The Linear Model in 4D-Var

Tim Payne

May 21st 2009

Introduction

Non-incremental 4D-Var cost function is

$$J(\delta\mathbf{x}) = \frac{1}{2}\sum_{i=1}^{N}(\mathbf{y}_i - \mathcal{H}_i(\mathcal{M}_i(\mathbf{x}_0 + \delta\mathbf{x})))^T R_i^{-1}(\mathbf{y}_i - \mathcal{H}_i(\mathcal{M}_i(\mathbf{x}_0 + \delta\mathbf{x}))) + \frac{1}{2}\delta\mathbf{x}^T B^{-1}\delta\mathbf{x} \quad (1)$$

where $\mathbf{x}_0$ is the background state at time t_0, $\mathcal{M}_i, \mathcal{H}_i$ are respectively the full model evolution to time t_i and the observation operator at time t_i, $\mathbf{y}_i$ is a vector of observations at t_i, and B, R_i are background and observation error covariance matrices.

Suppose for present that observation operators $\mathcal{H}_i$ are linear

Tangent linear approximation is to replace $\mathcal{M}_i(\mathbf{x}_0 + \delta\mathbf{x})$ by $\mathcal{M}_i(\mathbf{x}_0) + \mathcal{M}_i'(\mathbf{x}_0)\delta\mathbf{x}$

Can we do better? - ie, can we find $M(\mathbf{x}_0), \boldsymbol{\ell}(\mathbf{x}_0)$ so that $\boldsymbol{\ell}(\mathbf{x}_0) + M(\mathbf{x}_0)\delta\mathbf{x}$

(a) better approximates $\mathcal{M}_i(\mathbf{x}_0 + \delta\mathbf{x})$ than $\mathcal{M}_i(\mathbf{x}_0) + \mathcal{M}_i'(\mathbf{x}_0)\delta\mathbf{x}$ does, or

(b) 'performs better' in Eqn (1) than $\mathcal{M}_i(\mathbf{x}_0) + \mathcal{M}_i'(\mathbf{x}_0)\delta\mathbf{x}$ does

Best linear approximation for functions

On a functional level certainly better choices to make than tangent linear, so long as we know what distribution of increments is.

Seek

$$\mathbf{f}(\mathbf{x}+\boldsymbol{\delta}) \approx \tilde{\mathbf{f}}(\mathbf{x}) + \tilde{\mathbf{T}}(\mathbf{x})\boldsymbol{\delta}$$

in such as way as to minimise

$$E\{[\mathbf{f}(\mathbf{x}+\boldsymbol{\delta}) - \tilde{\mathbf{f}}(\mathbf{x}) - \tilde{\mathbf{T}}(\mathbf{x})\boldsymbol{\delta}]^T A[\mathbf{f}(\mathbf{x}+\boldsymbol{\delta}) - \tilde{\mathbf{f}}(\mathbf{x}) - \tilde{\mathbf{T}}(\mathbf{x})\boldsymbol{\delta}]\}$$

where the expectation is over $\boldsymbol{\delta}$.
This is achieved by choosing $\tilde{\mathbf{f}}(\mathbf{x}), \tilde{\mathbf{T}}(\mathbf{x})$ to solve the simultaneous equations

$$\tilde{\mathbf{f}}(\mathbf{x}) = E[\mathbf{f}(\mathbf{x}+\boldsymbol{\delta}) - \tilde{\mathbf{T}}(\mathbf{x})\boldsymbol{\delta}]$$

$$\tilde{\mathbf{T}}(\mathbf{x}) = E[(\mathbf{f}(\mathbf{x}+\boldsymbol{\delta}) - \tilde{\mathbf{f}}(\mathbf{x}))\boldsymbol{\delta}^T]\{E[\boldsymbol{\delta}\boldsymbol{\delta}^T]\}^{-1}$$

ie

$$\tilde{\mathbf{T}} = [E\{\mathbf{f}(\mathbf{x}+\boldsymbol{\delta})\boldsymbol{\delta}^T\} - E\{\mathbf{f}(\mathbf{x}+\boldsymbol{\delta})\}E\{\boldsymbol{\delta}\}^T][E\{\boldsymbol{\delta}\boldsymbol{\delta}^T\} - E\{\boldsymbol{\delta}\}E\{\boldsymbol{\delta}\}^T]^{-1}$$

Data assimilation cycle

We will consider the cycled system (where subscript t denotes true state) the ith stage of which has the form

$$\mathbf{x}_t^i \xrightarrow{\mathbf{h}} \hat{\mathbf{x}}_t^i \xrightarrow{\mathbf{g}} \mathbf{x}_t^{i+1} \qquad \leftarrow ith\ cycle$$

with observations valid at intermediate point in cycle

$$\mathbf{y}^i = \hat{\mathbf{x}}_t^i + \boldsymbol{\epsilon}_o = \mathbf{h}(\mathbf{x}_t^i) + \boldsymbol{\epsilon}_o$$

The non-incremental 4D-Var cost function for the ith stage of this system then is

$$J = \frac{1}{2}(\mathbf{y}^i - \mathbf{h}(\mathbf{x}_b^i + \delta\mathbf{x}))^T R_i^{-1}(\mathbf{y}^i - \mathbf{h}(\mathbf{x}_b^i + \delta\mathbf{x})) + \frac{1}{2}\delta\mathbf{x}^T B^{-1}\delta\mathbf{x}$$

We could apply the above improved approximation to

$$\mathbf{h}(\mathbf{x}_b^i + \delta\mathbf{x})$$

so that our linear version of $\mathbf{h}$ is 'closer' to true $\mathbf{h}$ and therefore our incremental J is closer to full J than using the TL approximation would give. (This regularisation of the model will be called 'Reg-M'.)

Issue 1 – Is the conditional mode what we want?

By using Reg-M we are coming closest to minimising full J, ie we are finding the maximum of posterior distribution (conditional mode), but is this what we want?

Suppose

$$\mathbf{x}_t - \mathbf{x}_b = \boldsymbol{\epsilon}_b \sim N(\mathbf{0}, B), \ \mathbf{y} - H(h(\mathbf{x}_t)) = \boldsymbol{\epsilon}_o \sim N(\mathbf{0}, R)$$

The posterior pdf of $\mathbf{x}$ given $\mathbf{y}$ is

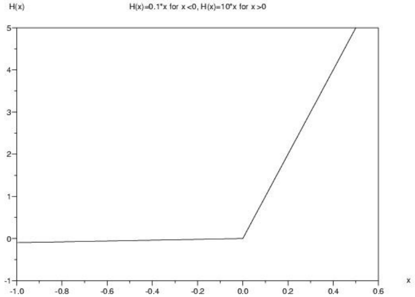
$$p(\mathbf{x}|\mathbf{y}) = \frac{e^{-J(\mathbf{x})}}{\int_{R^n} e^{-J(\mathbf{x})} d\mathbf{x}}$$

where $J(\mathbf{x})$ is

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{y} - H(\mathbf{h}(\mathbf{x})))^T R^{-1}(\mathbf{y} - H(\mathbf{h}(\mathbf{x}))) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_b) B^{-1}(\mathbf{x} - \mathbf{x}_b)$$

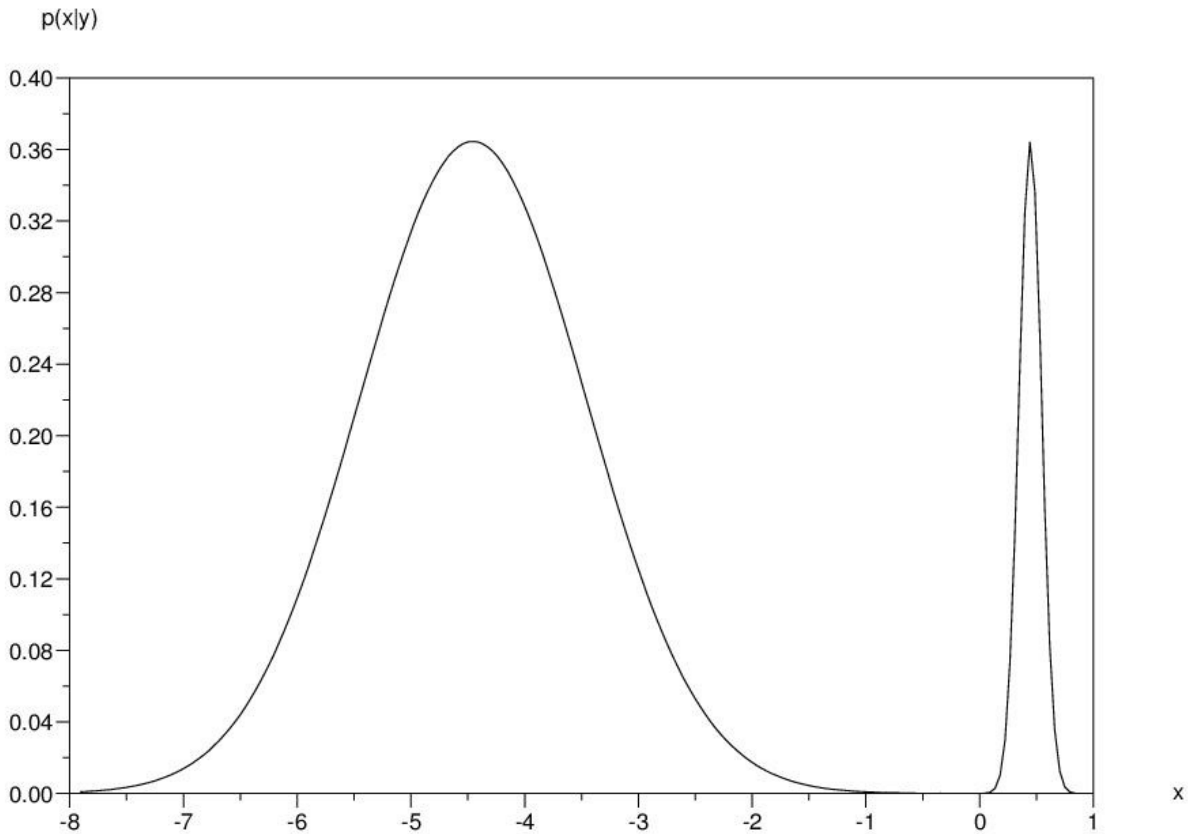
Small example: 3D-Var in 1D, with observation operator

$$H(x) = \begin{cases} 10x & if\ x \geq 0 \\ 0.1x & if\ x < 0 \end{cases}$$



Posterior pdf

$$x \sim N(-5, 1)$$
$$y = 5$$



4D-Var strategies

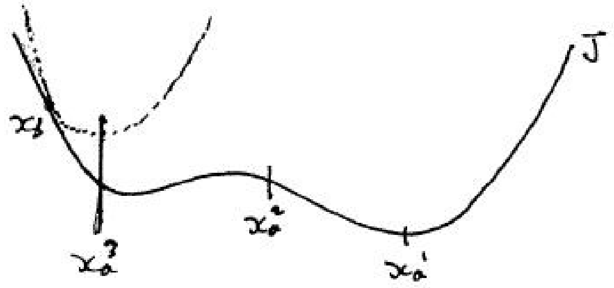
Now suppose forecast model is logistic map $h(x) = 4x(1-x)$, $g = I$.

$\sigma_o = 0.2$ and choose σ_b by ensemble method (using 4D-Var with TL).

At each cycle adopt one of three strategies:

Strategy 1: Global Minimum of J (Conditional Mode)
Strategy 2: Incremental 4D-Var with TL Model
Strategy 3: Conditional Mean $\mathbf{x}_a^e = E[\mathbf{x}|\mathbf{y}] = \int_{R^n} \mathbf{x}p(\mathbf{x}|\mathbf{y})d\mathbf{x} = \frac{\int_{R^n} \mathbf{x}e^{-J(\mathbf{x})}d\mathbf{x}}{\int_{R^n} e^{-J(\mathbf{x})}d\mathbf{x}}$

	$E[\vert\mathbf{x}_b - \mathbf{x}_t\vert^2]$
Minimise J (Conditional Mode)	0.098
Incremental 4D-Var with TL Model	0.081
Conditional Mean	0.069



.. So minimising true J is the worst thing to do!

Reg-A: Optimising linear model for analysis

Recall that in incremental 4D-Var

$$\mathbf{x}_a = \mathbf{x}_b + \boldsymbol{\delta}$$

where $\boldsymbol{\delta}$ is obtained in the inner loop by minimising, for some R, B

$$J(\boldsymbol{\delta}) = \frac{1}{2}\boldsymbol{\delta}^T B^{-1}\boldsymbol{\delta} + \frac{1}{2}(\mathbf{y} - \boldsymbol{\ell}(\mathbf{x}_b) - M(\mathbf{x}_b)\boldsymbol{\delta})^T R^{-1}(\mathbf{y} - \boldsymbol{\ell}(\mathbf{x}_b) - M(\mathbf{x}_b)\boldsymbol{\delta})$$

where $\boldsymbol{\ell}(\mathbf{x}_b) + M(\mathbf{x}_b)\boldsymbol{\delta}$ approximates $\mathbf{h}(\mathbf{x}_b + \boldsymbol{\delta})$
- eg in TL have $\boldsymbol{\ell}(\mathbf{x}_b) = \mathcal{H}(\mathcal{M}(\mathbf{x}_b))$ and $M(\mathbf{x}_b)\boldsymbol{\delta} = (\mathcal{H}\mathcal{M})'(\mathbf{x}_b)$

What about instead choosing $M(\mathbf{x}_b), \boldsymbol{\ell}(\mathbf{x}_b)$ to improve analysis directly?

- might do this by choosing analysis (i) to best approximate truth
or
(ii) to best approximate conditional mean (cf how successful this was in example above)

Reg-A: Optimising linear model for analysis

$\mathbf{x}_a$ minimising J on previous slide is

$$\mathbf{x}_a = \mathbf{x}_b + K(\mathbf{x}_b)(\mathbf{y} - \boldsymbol{\ell}(\mathbf{x}_b))$$

where

$$K(\mathbf{x}_b) = (B^{-1} + M^T R^{-1} M)^{-1} M^T R^{-1}$$

so choosing $M(\mathbf{x}_b), \boldsymbol{\ell}(\mathbf{x}_b)$ is equivalent to choosing $K(\mathbf{x}_b), \boldsymbol{\ell}(\mathbf{x}_b)$.

We first consider two problems where the objective is to choose $K(\mathbf{x}_b), \ell(\mathbf{x}_b)$ in such a way as to minimise the expected error in the analysis.

(Reg-A) Find matrix $K(\mathbf{x}_b)$ and vector $\ell(\mathbf{x}_b)$ which minimise the expected analysis error

$$E[\|\mathbf{x}_a - \mathbf{x}_t\|^2] = E[\|\mathbf{x}_b + K(\mathbf{x}_b)(\mathbf{h}(\mathbf{x}_b + \boldsymbol{\epsilon}_b) + \boldsymbol{\epsilon}_o - \ell(\mathbf{x}_b)) - \mathbf{x}_t\|^2]$$

where the expectation is over $\boldsymbol{\epsilon}_b, \boldsymbol{\epsilon}_o$ and $\|.\|$ denotes some norm.

Solution to Reg-A

For the norm (on analysis error etc) we will use

$$\|\mathbf{x}\|^2 = \mathbf{x}^T A \mathbf{x}$$

where A is some positive definite matrix. In Reg-A we seek to find vector $\boldsymbol{\ell}(\mathbf{x}_b)$ and matrix $K(\mathbf{x}_b)$ which minimises

$$E[(\mathbf{x}_t - \mathbf{x}_a)^T A(\mathbf{x}_t - \mathbf{x}_a)]$$

where

$$\mathbf{x}_a = \mathbf{x}_b + K(\mathbf{y} - \boldsymbol{\ell})$$

The solution to Reg-A is independent of the matrix A:

$$K = E[\{\mathbf{x}_t - \mathbf{x}_b - E[\mathbf{x}_t - \mathbf{x}_b]\}\mathbf{y}^T]E[\{\mathbf{y} - E[\mathbf{y}]\}\mathbf{y}^T]^{-1}$$

$$K\boldsymbol{\ell}(\mathbf{x}_b) = KE[\mathbf{y}] - E[\mathbf{x}_t - \mathbf{x}_b]$$

Reg-A': optimising linear model for conditional mean

As mentioned earlier, a popular single-value choice for the analysis is the conditional mean

$$\mathbf{x}_a^e = E[\mathbf{x}|\mathbf{y}] = \int_{R^n} \mathbf{x} p(\mathbf{x}|\mathbf{y}) d\mathbf{x}$$

as this is also the minimum variance solution. Note that unlike the maximum likelihood estimate the conditional mean is a function of the whole pdf

This suggests the second problem

(Reg-A$'$) find matrix $K(\mathbf{x}_b)$ and vector $\ell(\mathbf{x}_b)$ which minimise the expected error

$$E[\|\mathbf{x}_a - \mathbf{x}_a^e\|^2] = E[\|\mathbf{x}_b + K(\mathbf{x}_b)(\mathbf{h}(\mathbf{x}_b + \boldsymbol{\epsilon}_b) + \boldsymbol{\epsilon}_o - \ell(\mathbf{x}_b)) - \mathbf{x}_a^e\|^2]$$

where the expectation is again over $\boldsymbol{\epsilon}_b, \boldsymbol{\epsilon}_o$.

Solution to Reg-A’

Note that Reg-A$'$ has the identical formulation to Reg-A but with the conditional mean $\mathbf{x}_a^e$

$$\mathbf{x}_a^e = E[\mathbf{x}_t|\mathbf{y}] = \int \mathbf{x} p(\mathbf{x}|\mathbf{y}) d\mathbf{x}$$

replacing $\mathbf{x}_t$. The solution to Reg-A depended on $\mathbf{x}_t$ only through the terms $E[\mathbf{x}_t]$ and $E[\mathbf{x}_t\mathbf{y}^T]$. However

(i) Since $\mathbf{x}_a^e$ is a function of $\mathbf{y}$ only,

$$E[\mathbf{x}_a^e] = E[E[\mathbf{x}_t|\mathbf{y}]] = E[\mathbf{x}_t]$$

(ii)

$$E[\mathbf{x}_a^e\mathbf{y}^T] = \int\int \mathbf{x}\mathbf{y}^T p(\mathbf{y}) p(\mathbf{x}|\mathbf{y}) d\mathbf{x} d\mathbf{y}$$

and since by Baye’s theorem

$$p(\mathbf{x}|\mathbf{y})p(\mathbf{y}) = p(\mathbf{x}, \mathbf{y})$$

it follows that

$$E[\mathbf{x}_a^e\mathbf{y}^T] = \int\int \mathbf{x}\mathbf{y}^T p(\mathbf{x}, \mathbf{y}) d\mathbf{x} d\mathbf{y} = E[\mathbf{x}\mathbf{y}^T]$$

It follows that Reg-A and Reg-A$'$ have exactly the same solution!

Reg-A applied to Vis op

Out of interest we apply this strategy to our 3D-Var problem described above

To compute Reg-A need

$$E[\mathbf{x}_t - \mathbf{x}_b] \text{ and } E[(\mathbf{x}_t - \mathbf{x}_b)\mathbf{y}^T]$$

but would not normally have access to these quantities, so instead use

$$E[\mathbf{x}_a - \mathbf{x}_b] \text{ and } E[(\mathbf{x}_a - \mathbf{x}_b)\mathbf{y}^T]$$

Applied to 3D-Var vis problem impact of Reg-A is a minute improvement:

	$E[\lvert\mathbf{x}_b - \mathbf{x}_t\rvert^2]$
Minimise J (Conditional Mode)	0.098
Incremental 4D-Var with TL Model	0.081
Incremental 4D-Var using Reg-A	0.079
Conditional Mean	0.069

Expect larger impact using Reg-A on model

Issue 2: Will improving analysis improve forecasts?

Eg, 4D-Var cycled system in 1D, the ith stage of which has the form

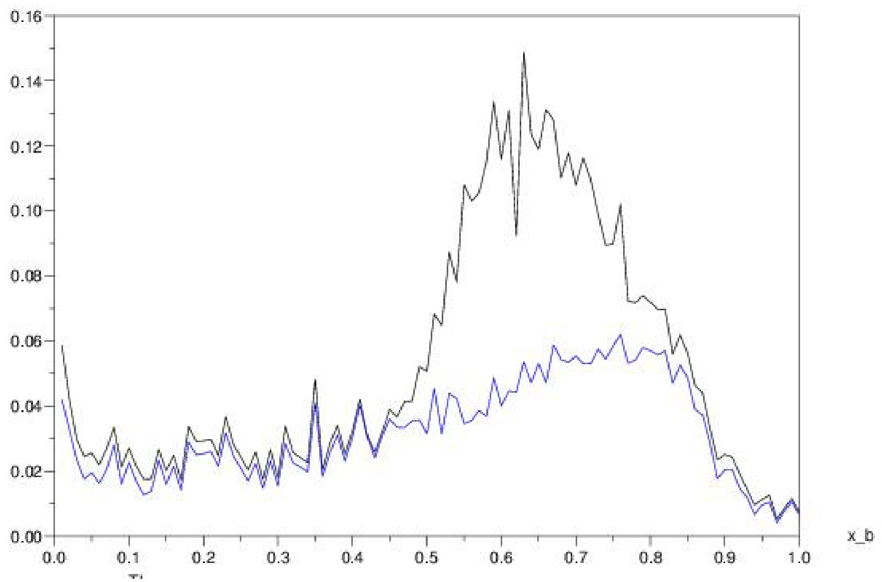
$$x_t^i \xrightarrow{h} \hat{x}_t^i \xrightarrow{g} x_t^{i+1}$$

h and g both logistic map $x \to 4x(1-x)$ on $[0,1]$. Observations valid at intermediate point in cycle $y^i = \hat{x}_t^i + \epsilon_o = h(x_t^i) + \epsilon_o$

Iterate DA cycle using 4D-Var with TL map; use this to compute distribution of background errors $x_b - x_t$

Use this distribution to compute for each x_b gain $K(x_b)$ and 'obs equivalent' $\ell(x_b)$ which minimise mean analysis error

Mean Square analysis error using TL (black) and Reg-A (blue)



Short forecast error

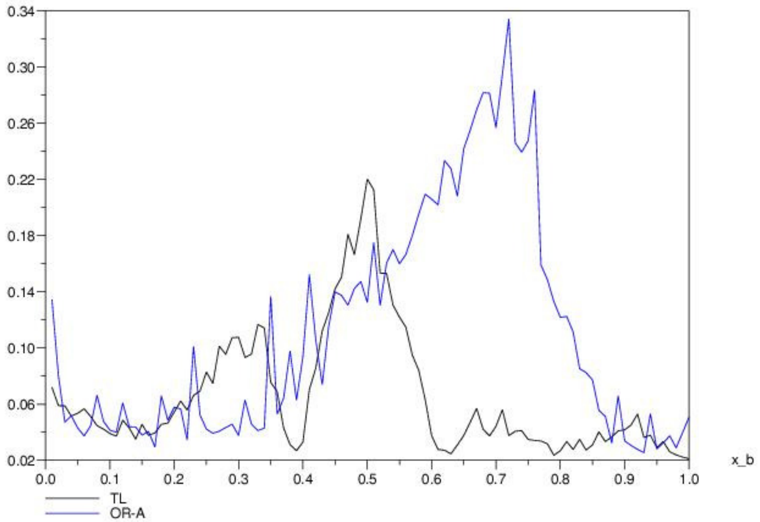
Using this $K(x_b)$ and $\ell(x_b)$ to calculate

$$x_a = x_b + K(x_b)(y - \boldsymbol{\ell}(x_b))$$

and hence background for next cycle

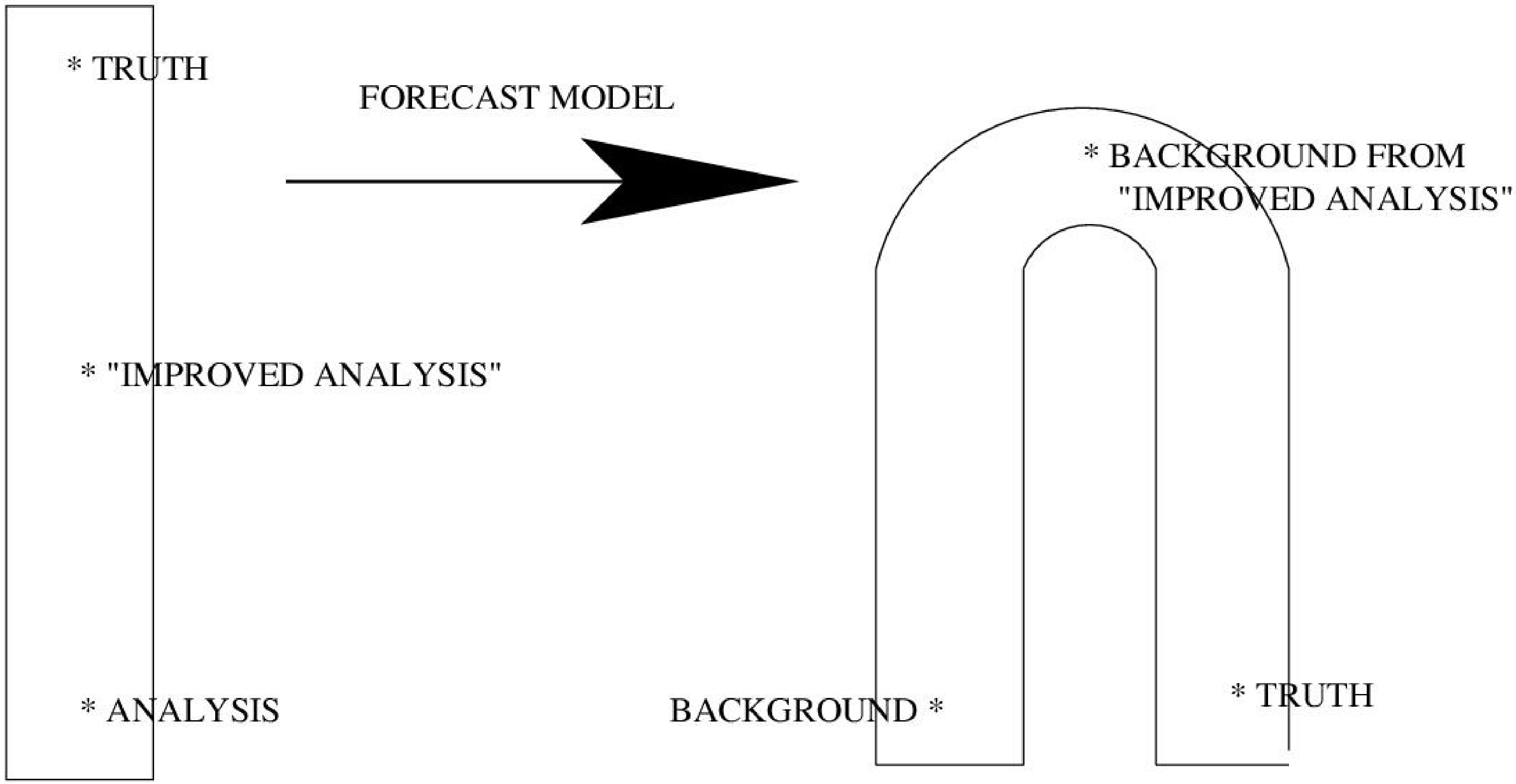
$$x_b^{new} = gh(x_a)$$

Mean Square short forecast error using TL (black) and Reg-A (blue)



Applying Reg-A improves analysis but can degrade next background and therefore subsequent analysis, so overall impact can be negative!

Schematic of why improving analysis may not improve short forecast



Reg-B: Optimising linear model for short forecast

Foregoing suggests the alternative optimization problem:

Find $\boldsymbol{\ell}$ and gain $K(\mathbf{x}_b)$, so that the *next* background best estimates the true state at that time, which since

$$\mathbf{x}_a = \mathbf{x}_b + K(\mathbf{y} - \boldsymbol{\ell})$$

means we seek to minimise

$$E[\|\mathbf{gh}\{\mathbf{x}_b + K(\mathbf{x}_b)(\mathbf{h}(\mathbf{x}_b + \boldsymbol{\epsilon}_b) + \boldsymbol{\epsilon}_o - \boldsymbol{\ell}(\mathbf{x}_b))\} - \mathbf{gh}(\mathbf{x}_t)\|^2]$$

over $K, \boldsymbol{\ell}$.

Unlike Reg-A this is in general a non-quadratic minimisation problem (potentially with more than one local minimum) which is solved numerically.

It is numerically more stable to use as independent variables $\{K, K\ell\}$ rather than $\{K, \ell\}$. We have used a conjugate-gradient method with iterative line searches (that is, non-exact line searches using cubic interpolation/extrapolation).

Reg-B applied to logistic map

Reg-B harder and more expensive to compute

Very effective even for cases for which TL well-suited (ie when increments small compared with scale of nonlinearites)

Eg, 4D-Var in 1D on $[0,1]$ with

$$h(x) = g(x) = 4x(1-x), \; H = I$$

(this example computed using true background errors)

	TL	Reg-B
Mean Square Analysis Error	0.039	0.010
Mean Square short forecast Error	0.045	0.020

Table 1: Mean square errors for cycled forecast-analysis system using logistic map

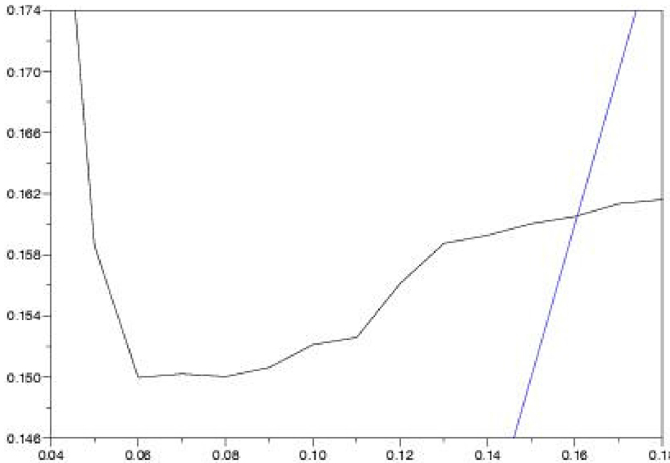
More Issues

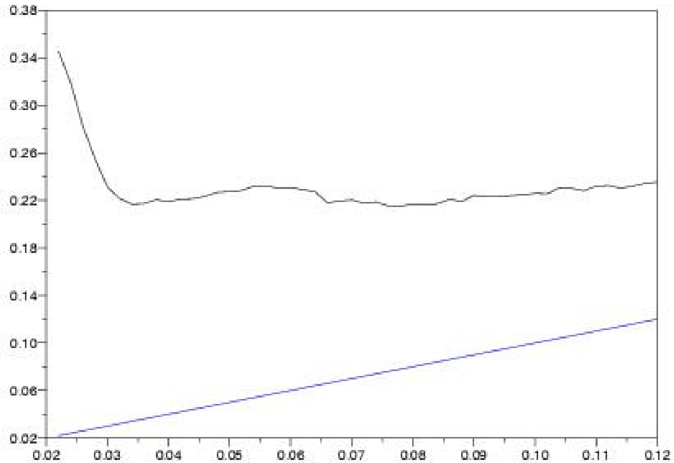
What pdf to use: $\mathbf{x}_b - \mathbf{x}_t$ or $\mathbf{x}_b - \mathbf{x}_a$ etc

What to do about variable obs network: strictly Reg-A or Reg-B would need to be recomputed every analysis

Lack of chain rule

Choice of B. To calculate $B = E[(\mathbf{x}_t - \mathbf{x}_b)(\mathbf{x}_t - \mathbf{x}_b)^T]$ need DA system for which need B.





σ_b^{in} on x-axis versus σ_b^{out} on y-axis for *left* shift and *right* logistic maps. The blue curve is the leading diagonal $\sigma_b^{out} = \sigma_b^{in}$.

Background covariance matrix B and linear model are inextricably intertwined. Reg-A and Reg-B effectively also determine B. For other methodologies for linear model conclusion may depend on how we choose B.

Reg-M revisited

Notwithstanding the many issues, if TL is poor enough then even our simplest regularisation, Reg-M, is very effective

Recall Reg-M uses best linear approximation of model directly (not requiring solution of matrix equation) and cheap to compute as gives rise to quadratic minimisation problem

Use $\mathbf{x}_a - \mathbf{x}_b$ in place of $\mathbf{x}_t - \mathbf{x}_b$

(Reg-M) Find PF Model $M(\mathbf{x}_b)$ and vector $\boldsymbol{\ell}(\mathbf{x}_b)$ which minimise the expected error

$$E[\|\mathbf{h}(\mathbf{x}_b + \boldsymbol{\delta}) - M(\mathbf{x}_b)\boldsymbol{\delta} - \boldsymbol{\ell}(\mathbf{x}_b)\|^2]$$

where the expectation is over analysis increments $\boldsymbol{\delta} = \mathbf{x}_a - \mathbf{x}_b$, the analysis being that produced by a TL model (one could then iterate).

Following the above the solution is

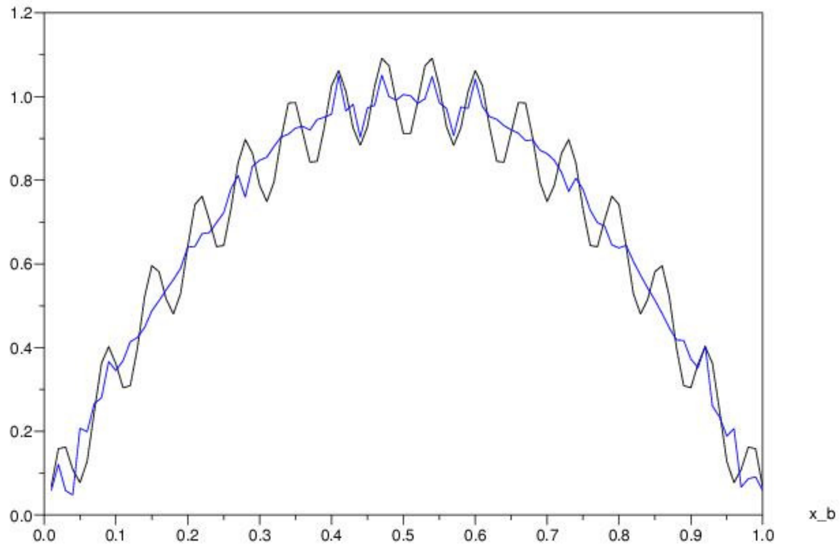
$$\begin{aligned} M(\mathbf{x}_b) = \{E[\mathbf{h}(\mathbf{x}_b + \boldsymbol{\delta})\boldsymbol{\delta}^T] - E[\mathbf{h}(\mathbf{x}_b + \boldsymbol{\delta})]E[\boldsymbol{\delta}]^T\}\{E[\boldsymbol{\delta}\boldsymbol{\delta}^T] - E[\boldsymbol{\delta}]E[\boldsymbol{\delta}]^T\}^{-1} \\ \boldsymbol{\ell}(\mathbf{x}_b) = E[\mathbf{h}(\mathbf{x}_b + \boldsymbol{\delta})] - M(\mathbf{x}_b)E[\boldsymbol{\delta}] \end{aligned}$$

TL versus Reg-M applied to modified logistic map

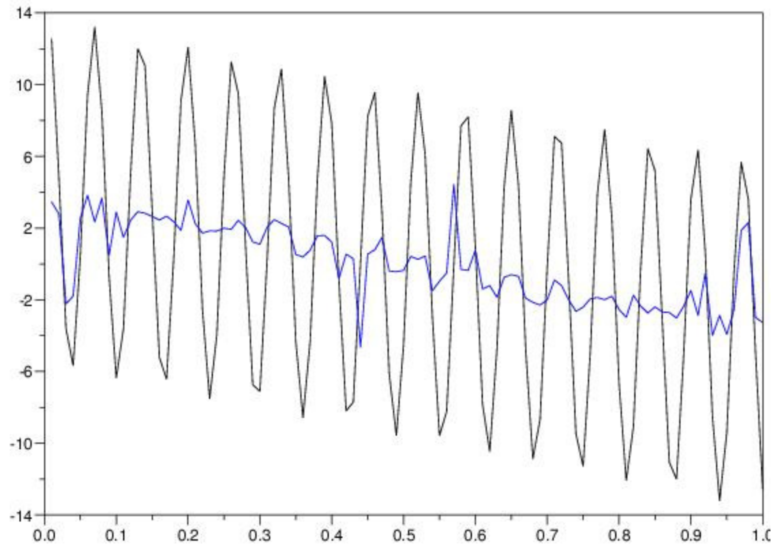
Example

$$
\begin{aligned}
h(x) = 4[x](1-[x]) + \gamma sin(K\pi[x])\ where\ [x] = x, mod\ 1 \\
g(x) = H(x) = I \\
y \sim N(h(x), 0.04) \\
x - x_b \sim N(0, 0.034)
\end{aligned}
$$

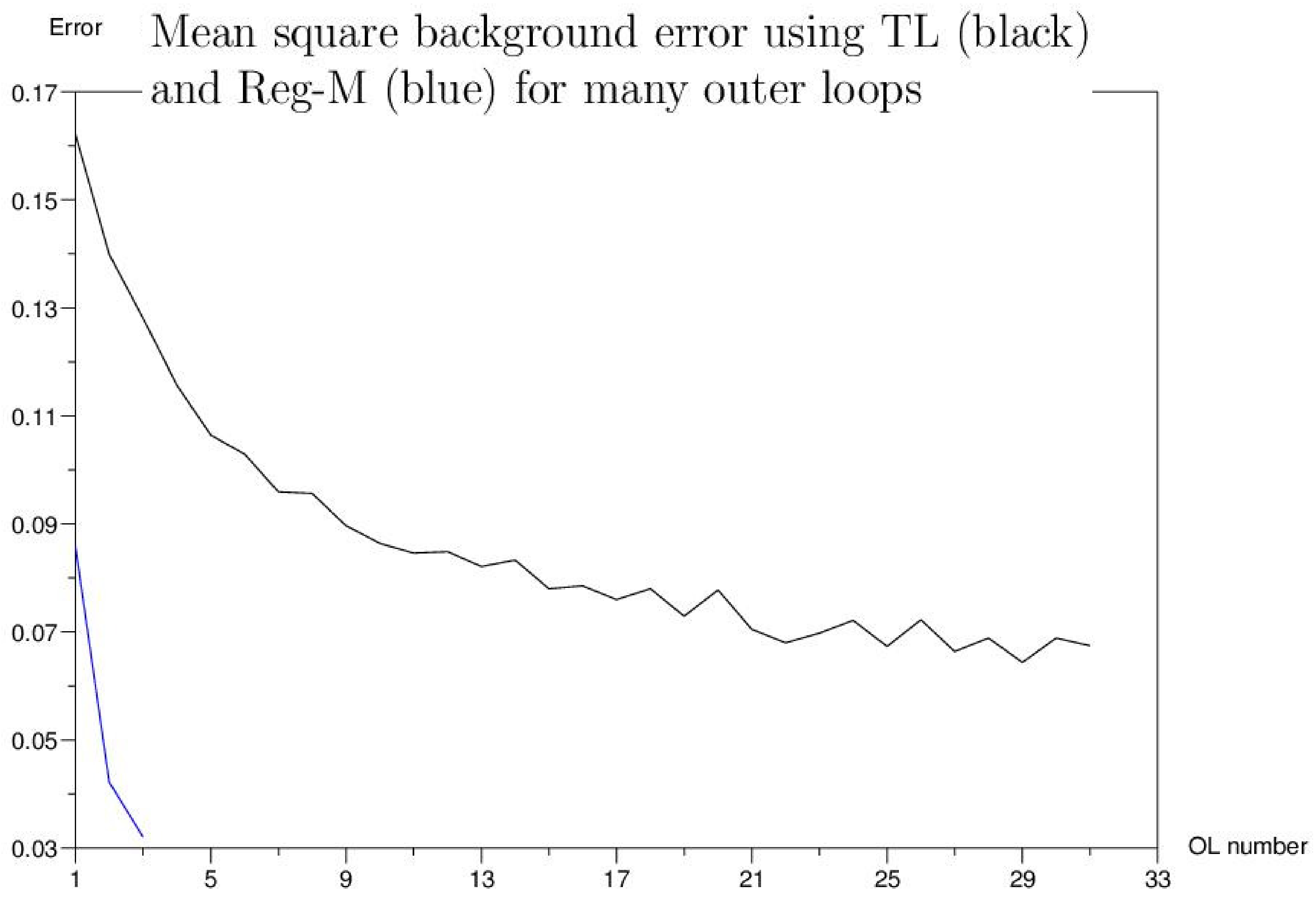
ℓ using TL (black) and Reg-M (blue)



PF using TL (black) and Reg-M (blue)



TL versus Reg-M with many outer loops



Summary

As an alternative to the TL approach we have considered other ways of formulating M, ℓ (ie the methods we have termed Reg-M,A,A$'$,B) which are obtained by solving optimisation problems in advance of the assimilation

Some issues need to be borne in mind:
non-incremental 4D-Var does not necessarily outperform incremental 4D-Var with a TL model, with the implication that finding a better approximation to the full model may not always be beneficial;
choosing the linear model to reduce analysis error may increase short forecast error and so be self-defeating

Reg-B will in practice always outperform TL but is very hard to implement

Reg-M is the simplest methodology and makes several approximations/assumptions. It is still *quite* hard to implement in realistic systems, but at least has a simple analytic form and is the same cost as TL in running.

Reg-M can be extremely effective if TL approximation poor