4D-Var in the presence of model error

Mike Cullen

3 August 2011

Acknowledgements

The 3 body model code was written by Gordon Inverarity.

Contents

This presentation covers the following areas

- Background
- Do we want the real R matrix?
- Do we want the real Q matrix?
- Do we need a long window?

Background

Analysis theory

Nearly all data assimilation theory is based on the idea of a statistically optimal analysis.

This shows that the relative weighting of different pieces of information depends on their errors.

If linear and Gaussian assumptions are made, this leads to the standard Var formulation.

More generally, it is sensible to give greater weight to more accurate pieces of information.

Model error

Standard theory allows for model error.

Increments to background state made to compensate for error growth and separately for model error.

Requires a model of model error, e.g. (ECMWF) a component which is Gaussian with zero mean and covariance Q, and an explicit non-random part.

Perfect stochastic model

If the non-random part can be expressed in terms of model variables and statistically calibrated parameters, this implies creation of a ‘perfect stochastic model’, a trajectory from which is then fitted to the observations in an optimal way.

If the non-random part can only be determined from the observations in each assimilation window, then it is a form of analysis increment which should be minimised in the normal way.

Do we want the real R matrix?

The real situation

The analysis to be used in NWP has to be consistent with the model's representation of the truth.

The deficiencies in the model representation should therefore not be included in the model error.

A 'perfect' model initial state can be written as $\mathbf{x}=\mathbf{Sz}$, where $\mathbf{z}$ is the truth and $\mathbf{S}$ a simplification operator.

If the model evolution operator is $\mathbf{M}$ and the true evolution operator $\mathbf{N}$, the model evolution error which has to be taken into account in DA is $\mathbf{MSz}-\mathbf{SNz}$.

Errors of representation

The simplification operator includes:

- Space-time averaging
- Use of model's discrete representation of hydrostatic and geostrophic balance
- Replacement of the effect of real orography by that of the model orography
- etc.

Effect on observation error

Observed innovations should not be used to drive the model state towards the true state **z** from the model ‘true’ state **Sz**.

Likely to happen if there are lots of good observations.

While some of the effects of **S** are the filtering of random noise, much is systematic.

Thus if they are treated by inflating the observation error, there will be correlations between the observation error from different platforms.

Forecast accuracy

In NWP, the analysis should be optimised on the basis of giving the best forecast.

If the model evolution error is zero, then the optimum analysis would be $\mathbf{x}=\mathbf{S}\mathbf{z}$ and a statistically optimal estimate of this is required.

If the evolution error is non-zero this may not be true.

It may be better to match time tendencies with observed time tendencies, a natural potential benefit of 4d-Var.

Do we want the real Q matrix?

Compensating model error

Consider cycled 4D-Var with all obs at the end of the window.

Standard Kalman filter equation for predicting background error for the next cycle, in standard notation, and assuming random model error, is

$$\mathbf{MAM}^{\mathrm{T}} + \mathbf{Q} = \mathbf{B}$$

Given plentiful good observations, the state at the end of the window will be close to the truth. If the model error is relatively large, this implies a negative correlation between $\mathbf{MAM}^{\mathrm{T}}$ and $\mathbf{Q}$, and a small $\mathbf{B}$.

Effects

Assume B remains constant from cycle to cycle. Then B will always be small and

$$\left|\mathbf{MAM}^{\mathrm{T}}\right| = \left|\mathbf{Q}\right|$$

If M amplifies perturbations, then

$$\left|\mathbf{A}\right| << \left|\mathbf{Q}\right|$$

And it is consistent with both these relations and the analysis equation to have

$$\left|\mathbf{A}\right| \leq \left|\mathbf{B}\right| << \left|\mathbf{Q}\right|$$

This behaviour is characteristic of a smoother rather than a filter

3 body model

Use classical 3 body model configured as sun/planet/moon system.

Slow timescale associated with planet/moon system's orbit round sun (7.2 time units). Labelled 'sun'

Fast timescale with moon's orbit round planet (.54 time units). Labelled 'moon'.

Choose assimilation period of 0.3 time units, about 50% of moon's orbital period

Experiments

Use cycled 4dVar, window length 0.3, regularisation matrix (‘background’ error) **C** estimated from statistics of (analysis-background) over set of 300 cycles.

Repeat recalculation and rerun cycles till assimilation fails because of ill-conditioned **C** matrix.

1 observation of all 12 variables per assimilation cycle

Show results in relative coordinates (8 d.o.f.), errors in ‘sun’ correspond to slow timescale, those in ‘moon’ to fast timescale. These coordinates affected by representation errors.

Model errors

Truth

- Sun mass = 1.0
- Planet mass = 0.1
- Moon mass = 0.01

Model

- Sun mass = 1.0
- Planet mass = 0.101
- Moon mass = 0.01

Model error small, so both slow and fast motions can be skilfully predicted.

After time 0.3, model error accumulation is 0.003 for the sun's position and 0.003 for the moon's position.

Error of representation due to shifted centres of mass, 0.004 for moon, negligible for sun.

Analysis and background errors

Comments using ‘best’ cycles.

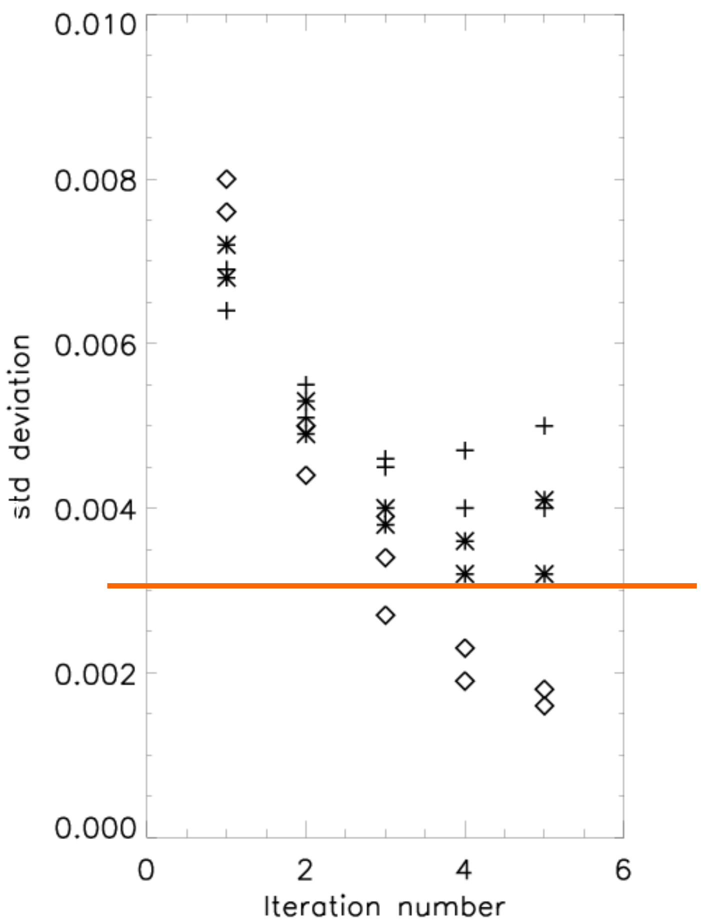
Sun: **A** 20% smaller than **B**, both larger than model error accumulation. Analysis increments and optimal **C** matrix smaller than model error accumulation.

Moon: **A** almost equal to **B**, analysis increments 30 times smaller, **A** and **B** comparable to representation error and larger than model error accumulation. Analysis increments and optimal **C** matrix 20 times smaller than model error accumulation.

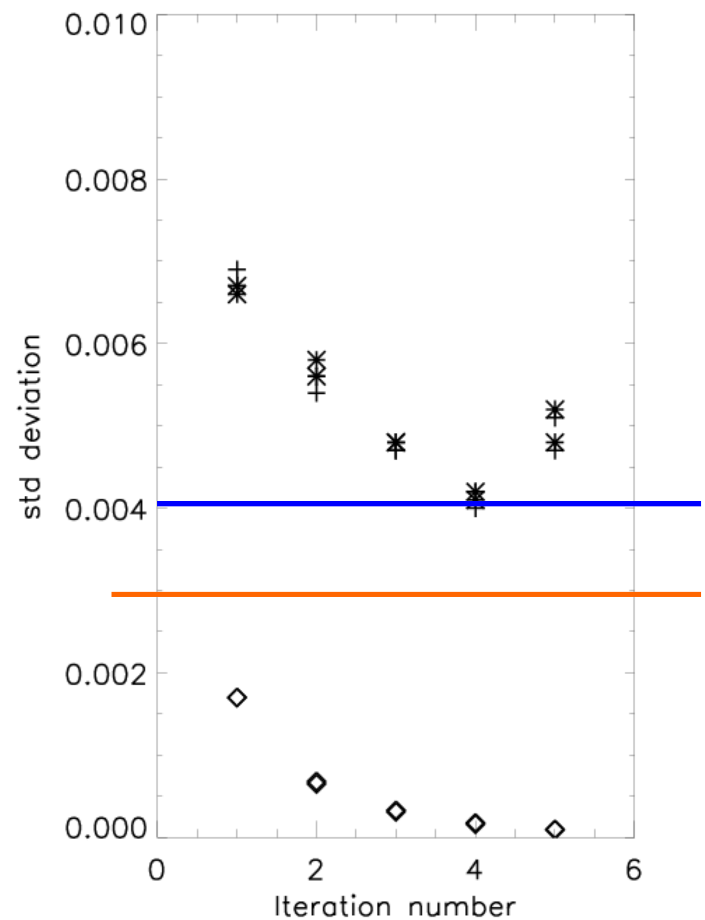
Met Office

+=(B-T),*=(A-T),◊=(A-B)

Sun



Moon



Model error

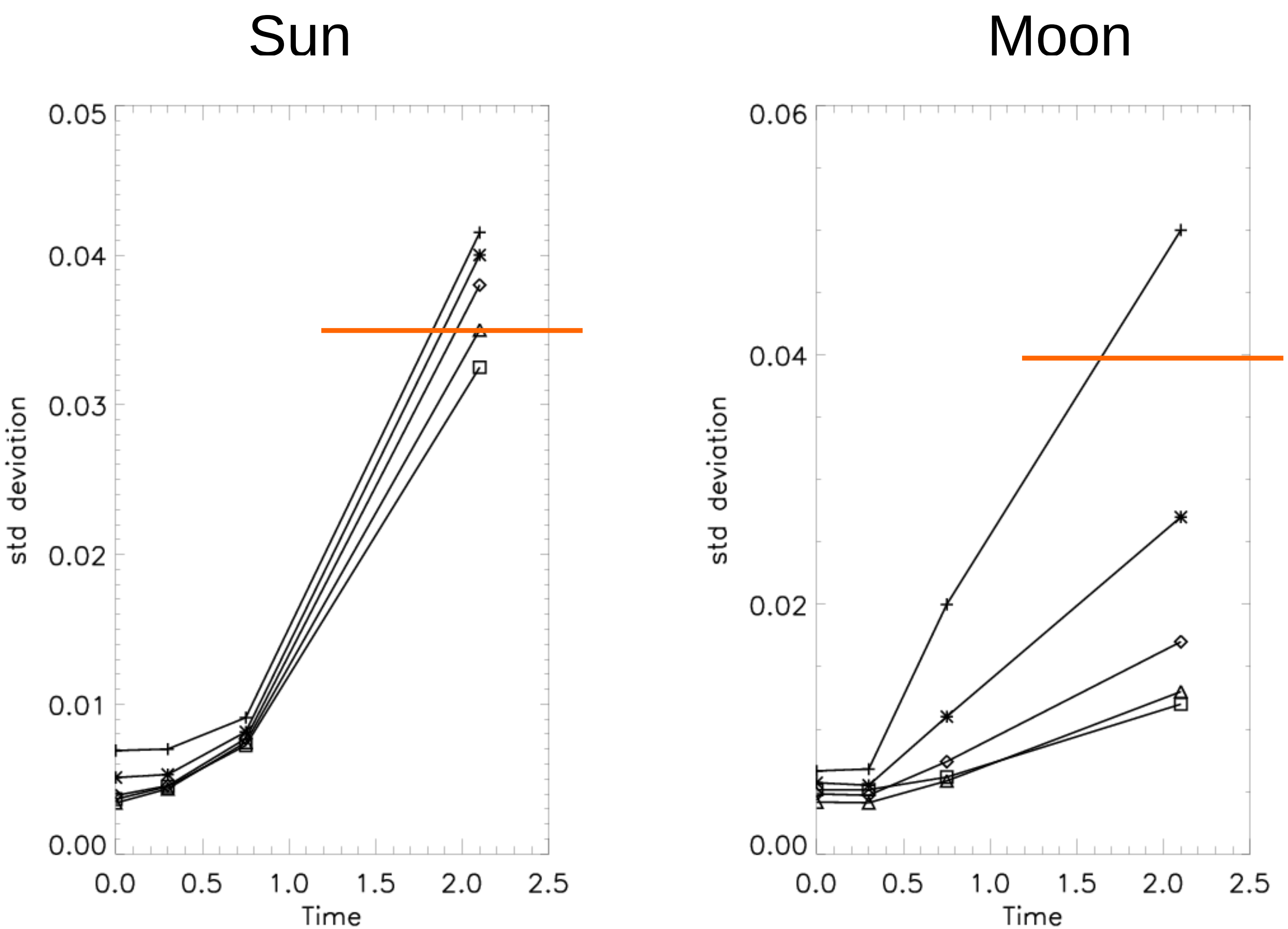
Repres. error

Effect lasts into forecast

Forecast error after 7 cycle lengths (t=2.1) for best choice of regularisation is similar to model error growth for sun and only 1/3 of that for moon.

Position error growth against time-experiment II

Model error



Effect depends on obs quality

Double the observation error.

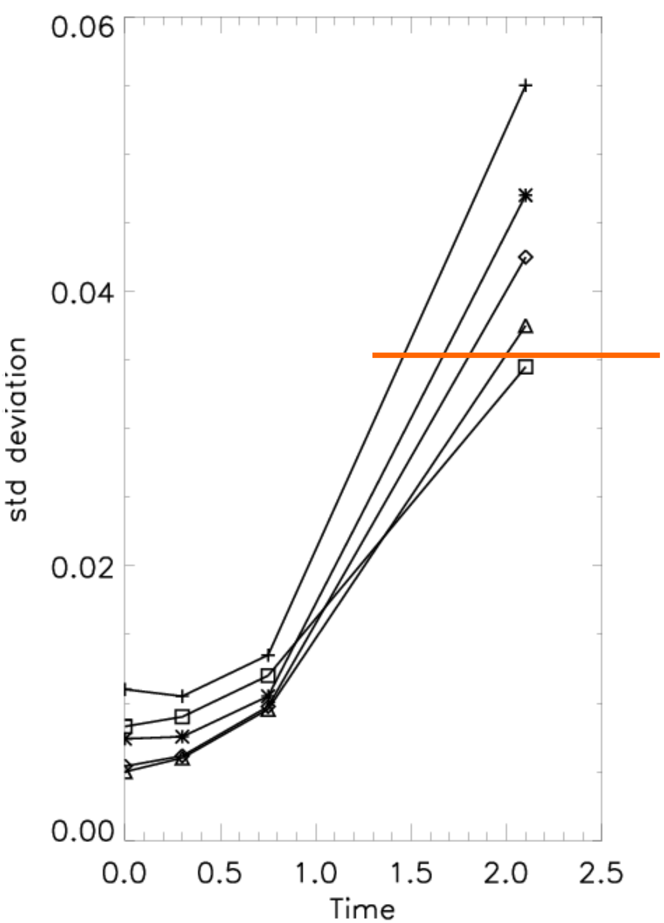
Optimum forecast error not much greater for sun, but doubled for moon.

Compensation effect not clear cut for sun, clearly present for moon.

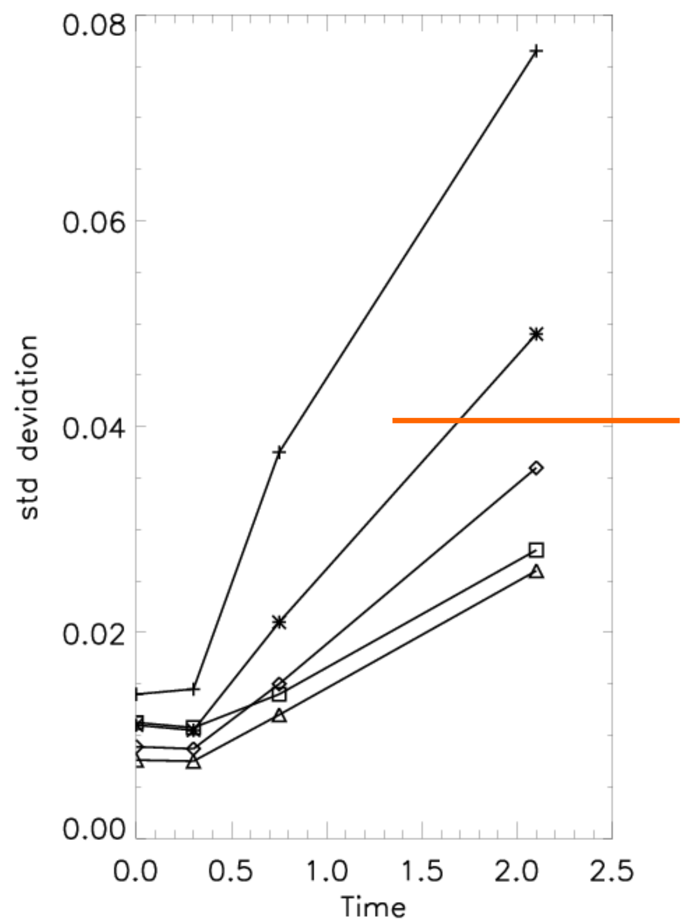
Doubled obs errors

Sun

Model error



Moon



Obs spread through cycle

Error for sun 15% larger than standard expt.

Error for moon 15% larger than standard exp but much less than model error accumulation.

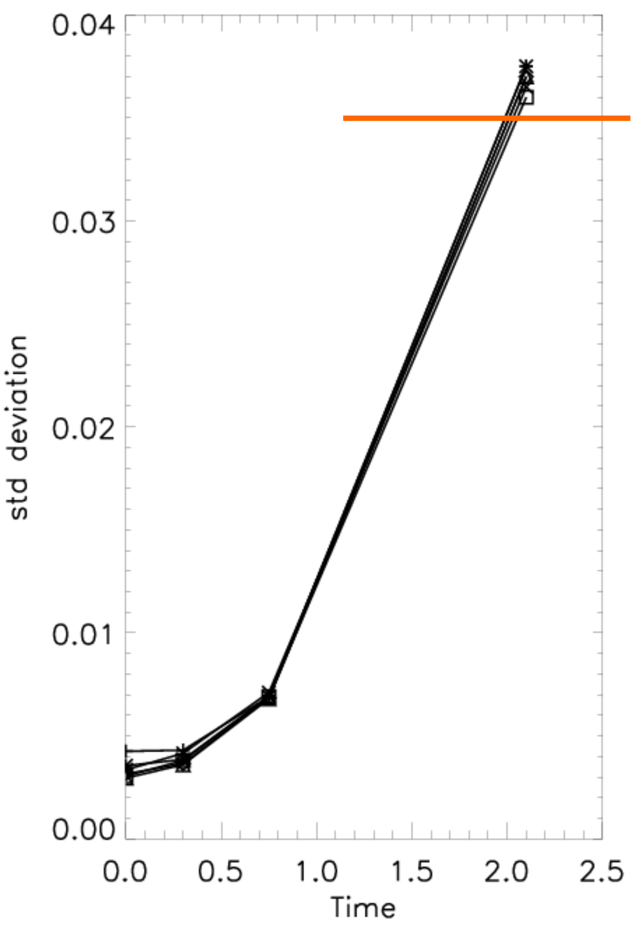
Compensation less effective but still present despite more realistic spread of observations.

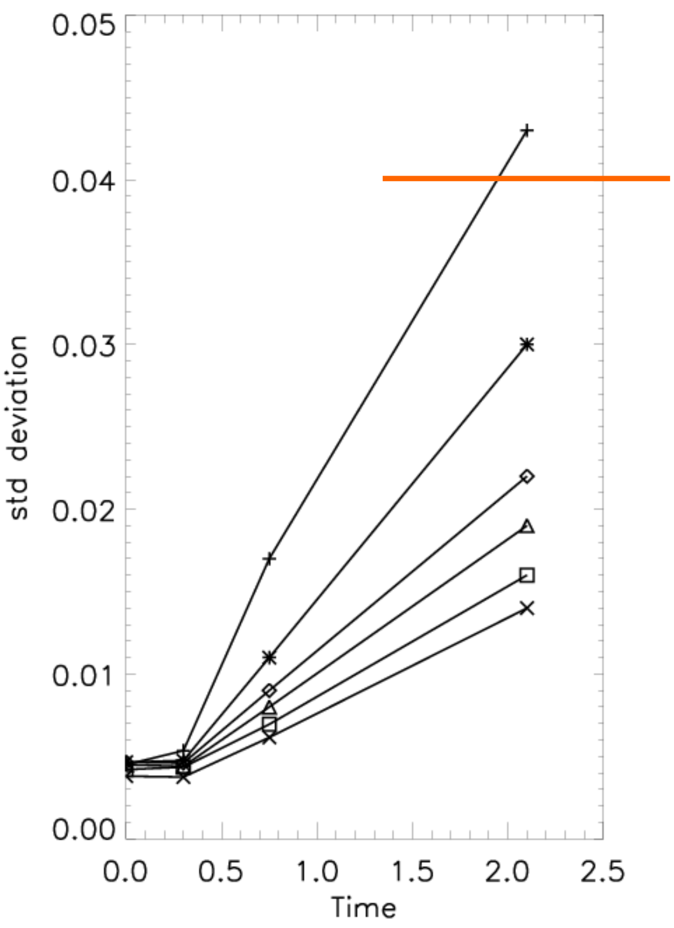
Increased obs-4 sets/cycle

Sun

Moon

Model error





Comments

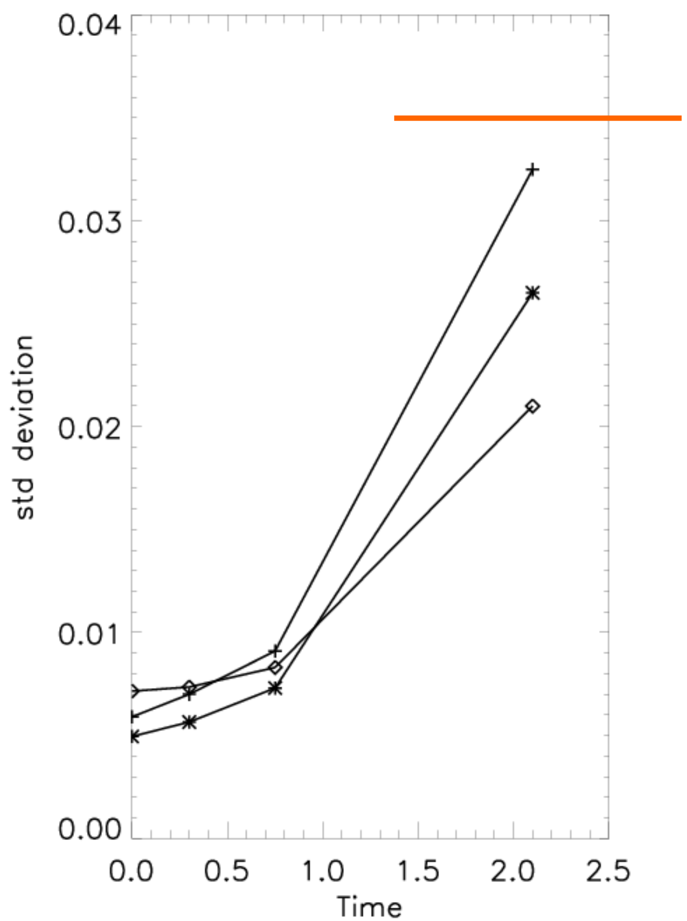
Error for sun only 2/3 of value in standard expt., but analysis error twice as large.

Error for moon nearly twice as large as in standard expt., analysis error 50% greater.

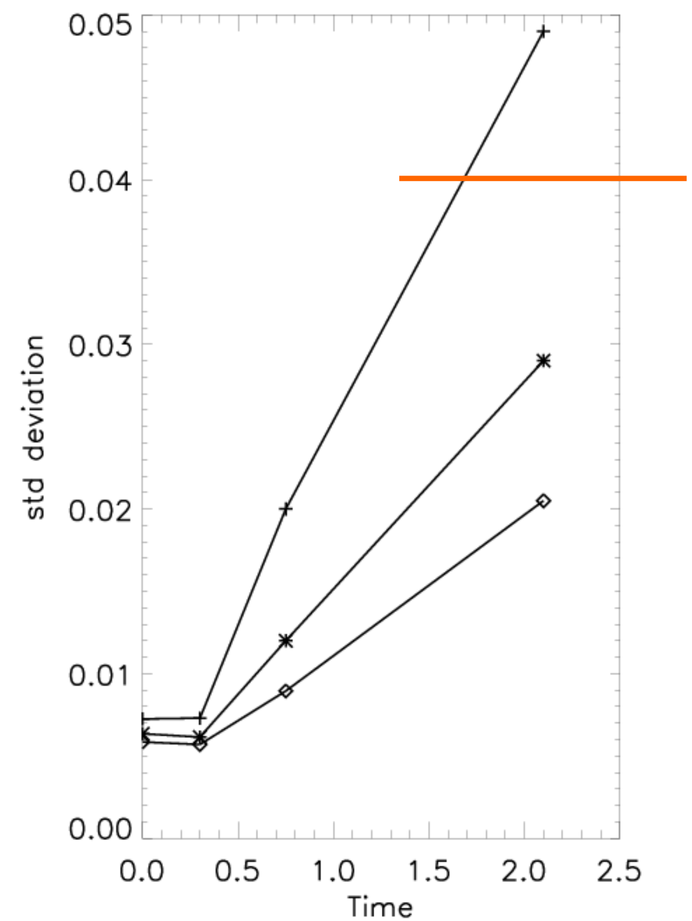
Consistent with control of performance of moon by obs.

Momentum only obs-standard **R**

Sun



Moon



Model error

Increased model error

Error for sun 5x larger, thus controlled by model error.

Analysis error similar to model error accumulation and greater than background.

Forecast error for moon only increased by 50%.

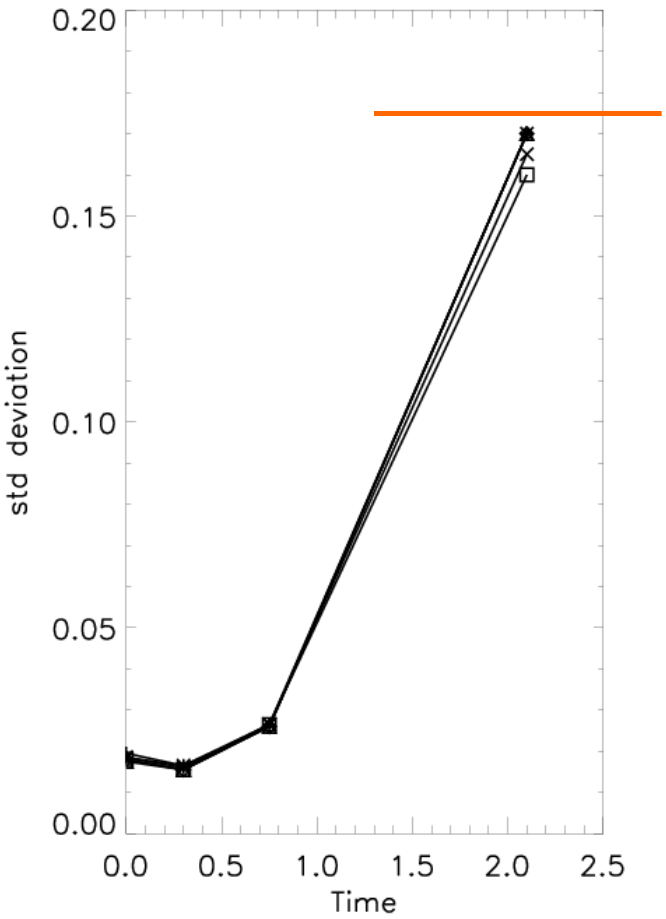
Analysis error much less than implied by model error growth.

Still mainly controlled by obs. Compensation can deal with it.

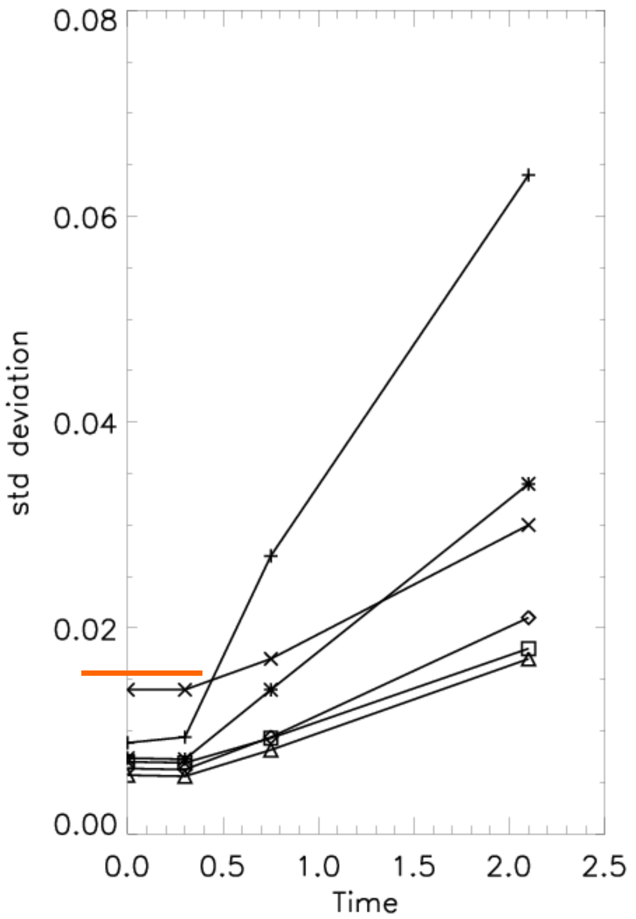
Increased model error (x5)

Model error

Sun



Moon



Do we need a long window?

Control of forecast error

Forecast error will not grow if

$$\frac{d}{dt}\left(\mathbf{Sz}\right) = \mathbf{S}\,\frac{d\mathbf{z}}{dt}$$

Information on d**z**/d*t* is only available from time sequences of similar observations. Since this information needed for timescales>12hrs, appropriate for balanced motion, a window length >12 hrs is needed, so weak constraint 4D-Var has to be used.

Results shown above, however, indicate that this effect can be achieved with a short window strong constraint 4D-Var.

Basic mechanism

Background state is controlled by obs used in previous cycle.

If only minimal increments made to fit next set of observations, then the trajectory over multiple cycles will fit observations over long periods with small corrections and

$$\frac{d}{dt}\left(\mathbf{Sz}\right) \approx \mathbf{S}\frac{d\mathbf{z}}{dt}$$

If large increments are made this will no longer be true

Also depends on **S** not being strongly time dependent

Toy calculation

Assume model equation is

$$\frac{d}{dt}\left(x,\dot{x}\right)=\left(\dot{x},-b^2\left(x+a\right)\right)$$

and truth is given by

$$\frac{d}{dt}\left(x,\dot{x}\right)=(\dot{x},-x)$$

Model error is given by model-truth at $\left(x,\dot{x}\right)$

Solutions

True solution if x=0,dx/dt=1 at t=0 is

$$x = \sin(t)$$

Model solution if x=0,dx/dt=1 at t=0 is given by

$$x + a = \sqrt{(a^2 + b^{-2})}\,\sin\left(bt - \sin^{-1}\left(\frac{a}{\sqrt{(a^2 + b^{-2})}}\right)\right)$$

but model solution if x=-a,dx/dt=b at t=0 is

$$x + a = \sin(bt)$$

Initial phases then match

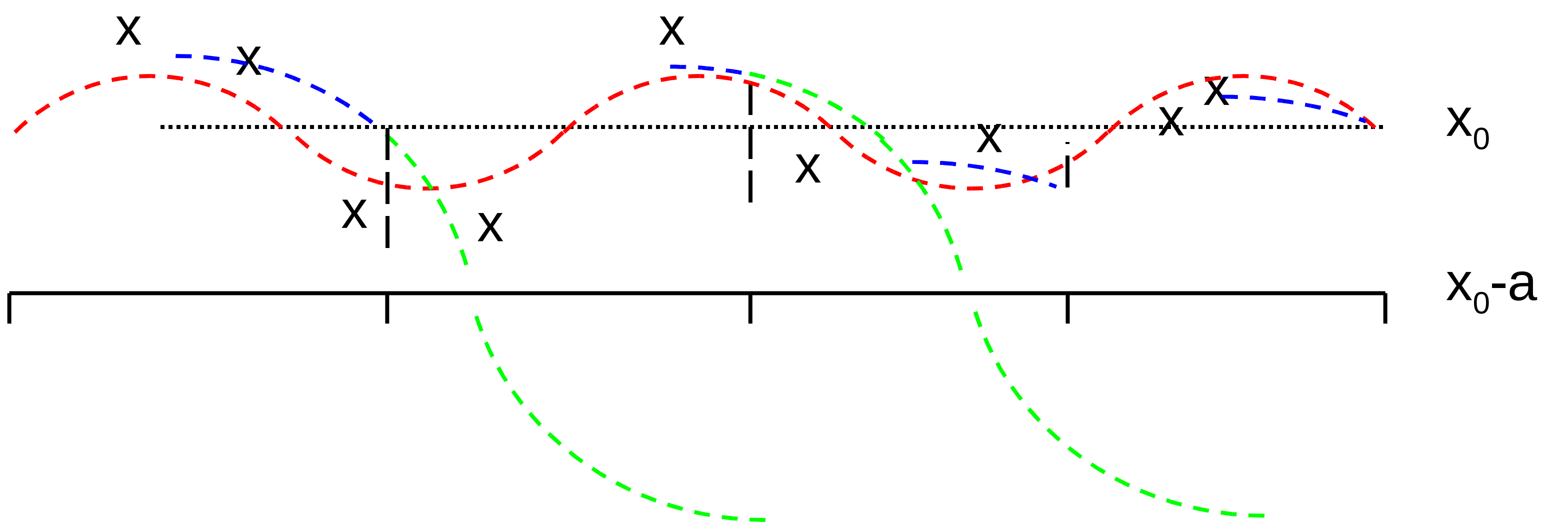
Solution using long window

If cost minimised over whole long window with subwindows, and same number of observations each subwindow, then expect similar final cost for each subwindow.

Case with large Q

Observations fitted closely in each subwindow

Large forecast errors possible



Does a long window help

For large Q the solution is the same if minimised over each subwindow separately, (forgetting factor).

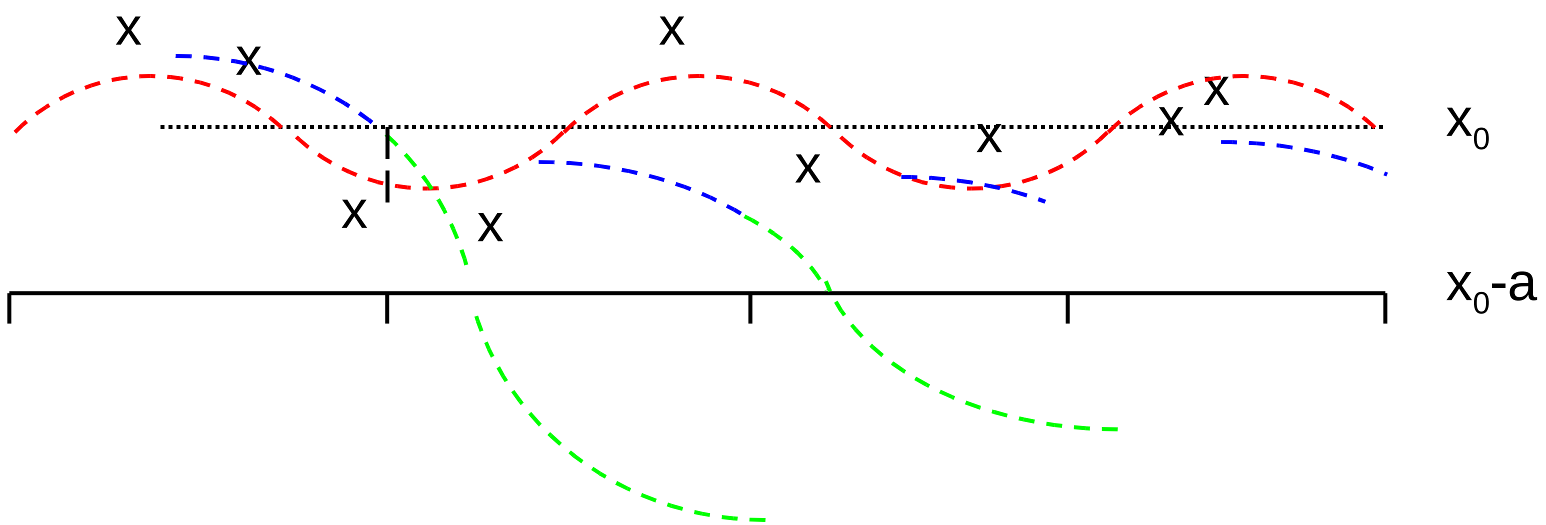
For Q=0 observations over all subwindows determine the result.

Now consider sequential minimisations with small Q.

Sequential with small Q

Minimise changes in trajectory and gradient at each boundary

Solution forced closer to x=-a by restriction on changes in gradient. Forecast errors thus smaller.



Summary

Summary

Three aspects of controlling forecast error growth in the presence of model error.

These issues matter if the model error is not purely random and in particular is correlated in time. Otherwise standard theory is fine.

Experience is that model error has a substantial non-random component which is correlated in time.

All relate to idea of analysing on model attractor. If the model error is purely random then the model attractor will be the same as the true attractor.

Issue I

Use of a smoother rather than a filter allows model error to be compensated by analysis error over a time period.

If model errors correlated in time this can continue into the forecast.

Forecast errors can then be smaller than if started from the true state.

Issue II

Non-random error associated with errors of representativity.

Forecasts started from true state will be worse.

Necessary to prevent observations driving analysis too far towards the truth, need to realise observation ‘errors’ from different platforms are correlated.

Issue III

Forecast error growth also controlled if time derivative of the model state can be matched to the true time derivative inferred from observations.

Requires the errors of representativity to vary slowly in time.

Then use long window, but don't allow large increments between subwindows.

If short windows are used, but only small increments allowed, then can achieve effect of a long window.

Questions and answers