4D-Var in the presence of model error

Mike Cullen

24 October 2011

Acknowledgements

The 3 body model code was written by Gordon Inverarity.

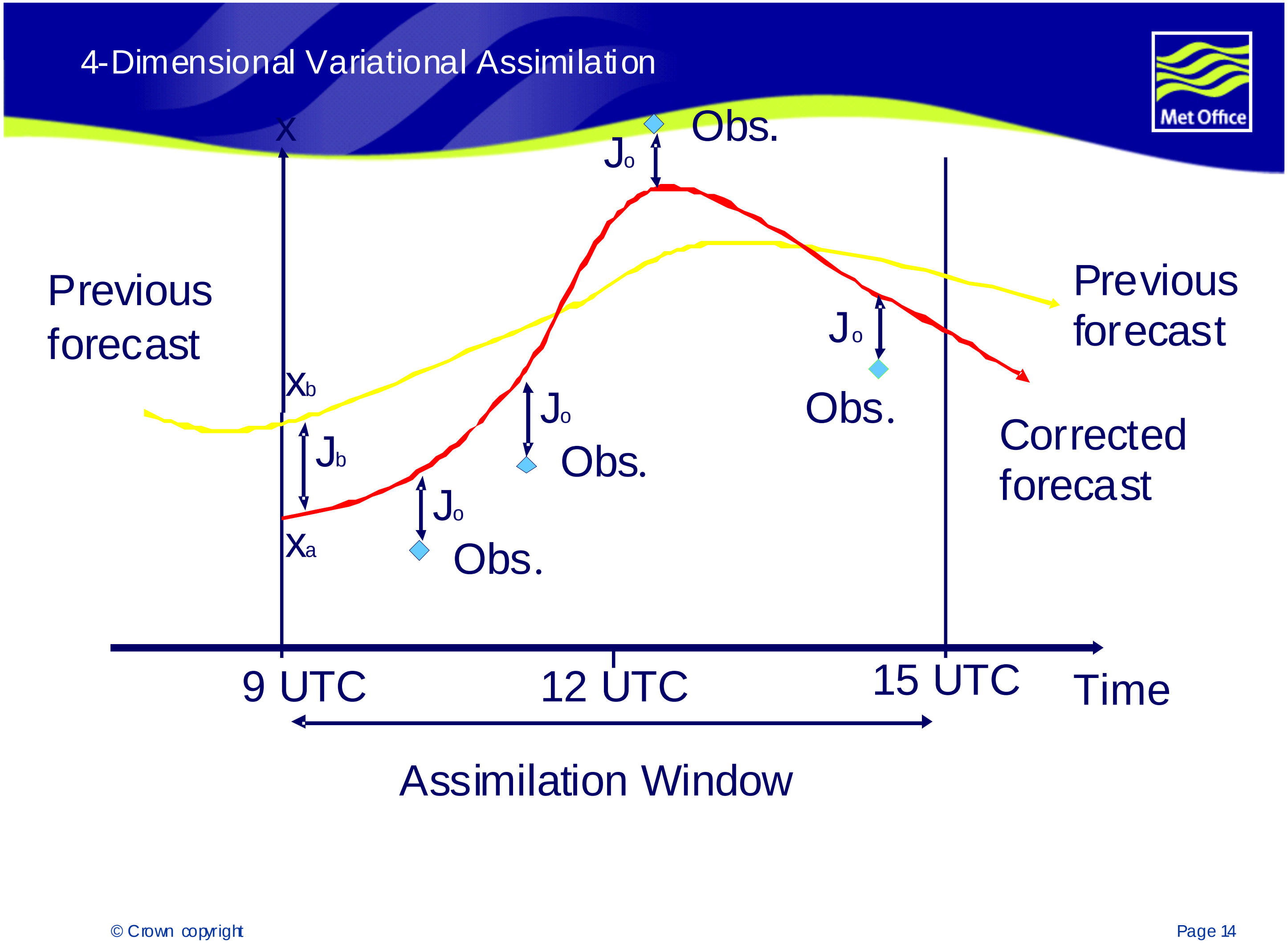
Contents

This presentation covers the following areas

- Background
- Theory
- Toy model results

Background

4dVar in meteorology



Smoothing and filtering

Let $\mathbf{x}$ be a state vector at time t and $\{y(t_i)\}$ be a collection of observations at times t_i.

Then if we evaluate $p(\mathbf{x}|\{y(t_i)\})$ this is a filter if $t_i<t$ for all i and a smoother if $t_i>t$ for some i.

If $t_i>t$ for some i but $t_i<t+\delta$ for all i, then we have a fixed lag smoother with lag δ.

Operational 4dVar schemes in meteorology are fixed lag smoothers with lags typically 6-12 hours.

Effect of smoothing

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

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

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 $\mathbf{P}$.

Errors of representation

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 $\mathbf{x}=\mathbf{S}\mathbf{z}$, 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 $\mathbf{MSz}-\mathbf{SNz}$.

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. Instead could treat by reducing background error.

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.

Requires using strong background constraint and long window (>12 hrs) to bring in observations from earlier times.

Evolution of error over window

Therefore assume that error growth during 4dVar window is governed by an evolution operator which includes the combined effect of model error and model evolution of error. Thus it will depends on the observation distribution and the assumed background error. The latter will not be equal to **P** as this is not evolved explicitly by 4dVar.

Previous arguments suggest that this growth, and hence **P**, may be reduced if the background constraint is strengthened.

Key results in this talk

1. The effect of a long window can be achieved by cycling with a short window.

2. The best choice ‘background’ term used in 4dVar is the true background error **P** compared at the observation time, not the start of the window.

3. There can be a strong feedback between the choice of background error and the true **P**, making a priori estimates of background error impossible.

Illustrate with theory and toy model results.

Theory

Short and Long window 4dVar

Met Office

Show that repeated cycles of 4dVar converge to the solution of long-window weak constraint 4dVar; assuming uniform distribution of observations between subwindows.

First analyse cycled short window system (i.e. each subwindow analysed successively rather than simultaneously), assuming the true time evolving error covariance is **P** and a model error covariance **Q** is applied to determine the increment at the beginning of each subwindow. **A** is the analysis error at the start of each subwindow after 4dVar has been carried out.

Analysis error on subwindow

The gain matrix on each subwindow, where **R** is the observation error covariance matrix, **H** is the observation operator including time selection, and **M** the tangent linear model evolution, is

$$\mathbf{K}=(\mathbf{Q}^{-1}+\mathbf{M}^{T}\mathbf{H}^{T}\mathbf{R}^{-1}\mathbf{H}\mathbf{M})^{-1}\mathbf{M}^{T}\mathbf{H}^{T}\mathbf{R}^{-1}$$

and the analysis error is

$$\mathbf{A}=\mathbf{P}+\mathbf{K}(\mathbf{R}+\mathbf{H}\mathbf{M}\mathbf{P}\mathbf{M}^{T}\mathbf{H}^{T})\mathbf{K}^{T}-\mathbf{K}\mathbf{H}\mathbf{M}\mathbf{P}-\mathbf{P}\mathbf{M}^{T}\mathbf{H}^{T}\mathbf{K}^{T}$$

Analysis of scalar case

Use lower case letters to represent variances of scalars, letter corresponds to that of the matrix in the general case. To maximise the smoothing effect assume obs are at the end of the window, assume obs operator is the identity and m represents the error growth under the action of the model up to the observation time.

Assume growth of error during subwindow, prior to fitting the obs, is by a factor λ and an additive term μ. This may be less than the model evolution because of continued cancellation of evolution error and model error if the model error is correlated in time.

Evolution of *p*

The evolution of *p* from the start of one subwindow to the next is given by

$$p_{n+1} = \frac{m^2 q}{r + m^2 q}\left(\frac{m^2 q}{r + m^2 q}\left(r + \lambda^2 p_n + \mu\right) - 2\left(\lambda^2 p_n + \mu\right)\right)$$

This calculation uses the fact at 4dVar with obs at the end of the window is equivalent to 3dVar with a background covariance matrix evolved to the end of the window.

Evolution equation

This can be considered as a discretisation with unit timestep of the equation

$$\frac{\partial p}{\partial t} = \alpha p + \beta$$

where

$$\alpha = \lambda^2 \left(\frac{r^2}{\left(r + m^2 q\right)^2} \right) - 1$$

$$\beta = \frac{\mu r^2}{\left(r + m^2 q\right)^2} + r \left(\frac{m^2 q}{r + m^2 q} \right)^2$$

Solution

The solution is

$$p = \alpha^{-1}\left(\left(\alpha p_0 + \beta\right)\mathrm{e}^{\alpha t} - \beta\right)$$

if α>0, p → ∞ as t → ∞

if α<0, p → $-\alpha^{-1}\beta$ as t → ∞. This requires

$$\lambda^2 < \left(\frac{r + m^2 q}{r}\right)^2$$

Long windows

If the problem is simultaneously solved over a large number of subwindows, and each has the same distribution of observations and the same growth rates, then the solution will be $p=-\alpha^{-1}\beta$.

Thus a sequential short window algorithm will converge to the long window solution on a timescale $|\alpha|^{-1}$. Convergence most rapid for large m and q.

Optimum choice of q

Choose q to minimise p at steady state.

$$p = -\alpha^{-1}\beta = \frac{\mu r^2 + m^4 q^2 r}{\left(r + m^2 q\right)^2 - \lambda^2 r^2}$$

$$q \to \infty \Rightarrow p \to r$$

$$q \to 0, \lambda < 1 \Rightarrow p \to \frac{\mu}{1-\lambda^2}$$

Case μ=0

$$p = -\alpha^{-1}\beta = \frac{m^4 q^2 r}{(r + m^2 q)^2 - \lambda^2 r^2}$$

$$q \to \infty \Rightarrow p \to r$$

$$q \to 0, \lambda < 1 \Rightarrow p \to 0$$

Corresponds to assimilation in the unstable subspace in perfect model case. Also applies with model error if its' effect is purely to modify growth rates.

Calculation of optimum *q*

Ignore dependence of λ and μ on q.

$$\frac{\partial p}{\partial q} = \frac{2qr\lambda^2 m^2 + 2\mu m^2\left(r + m^2 q\right)}{\left(r + m^2 q\right) - r^2\lambda^2} - \frac{2m^2\left(r + m^2 q\right)\left(q^2 r\lambda^2 m^2 + \mu\left(r + m^2 q\right)^2\right)}{\left(\left(r + m^2 q\right) - r^2\lambda^2\right)^2}$$

$$\frac{\partial p}{\partial q} = 0 \Rightarrow$$

$$\left(q - \mu\right)\left(r + m^2 q\right) = q\lambda^2 r$$

Analysis of solution

Solution for *q* is

$$2m^2q = \mu m^2 + \left(\lambda^2 - 1\right)r \pm \sqrt{\left(\mu m^2 + \left(\lambda^2 - 1\right)r\right)^2 + 4\mu m^2 r}$$

Only positive root relevant. For μ=0

$$m^2q = \left(\lambda^2 - 1\right)r$$

in which case we can show

$$\lambda^2 p = m^2 q$$

which is consistent with standard theory.

Further comments on solution

p and q are both functions of the assimilation process and cannot be determined in advance. As stated they depend on λ,μ and r, but the smoothing process will also make λ,μ functions of q.

In case μ=0 will only have $p=q$ at start of window if m=λ, i.e. true error during window grows at same rate as model.

In smoothing scenario likely that $\lambda<m$ for $m>1$. Though $p=q$ at the end of the window, will have $q<p$ at the beginning.

Additional comments

Can show that $\partial p/\partial q$ is increased if $\partial\lambda/\partial q>0$ and so optimum q decreases. In case $\mu=0$, can show $\partial\lambda/\partial q>0$ implies $\partial p/\partial q>0$ for all q.

Reduced q will bring observations from earlier times into play, and thus control errors in time derivative, so expect $\partial\lambda/\partial q>0$ for q large enough to prevent filter divergence.

Potentially allows p and q to collapse to very small values.

Predictions to test

1. Short window converges to long window. This implies that a short window analysis can make use of previous observations. (key result 1)

2. Smoothing. The analysis error does not grow as fast during the window as the model evolution of the initial error.

3. Optimum q is smaller than p for large model growth m, but evolved q similar to evolved p at the end of the window. (key result 2)

4. Overall error only limited by r. Optimum q a function of r and error growth properties which depend on q as well as model (key result 3).

Toy model results

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, model error covariance **Q** estimated from statistics of (analysis-background) over set of 300 cycles.

Repeat recalculation and rerun cycles till assimilation fails because of ill-conditioned **Q** 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

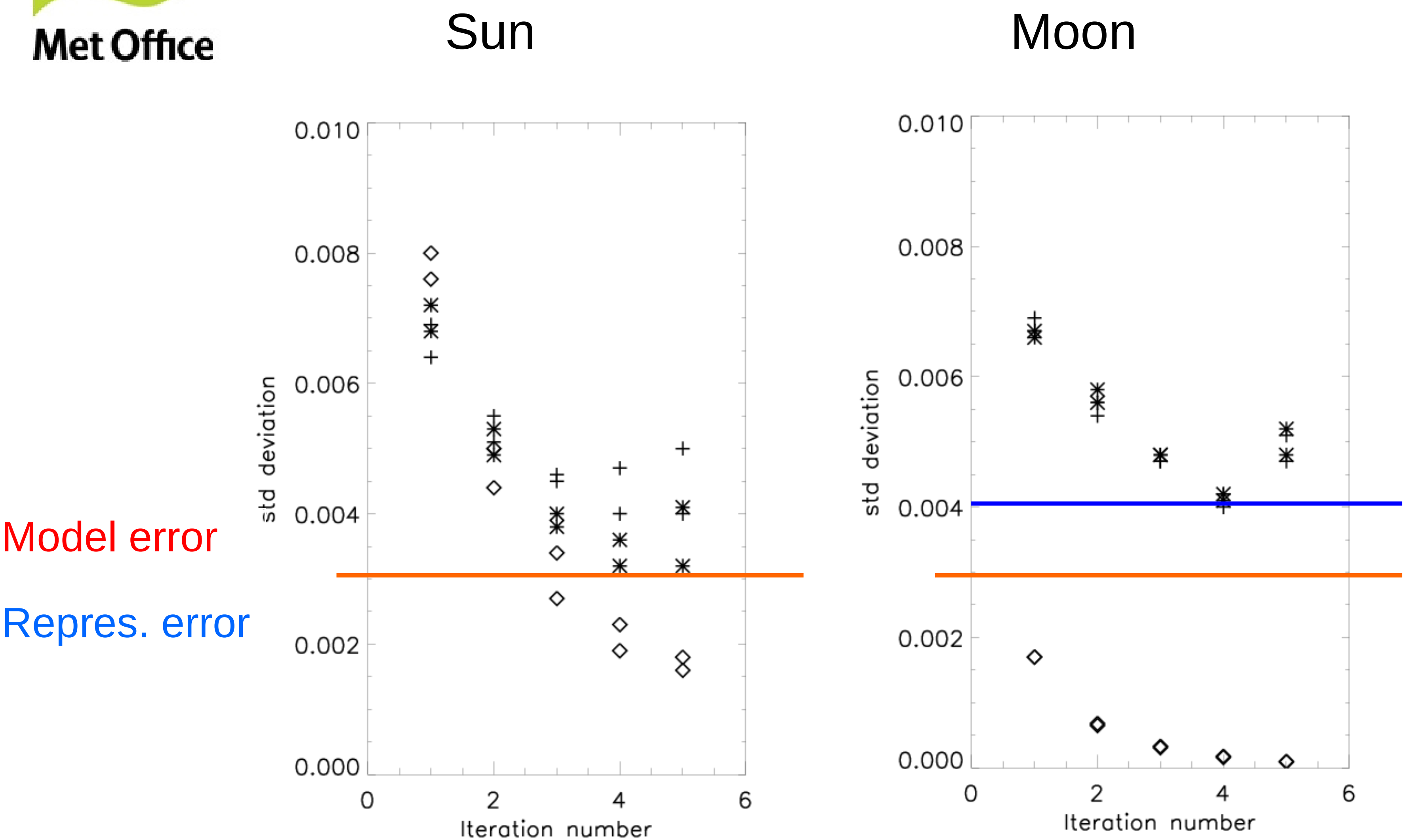
Minimum error given by error of representation.

Comments for 'best' cycles.

Fast mode; errors close to minimum, analysis error almost equal to background, analysis increment very small. Both p and q have collapsed.

Slow mode: analysis error less than background and analysis increment smaller than either.

+=(B-T),*=(A-T),◊=(A-B)



Sun
Moon
Model error
Repres. error
0.010
0.008
0.006
0.004
0.002
0.000
std deviation
0
2
4
6
Iteration number

Check evolved analysis error

Slow mode. Evolved analysis increment similar to model error growth. Evolved analysis error much larger, and much larger than background error for next cycle.

Fast mode: Evolved analysis increment 50% of model error growth. Evolved analysis error much larger , and much larger than background error for next cycle.

Both cases: Behaves as expected for smoother. Representation error should be excluded from analysis error.

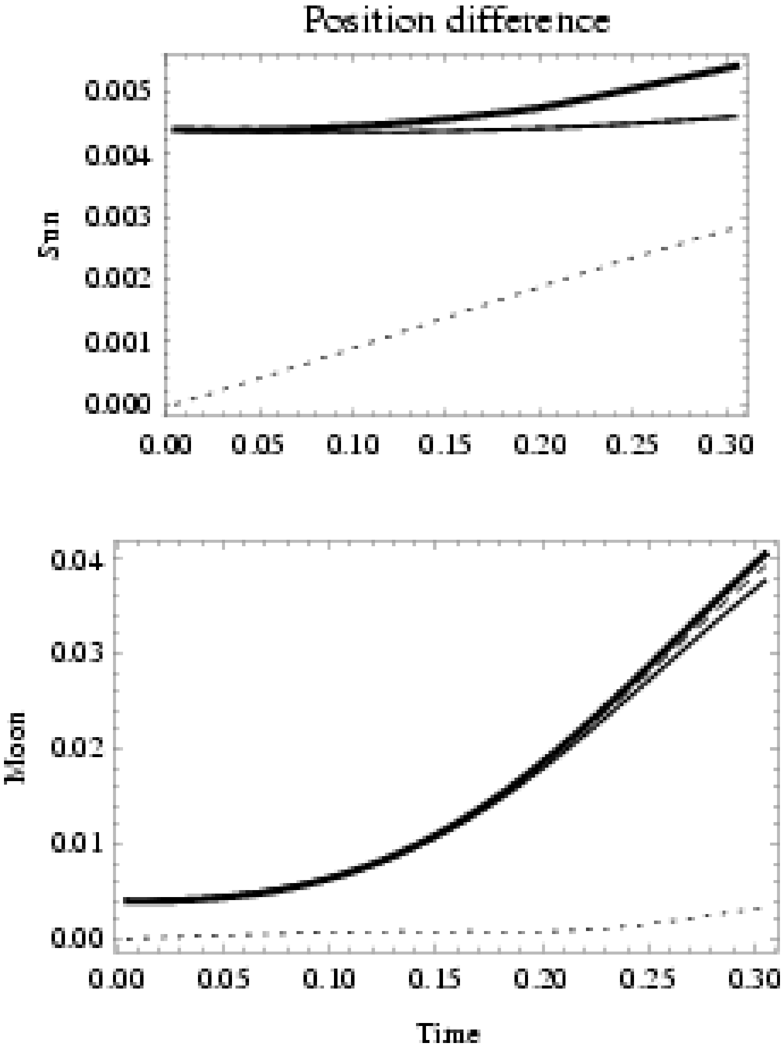
Evolution of errors

- - - - - - - - -
Evolution of analysis error (Best case)

———————
Evolution of analysis error (linear model)

....
Evolution of model error

———————
Total evolution



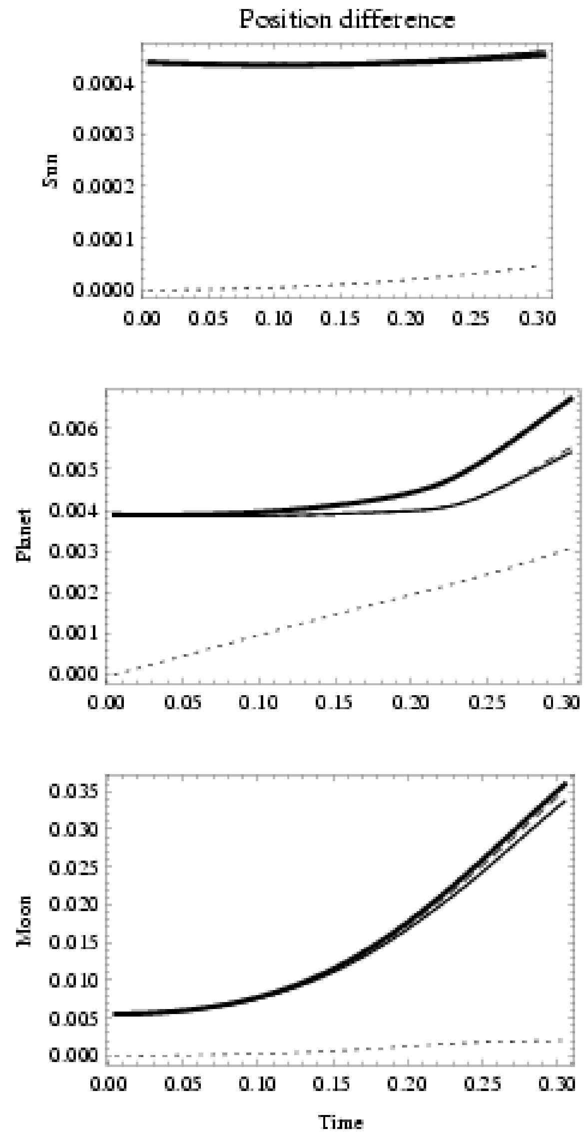
Comparison with obs error

Evolution of analysis error (Best case)

Evolution of analysis error (linear model)

Evolution of model error

Total evolution



Obs error

0.0008

0.02

0.001

Comments

Obs related to slow mode fitted well within obs error.

Obs related to fast mode not all so fitted. Obs error for moon is 25% of representation error, analysis error limited by the latter. Better performance for small *q* may also be related to this.

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.

Implies strong cancellation between model error and perturbation growth after the end of the window, so strong time correlation in model error (not surprising in this case-but also true in real systems).

Reduction of forecast error below that from truth state implies use of *d/dt* information from obs, which mean use of obs over more than one window.

Effect on choice of q

Actual growth of error between t=0.3 and 0.6 determines optimum *q*.

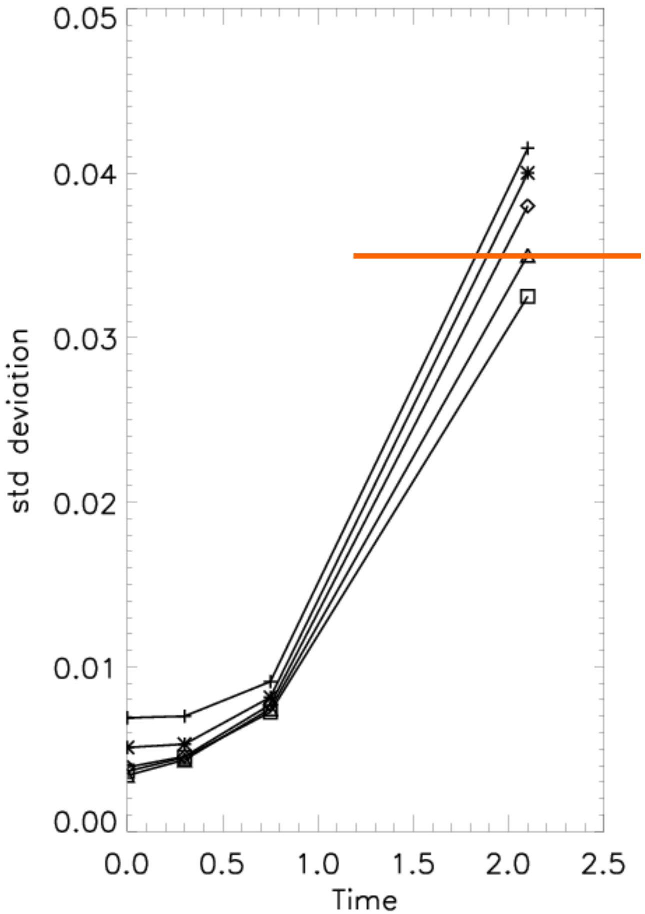
Even if we assume no growth between t=0 and 0.3, as shown, then evolved error under model at t=0.6 will be 0.005 for slow mode and 0.04 for fast mode, actual errors are 0.006 for slow mode and 0.005 for fast mode.

Hence optimum *q* for fast mode much less than *p*.

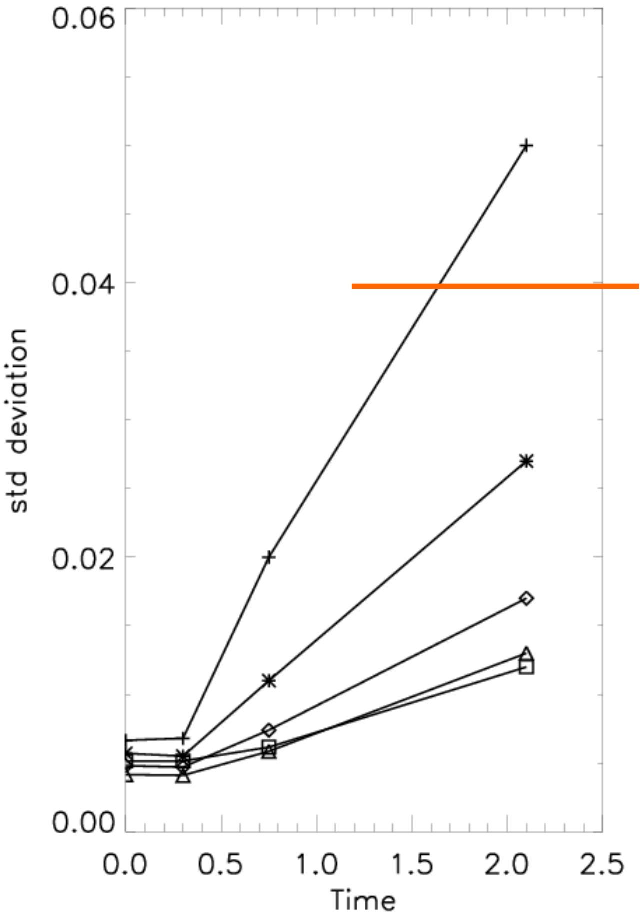
Position error growth against time-experiment II

Model error

Sun



Moon



Control of error by obs error

Double the observation error.

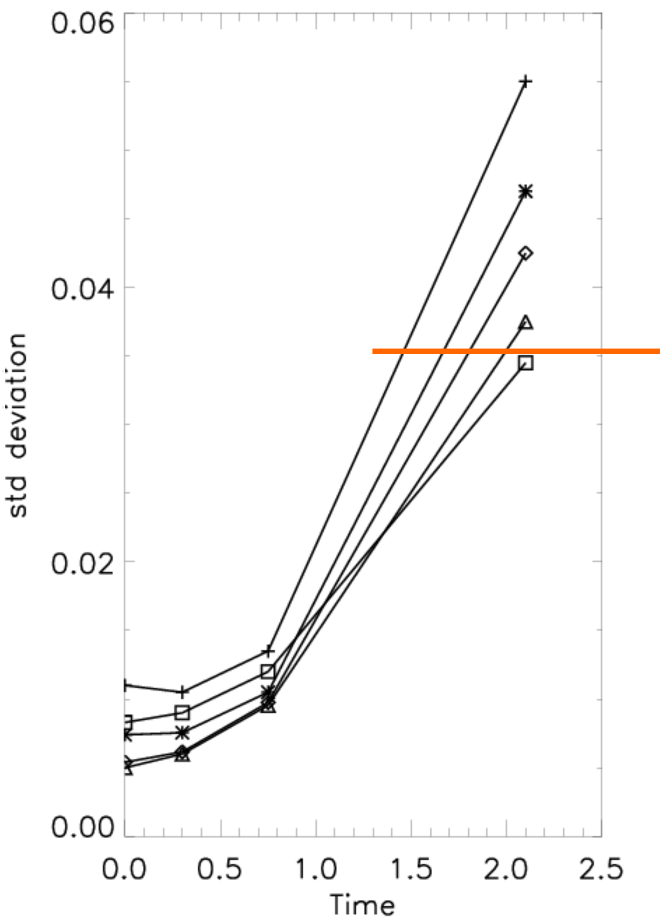
Optimum forecast error not much greater for sun, but doubled for moon.

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

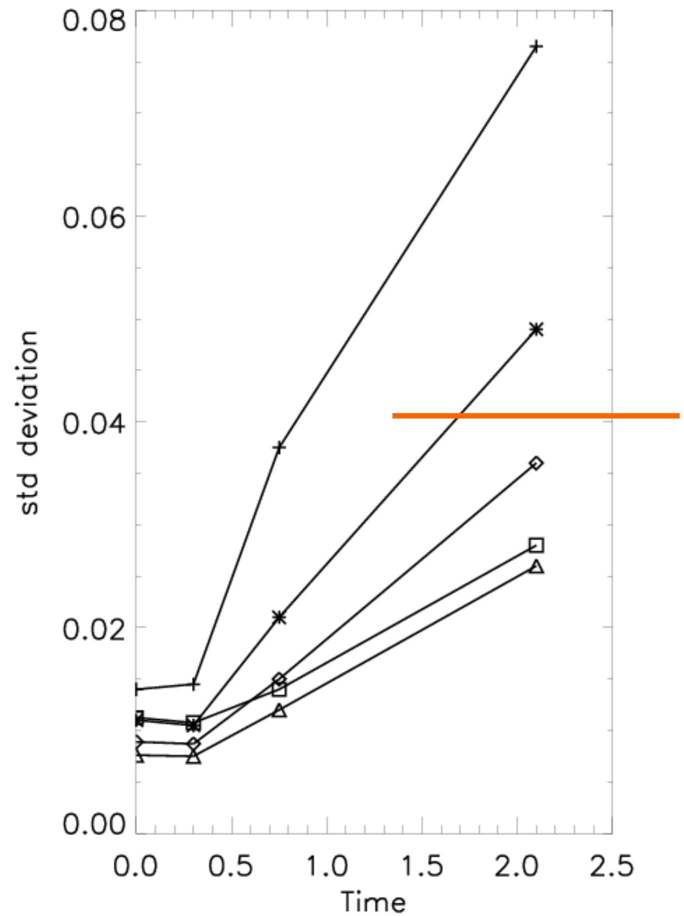
Doubled obs errors

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.

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

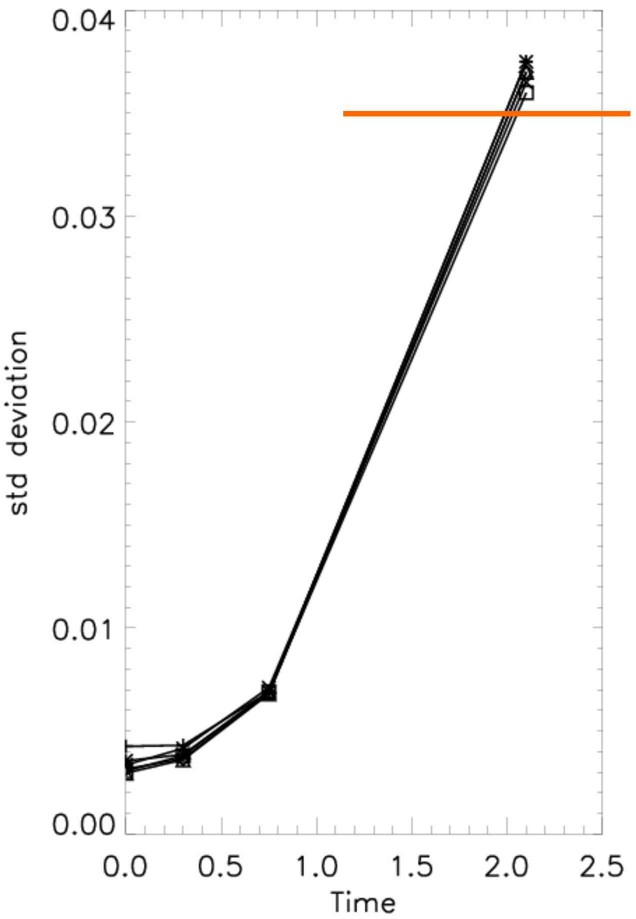
Analysis error held closer to minimum value for all choices of **Q** (more obs).

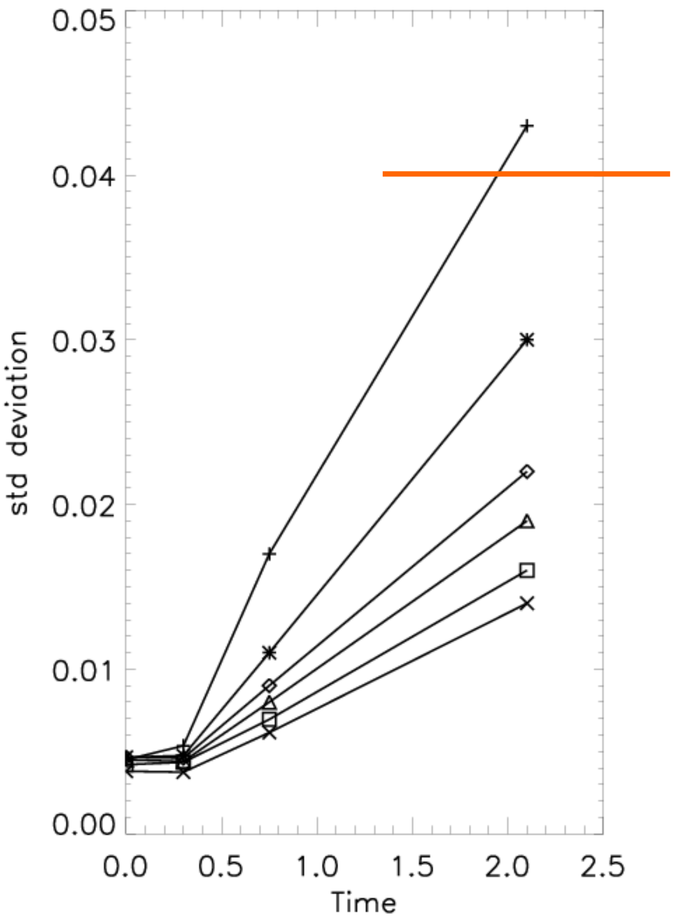
Increased obs-4 sets/cycle

Sun

Moon

Model error





Reduced obs

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

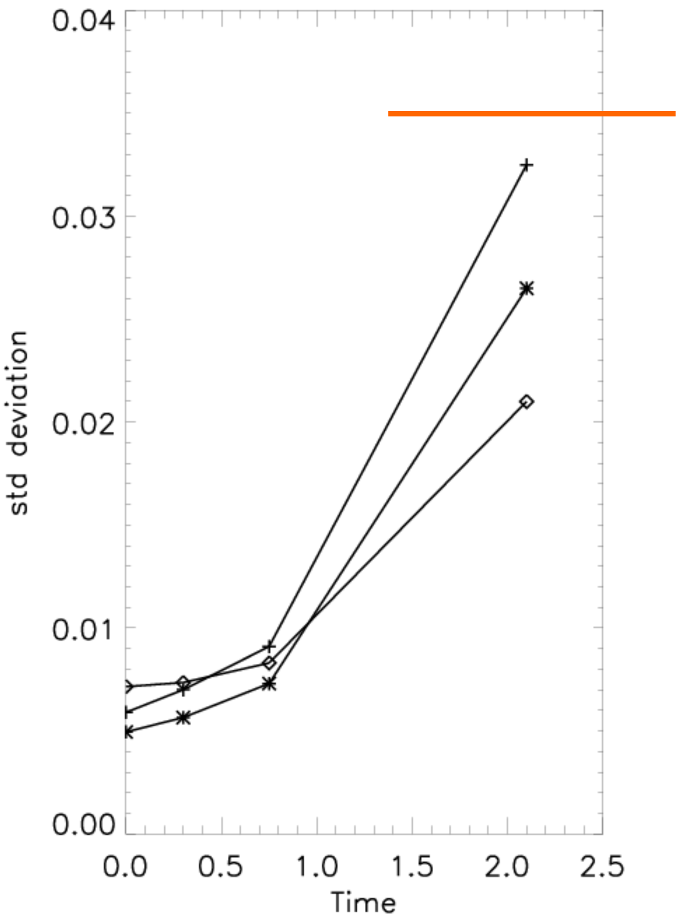
Error for moon nearly twice as large as in standard expt., analysis error 50% greater.

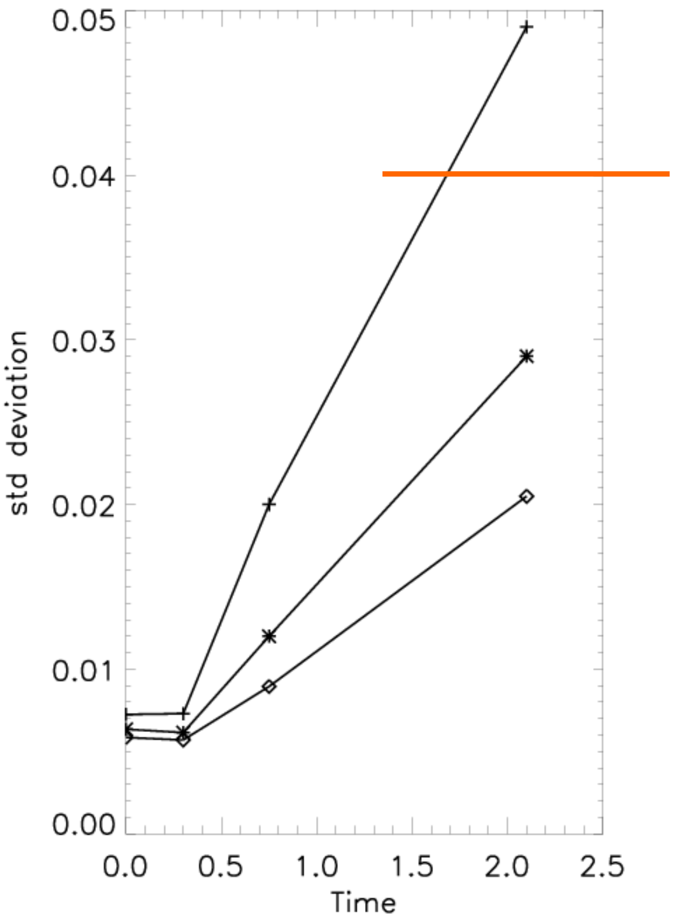
Consistent with control of performance of moon by obs.

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.

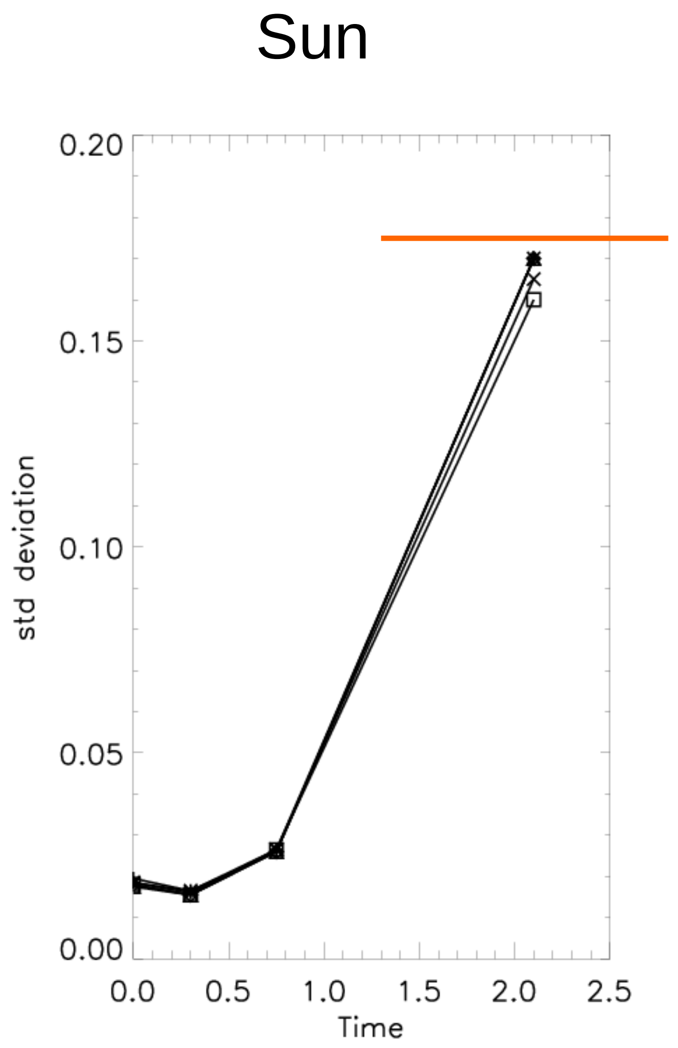
Forecast error for moon only increased by 50%.

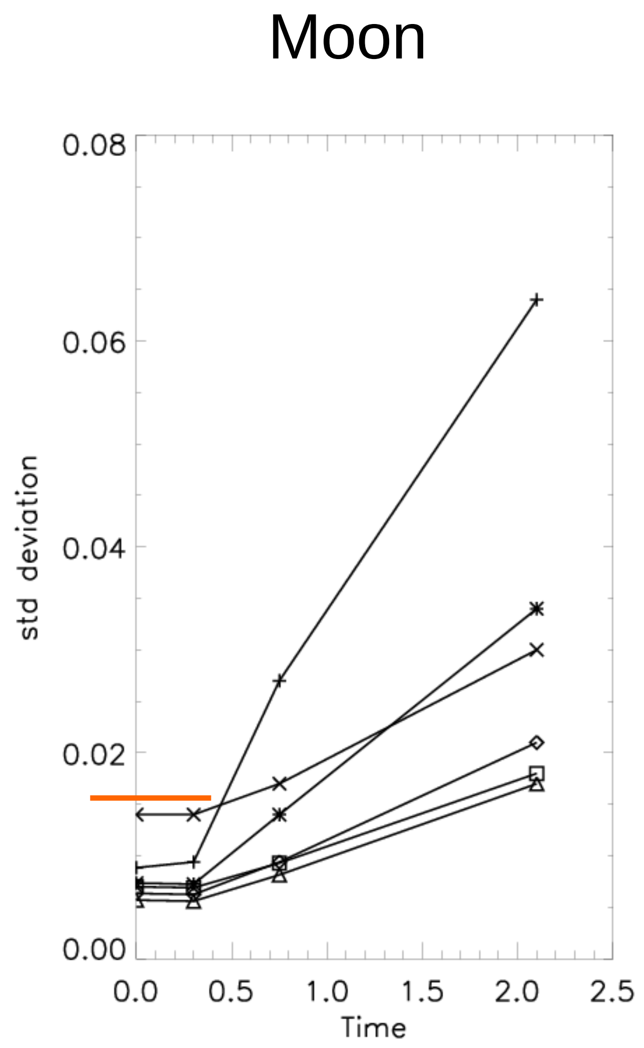
Analysis error much less than implied by model error growth.

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

Increased model error (x5)

Model error





Summary

Summary

Short window converges to long window. This implies that a short window analysis can make use of previous observations.

Smoothing. The analysis error does not grow not as fast during the window as the model evolution of the initial error.

Optimum q is smaller than p for large model growth m, but evolved q similar to evolved p at the end of the window.

Overall error only limited by r. Optimum q a function of r and error growth properties which depend on q as well as model.

Questions and answers