$$

the observation model is now

$$\mathbf{y}_k = \underline{H}_k \underline{\mathbf{x}}_k + \mathbf{v}_k$$

EKF with time correlated linearisation error

If we write down a standard KF for (8) then we would have

$$\underline{Q} = \begin{pmatrix} cov(\mathbf{w}_k, \mathbf{w}_k) & cov(\mathbf{w}_k, \boldsymbol{\eta}_k) \\ cov(\boldsymbol{\eta}_k, \mathbf{w}_k) & cov(\boldsymbol{\eta}_k, \boldsymbol{\eta}_k) \end{pmatrix} = \begin{pmatrix} Q_k^M & Q_k^{MP} \\ Q_k^{PM} & Q_k^P \end{pmatrix}$$

The enhanced KF is then: for $k = 0, 1, .., n$

Predict

$$\begin{aligned} \underline{\hat{\mathbf{x}}}_{k|k-1} &= \underline{G}_k \underline{\hat{\mathbf{x}}}_{k-1|k-1} + \underline{\mathbf{u}}_{k-1} \\ \underline{P}_{k|k-1} &= \underline{G}_k \underline{P}_{k-1|k-1} (\underline{G}_k)^T + \underline{Q}_k \end{aligned} \tag{9}$$

Update

$$\begin{aligned} \underline{K}_k &= \underline{P}_{k|k-1} \underline{H}_k^T (\underline{H}_k \underline{P}_{k|k-1} \underline{H}_k^T + R_k)^{-1} \\ \underline{\hat{\mathbf{x}}}_{k|k} &= \underline{\hat{\mathbf{x}}}_{k|k-1} + \underline{K}_k (y_k - \underline{H}_k \underline{\hat{\mathbf{x}}}_{k|k-1}) \\ \underline{P}_{k|k} &= (I - \underline{K}_k \underline{H}_k) \underline{P}_{k|k-1} \end{aligned} \tag{10}$$

Parameters for filter including lineariston error

This filter has new parameters Q_k^{MP}, Q_k^P, A_k, associated with our model for the evolution of linearisation error. These are both inputs and outputs from the filter.

We will set

$$\underline{Q} = \begin{pmatrix} Q_k^M & 0 \\ 0 & Q_k^P \end{pmatrix}$$

ie, neglect $Q_k^{MP} = cov(\mathbf{w}_k, \boldsymbol{\eta}_k)$

$Q_k^M = E[\mathbf{w}_k \mathbf{w}_k^T]$ is full model error (important, but not the subject of this talk!).

We will neglect dependence on k leaving us with the need to determine parameters A and Q^P.

Determination of linearisation error parameters

This leaves need to estimate A and Q^P

If we set them in some fashion and run the KF (9,10) for $N \gg 1$ time steps and for $k = 1, .., N$ set

$$\boldsymbol{\Xi}(:, k) = f_k(\mathbf{x}_k) - f_k(\hat{\mathbf{x}}_{k|k}) - G_k(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k})$$

we can form the covariance matrix

$$\boldsymbol{\Upsilon} = \boldsymbol{\Upsilon}_0 = \boldsymbol{\Xi}\boldsymbol{\Xi}^T / N$$

Similarly, we may estimate cross correlation linearisation error matrices $\boldsymbol{\Upsilon}_j$ by setting

$$\boldsymbol{\Xi}_j = \boldsymbol{\Xi}(:, 1 : N - j)$$
$$\boldsymbol{\Xi}^j = \boldsymbol{\Xi}(:, 1 + j : N)$$

and

$$\boldsymbol{\Upsilon}_j = \boldsymbol{\Xi}^j \boldsymbol{\Xi}_j^T / (N - j) \quad (11)$$

$\boldsymbol{\Upsilon}_j$ is the covariance between the linearisation error on a given time step and the linearisation error on a time step j time steps away.

Determination of linearisation error parameters, cont'd

Comparing with foregoing, so long as $\boldsymbol{\Upsilon}_{j+1}\boldsymbol{\Upsilon}_j^{-1}$ approximately independent of j we can set

$$A = \boldsymbol{\Upsilon}_{j+1}\boldsymbol{\Upsilon}_j^{-1}$$
$$Q^P = \boldsymbol{\Upsilon}_0 - A\boldsymbol{\Upsilon}_0 A^T$$

In practice $\boldsymbol{\Upsilon}_{j+1}\boldsymbol{\Upsilon}_j^{-1}$ will not be entirely independent of j so we will need to make some approximation.

In summary, we may estimate A, Q^p by running the filter (9,10) with arbitrary A, Q^p (eg $A = 0$, $Q^P = 0$), measure $\boldsymbol{\Upsilon}_j$ as given by (11) and use these empirical values to estimate A, Q^p, and repeat. So long as the process converges then input A, Q^p and measured covariances $\boldsymbol{\Upsilon}_j$ will be consistent.

Example: L95, nearly perfect full model, persistence for linear model

Example - for the truth *and* full model f we use L95 with $n = 25$ variables, that is f is a single timestep integration (by fourth order Runge-kutta with time step 0.05) of

$$\frac{dx_i}{dt} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F$$

For the linear model G we go to an interesting extreme and set $M_P = Id$, ie, the linear model solves

$$\frac{dx_i}{dt} = 0 \qquad (!)$$

We follow above iterative procedure to obtain parameters.

Example: L95, nearly perfect full model, persistence for linear model,cont'd

We go one stage further with the approximation for time correlations, and set $A = \alpha I$, that is

$$\Upsilon_{j+1}\Upsilon_j^{-1} \approx \alpha I$$

where

$$\alpha = \frac{1}{2}\frac{\|\Upsilon_0 \circ \Upsilon_1\|}{\|\Upsilon_0\|^2} + \frac{1}{2}\sqrt{\frac{\|\Upsilon_0 \circ \Upsilon_2\|}{\|\Upsilon_0\|^2}}$$

As above begin with standard EKF (Q^M tiny but non-zero, optimised for TL as in Fisher et al 2007, and $A = Q^P = 0$), obtain first estimates for $A_k = \alpha * Id$, Q^P, and iterate ...

Example: L95, nearly perfect full model, persistence for linear model, results

Cycle	Mean square analysis error for time steps 100-5000 time-correlated KF	α	Mean square analysis error for time steps 100-5000 time-uncorrelated KF
0	20.6	0	20.6
1	1.05	0.71	0.58
2	0.188	0.77	0.33
3	0.112	0.80	0.20
4	0.093	0.77	0.19
5	0.083	0.77	0.17

cf mean square analysis error using exact TL (and optimal Q^M) of 0.0207

Variational version: weak constraint 4D-Var allowing for time correlated linearisation error

To cut a long-ish story short:

In the limit as $Q^M \to 0$ we need to minimise

$$J(\delta \mathbf{x}_0, .., \delta \mathbf{x}_m, \delta \boldsymbol{\zeta}_0, .., \delta \boldsymbol{\zeta}_{m-1}) =$$

$$\frac{1}{2}\delta \mathbf{x}_0^T B_x^{-1} \delta \mathbf{x}_0 + \frac{1}{2}\sum_{i=0}^{m} (\mathbf{y}_i - H_i(\mathbf{x}_i^g + \delta \mathbf{x}_i))^T R_i^{-1} (\mathbf{y}_i - H_i(\mathbf{x}_i^g + \delta \mathbf{x}_i)) +$$

$$\frac{1}{2}\delta \boldsymbol{\zeta}_0^T B_\zeta^{-1} \delta \boldsymbol{\zeta}_0 + \frac{1}{2}\sum_{i=0}^{m-2} (\delta \boldsymbol{\zeta}_{i+1} - A_i \delta \boldsymbol{\zeta}_i)^T Q^{P^{-1}} (\delta \boldsymbol{\zeta}_{i+1} - A_i \delta \boldsymbol{\zeta}_i)$$

subject to

$$\delta \boldsymbol{\zeta}_i = \delta \mathbf{x}_{i+1} + \mathbf{x}_{i+1}^g - M_i^p \delta \mathbf{x}_i - M_i \mathbf{x}_i^g,\ i = 0, .., m-1$$

where B_ζ is the prior for linearisation error $\boldsymbol{\zeta}$, corresponding to B as the prior error for $\mathbf{x}$, ie

$$\mathbf{x}_{0|-1} \sim \mathcal{N}(\overline{\mathbf{x}}_0, B_x),\quad \boldsymbol{\zeta}_{0|-1} \sim \mathcal{N}(\mathbf{0}, B_\zeta)$$

Remarks on variational form

If we write down the problem to be solved in the form find $\delta\mathbf{x}_0, .., \delta\mathbf{x}_m$ such that

$$J'(\delta\mathbf{x}_0, .., \delta\mathbf{x}_m) = \mathbf{0}$$

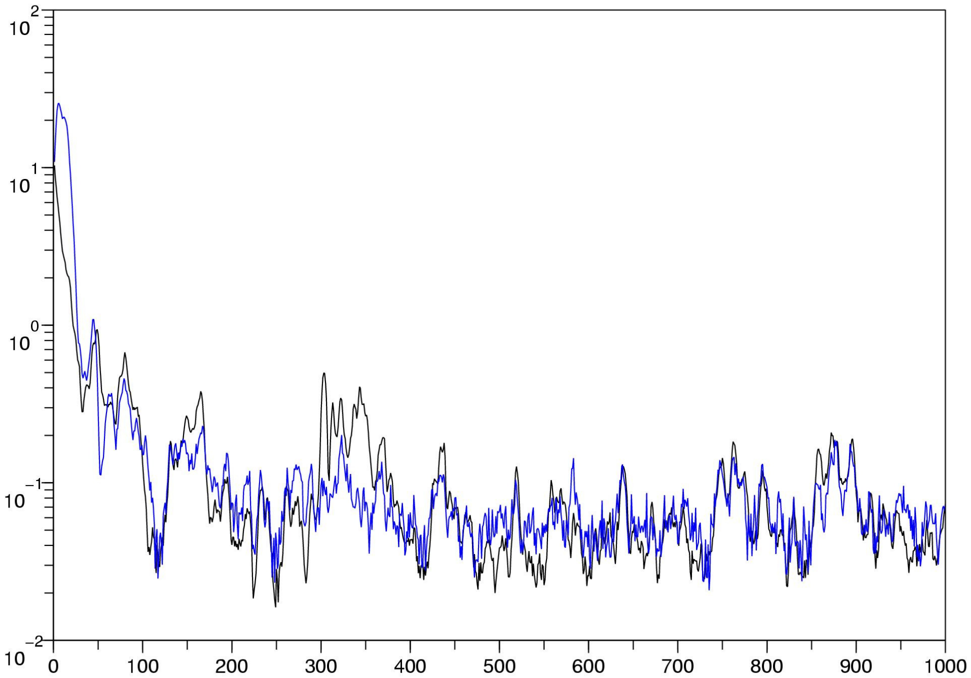
then whereas for standard weak constraint 4D-Var we had to solve a *block tridiagonal system*, we now need to solve a *block penta-diagonal one*.

Following Fisher et al, in the limit as window length $\rightarrow \infty$ we don't need prior for $\mathbf{x}$ or ζ

This is useful as we all know how sensitive results are to choice of B - so here we remove B entirely

If we minimise this cost function we obtain mean square analysis error as shown in blue on next slide

Long window weak constraint 4D-Var allowing for linearisation error, same example



4D-Var (blue) has mean square error of 0.074, slightly smaller than mean square error of 0.082 for EKF (black)

Remarks

Linearisation error has multiple sources, including missing physical processes and lower resolution. It is very significant in incremental 4D-Var, but unlike model error we have complete knowledge of it

We have shown that it is possible to account for it in the Extended Kalman Filter and incremental 4D-Var

In later iterations of the outer loop we expect analysis increments and hence linearisation error covariances to be smaller

In our example where the full model was L95 and the linear model was persistence, a simple allowance for linearisation error reduced RMS analysis error by a factor of 20, to only double what it would have been with exact tangent-linear.

One can view this variously as providing scope for:
- improved performance
- getting away with simpler (and hence cheaper) linear models
- providing insight into the relation between model error and linearisation error

The End