Optimizing Data Assimilation for Re-Analysis

Mike Fisher

Optimizing DA for Re-Analysis

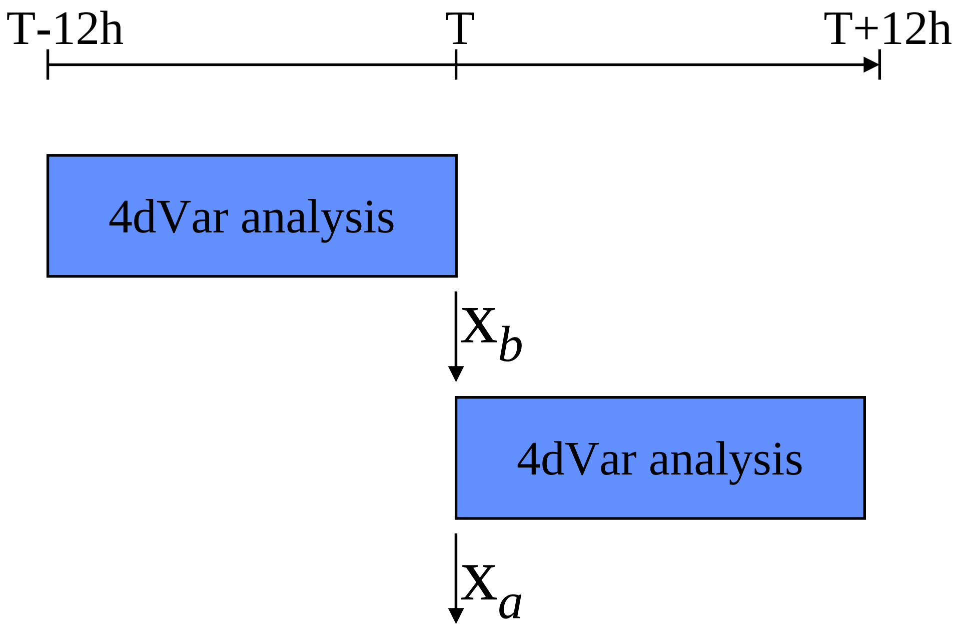
- Developing a data assimilation system is a major undertaking.
- For this reason, it is likely that most future re-analysis projects will continue to be based on NWP systems.
- But, NWP analysis systems are optimized for forecasting. They are not optimal for retrospective analysis.
  - (There is some advantage for a NWP centre in seeing how its current analysis system would have performed in the past. But, this does not outweigh the advantages of a high quality re-analysis.)
- Can we adapt NWP analysis systems to make them more optimal for re-analysis purposes?

Optimizing DA for Re-Analysis

- **The most obvious difference between analysis-for-NWP and retrospective analysis is that retrospective analysis can make use of future observations.**
- **I.e. re-analysis is a <u>smoothing</u> problem, whereas forecasting is a <u>filtering</u> problem.**
- **Can we recognize this distinction without having to develop radically different assimilation schemes for re-analysis and NWP?**

Optimizing DA for Re-Analysis

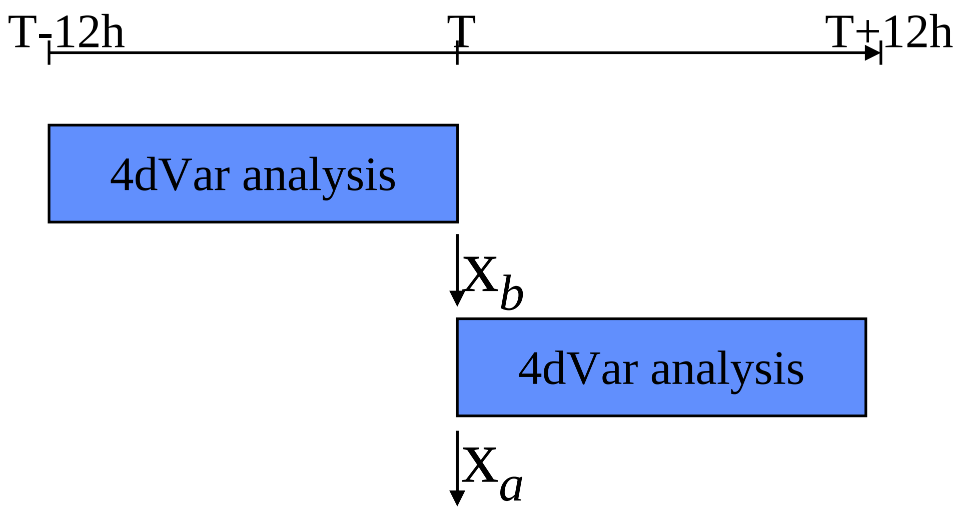
- **A simple optimization…**



This analysis knows about past observations, and observations 12h into the future

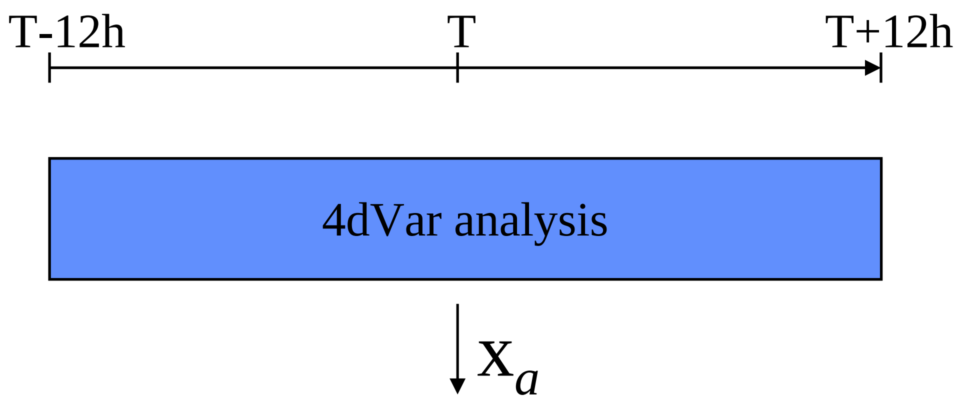
Optimizing DA for Re-Analysis

- **This simple optimization is far from optimal.**
- **Future observations are only taken into account up to 12h ahead.**
- **Information from the past is brought into the analysis via the prescribed background error statistics. These provide a crude approximation (B) to the true covariance matrix of background error (P^b).**

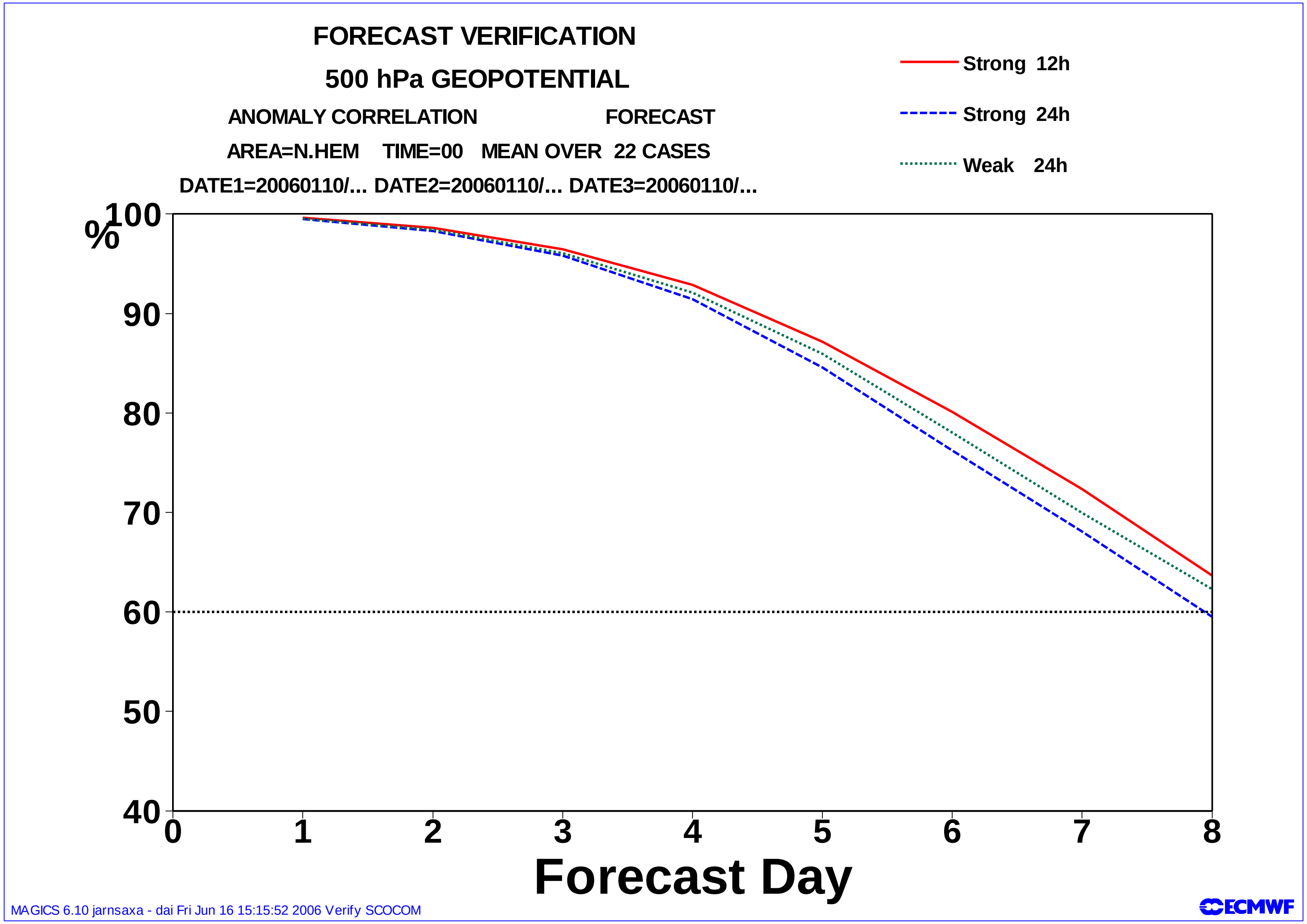


Optimizing DA for Re-Analysis

- **We can make the transfer of past information more optimal by making B more accurate.**
  - **For example, use some kind of approximate Kalman Filter.**
- **Alternatively, we can eliminate the need to specify B at the analysis time by analysing past and future observations simultaneously in a single, long analysis window.**

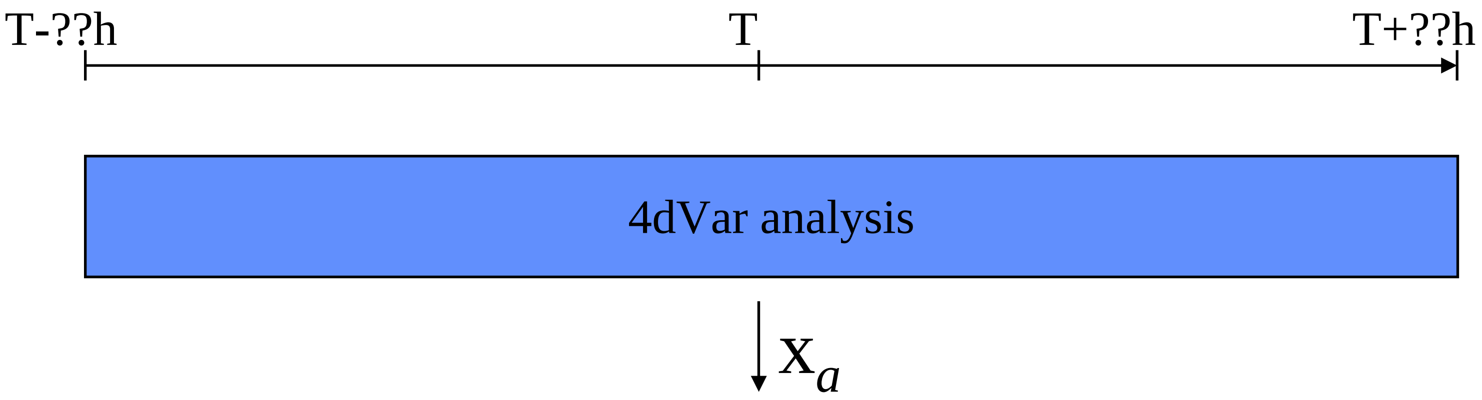


Optimizing DA for Re-Analysis



Optimizing DA for Re-Analysis

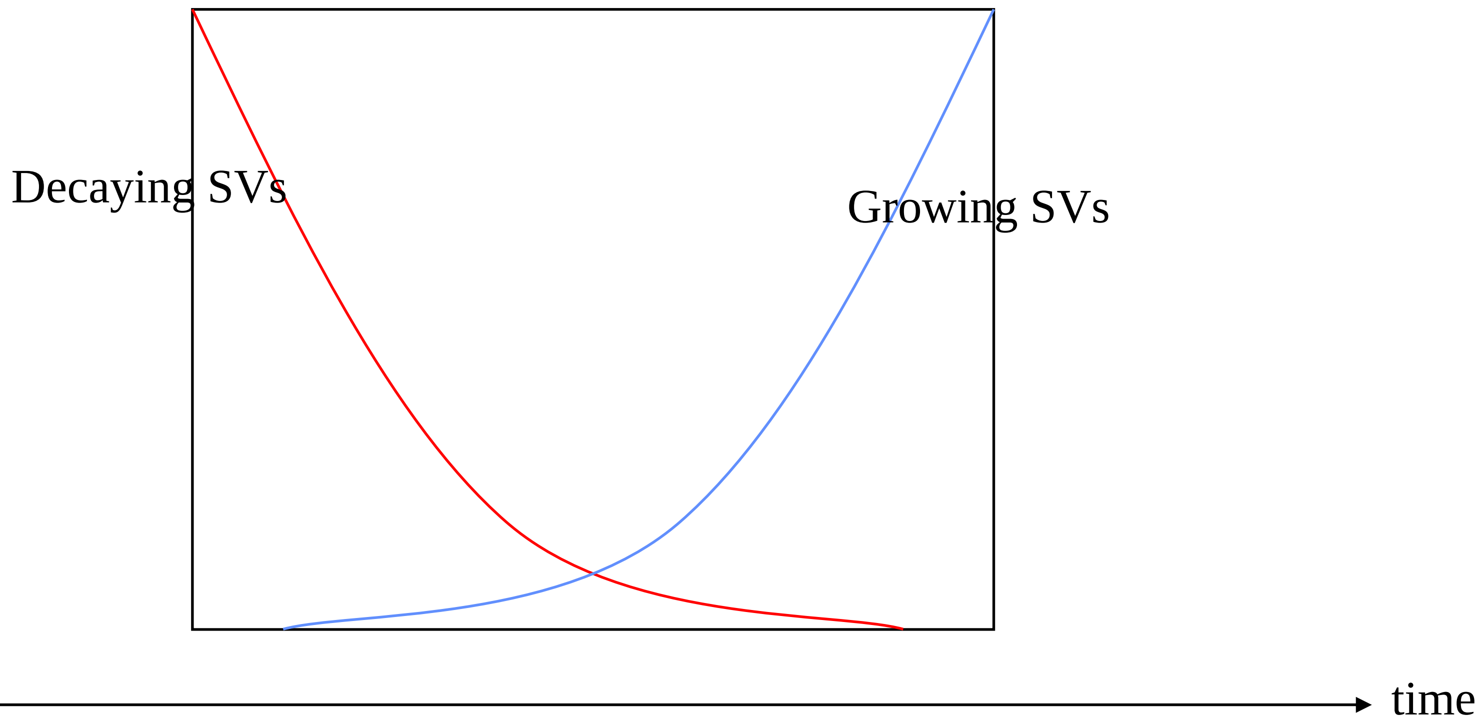
- **But, centred 24h 4dVar still only uses observations up to 12h into the future.**
- **Ideally, the analysis window should encompass <u>all</u> observations that are capable of influencing the analysis.**



- **How long a window do we need?**
- **Does it make sense to run 4dVar with very long windows?**

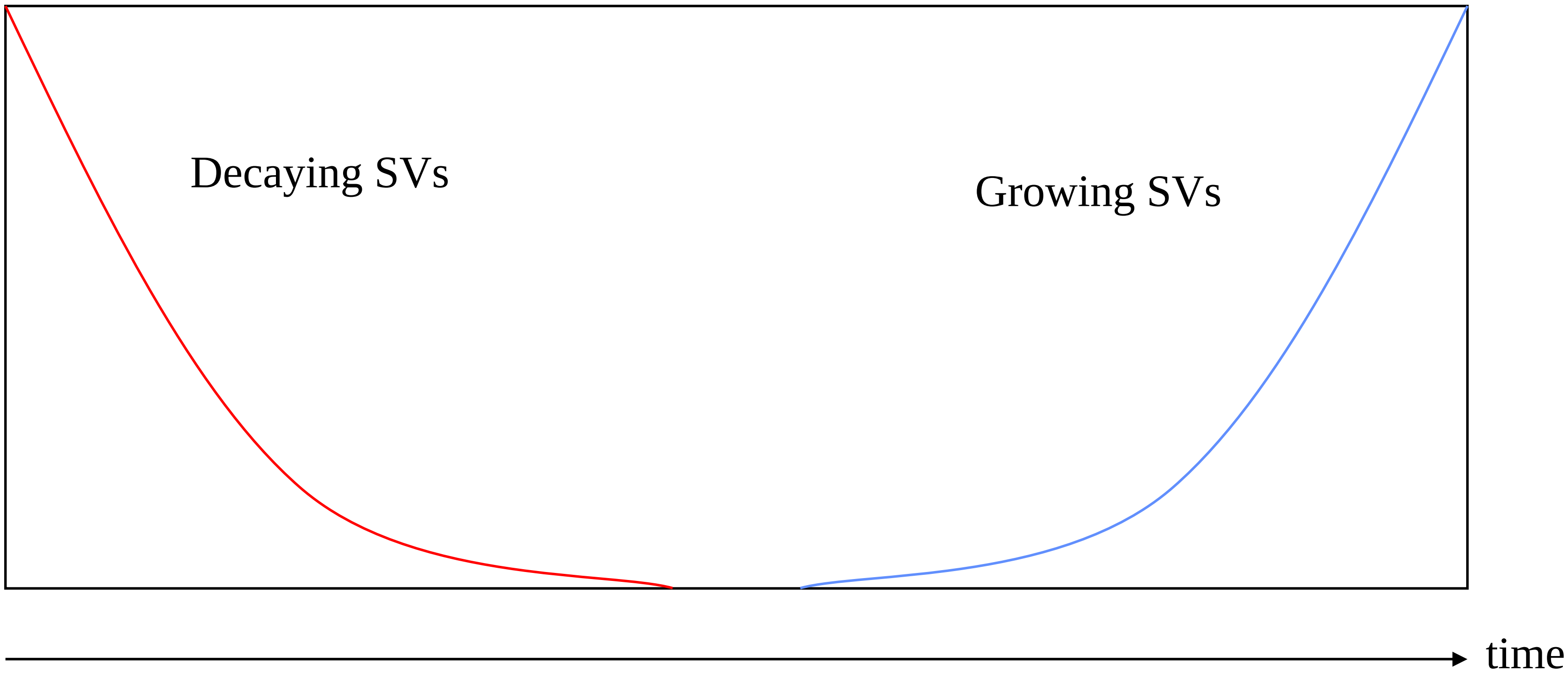
Optimizing DA for Re-Analysis

- **The analysis finds it easiest to use decaying "modes" to fit observations at the start of the window, and growing "modes" to fit observations at the end of the window.**



Optimizing DA for Re-Analysis

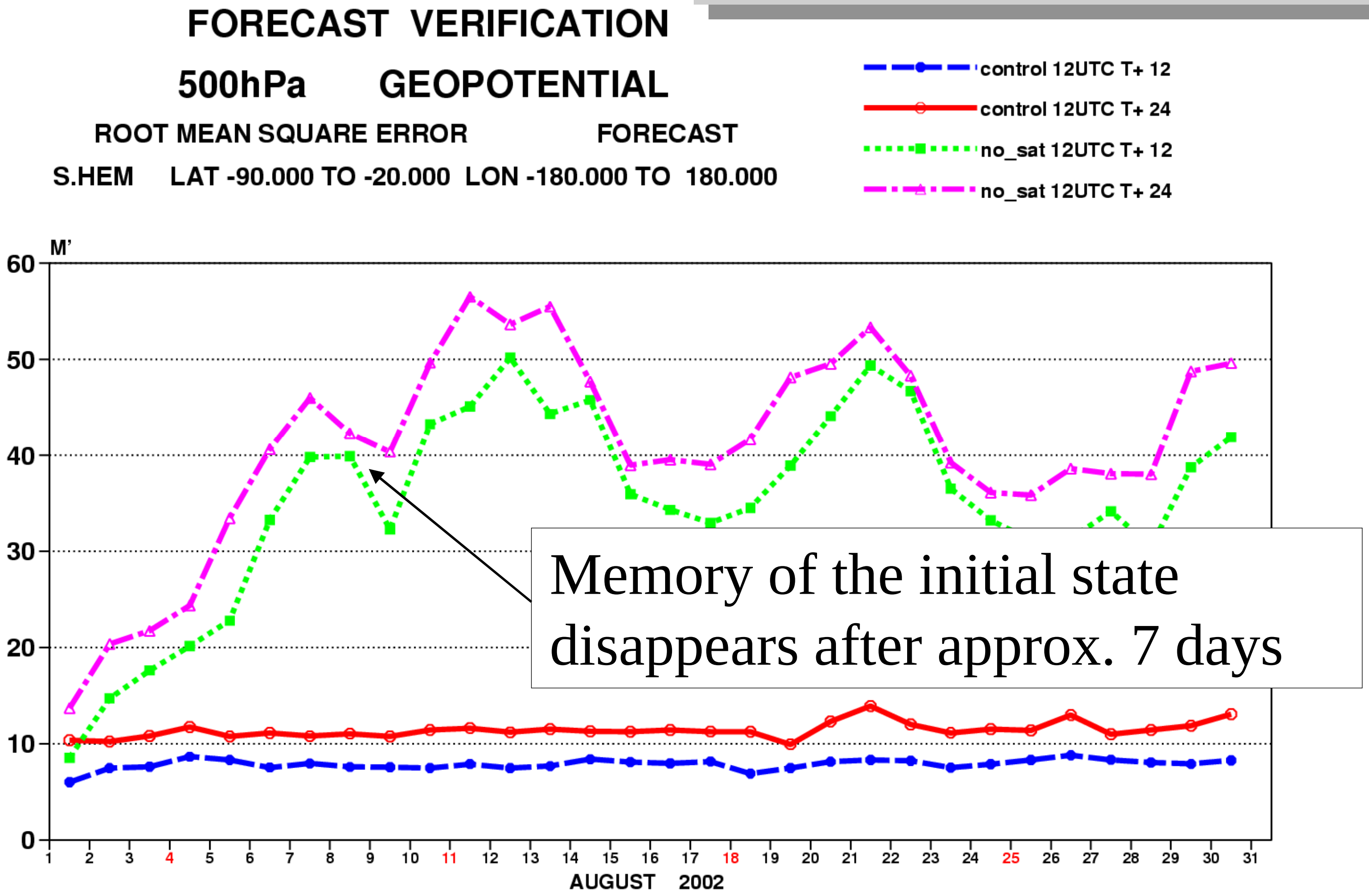
- **If the window is very long, then observations at the ends of the window will not influence the analysis at the central time.**
  - **No analysis scheme can make use of observations in the distant past or distant future: limited memory is due to the dynamics.**



ECMWF

Limited Memory

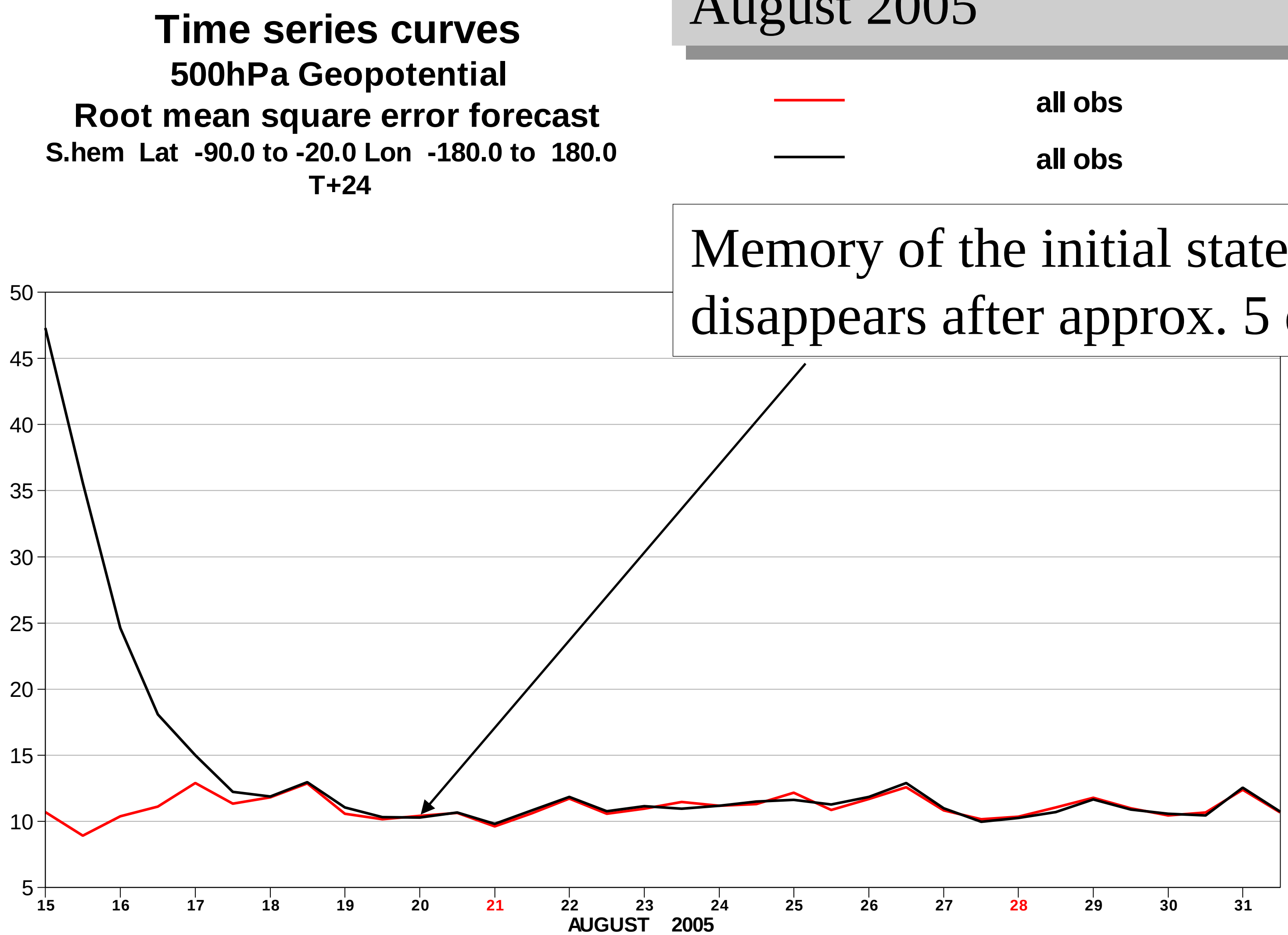
Analysis experiments started with/without satellite data on 1st August 2002



from: Graeme Kelly

Limited Memory

Analysis experiments started different initial states on 15th August 2005



Memory of the initial state disappears after approx. 5 days

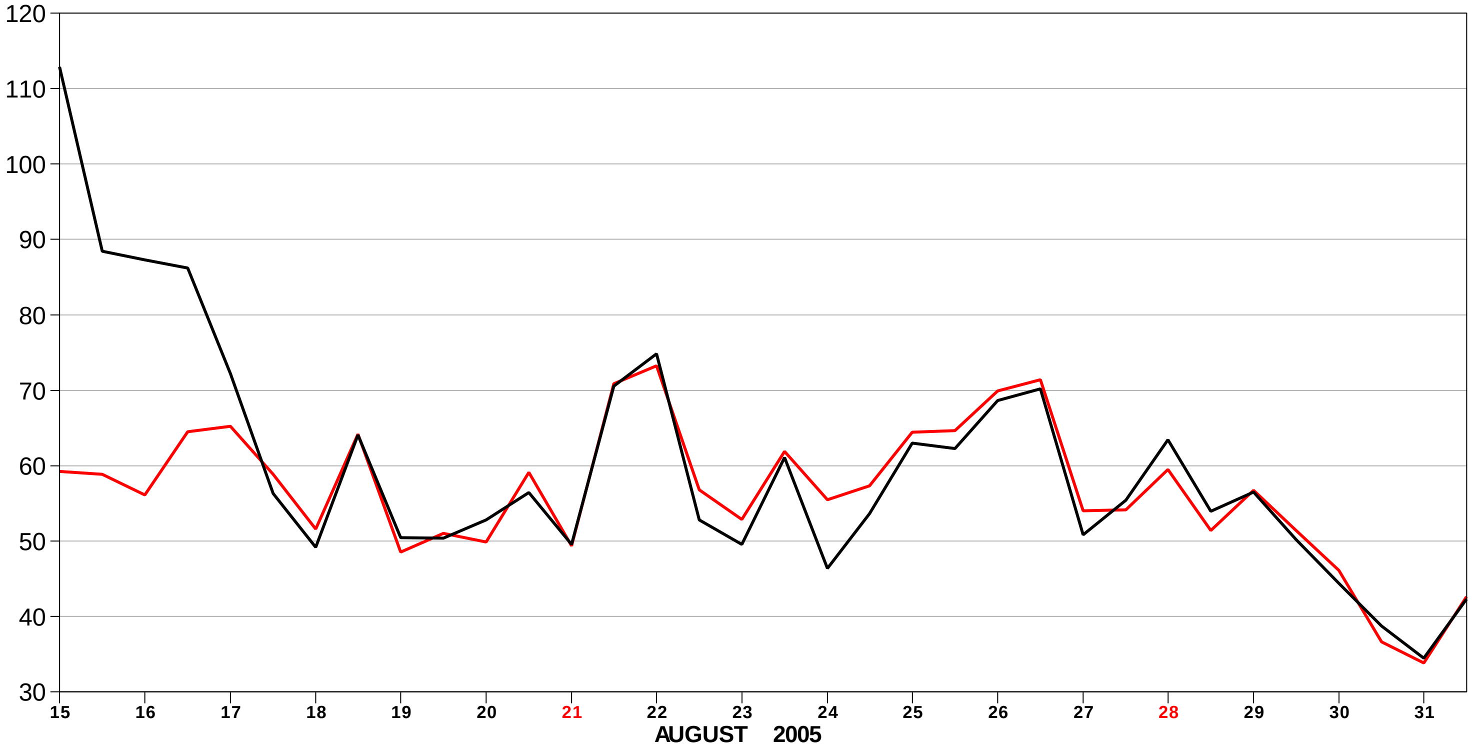
Limited Memory

Analysis experiments started different initial states on 15th August 2005

Time series curves
500hPa Geopotential
Root mean square error forecast
S.hem Lat -90.0 to -20.0 Lon -180.0 to 180.0
T+120

all obs

all obs



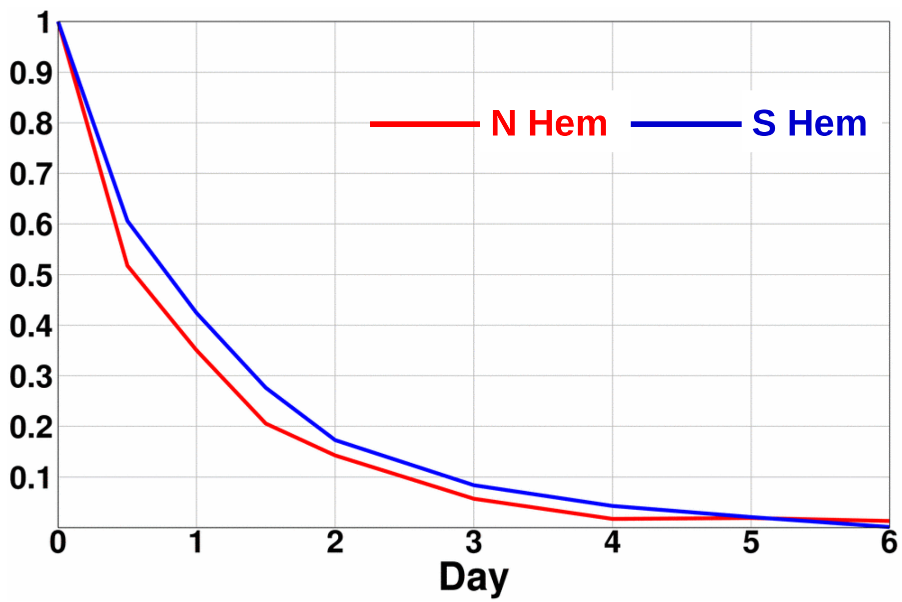
ECMWF

The Analysis forgets the initial state

- **"It takes of the order of four to five days for sufficient observations to have been assimilated for 500hPa height analysis differences to lose virtually all memory of earlier differences in background forecasts." (Simmons, 2003)**

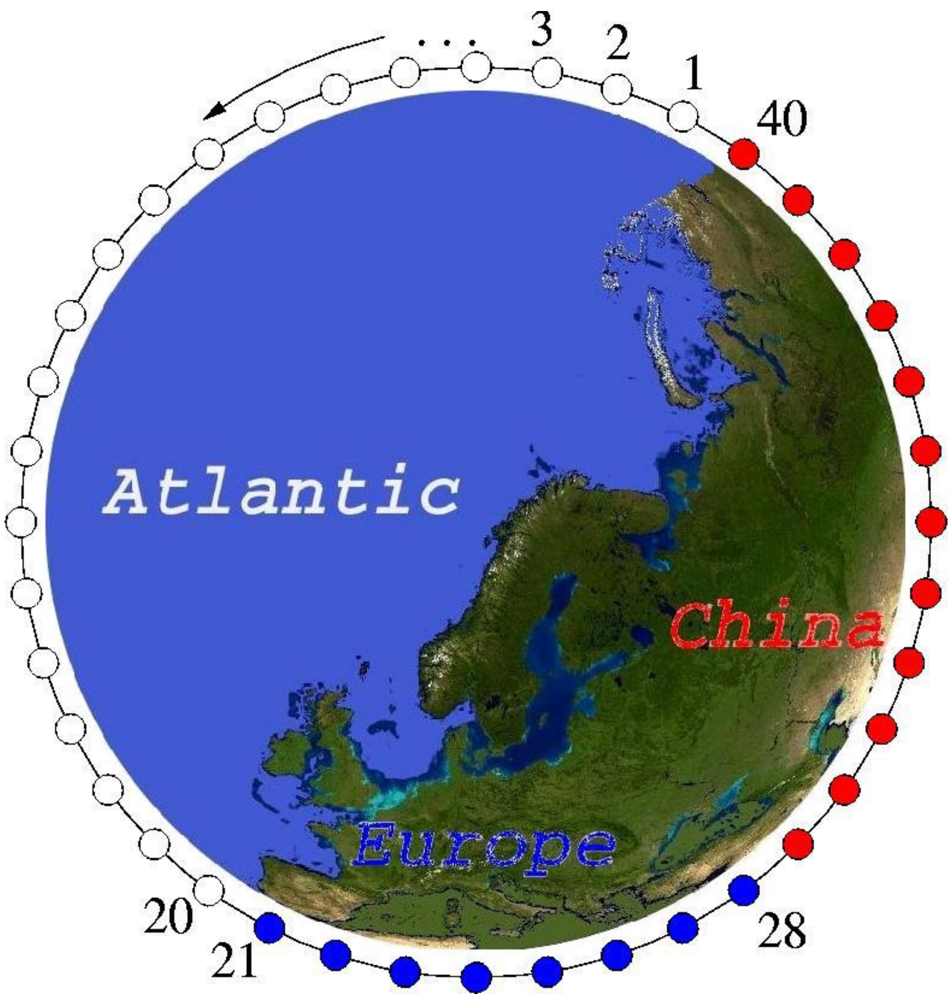
Correlation between differences in forecasts and differences in verifying analyses

(ECMWF minus UKMO)



from: Simmons (2003, proc. ECMWF Seminar)

Martin Leutbecher's "Planet L95" EKF



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

$$\text{with } i = 1, 2 \ldots 40$$

$$x_0 = x_{40}$$

$$x_{-1} = x_{39}$$

$$x_{41} = x_1$$

$$F = 8$$

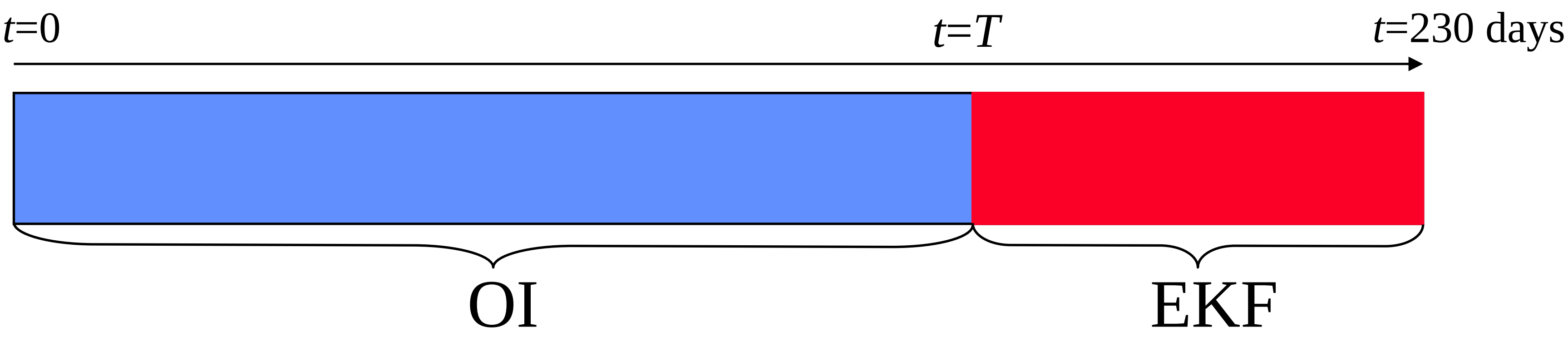
(Lorenz, 1995, ECMWF Seminar on Predictability,
and Lorenz and Emanuel, 1998)

unit time ~ 5 days

Chaotic system: 13 positive Lyapunov exponents.

The largest exponent corresponds to a doubling time of 2.1 days.

An analysis/forecast system is cycled for 230 days.
Observations are assimilated every 6 hours.
At some time $t=T$, the analysis changes from OI to EKF.
How does the quality of the final analysis vary with T?
How does the final covariance matrix vary with T?



OI

$$\mathbf{x}_k^a = \mathbf{x}_k^b + \tilde{\mathbf{K}}_k \left(\mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k^b \right)$$

$$\mathbf{x}_{k+1}^b = \mathcal{M}_{t_k \to t_{k+1}} \left(\mathbf{x}_k^a \right)$$

$$\tilde{\mathbf{K}}_k = \mathbf{B}\mathbf{H}_k^T \left(\mathbf{R}_k + \mathbf{H}_k \mathbf{B} \mathbf{H}_k^{\mathrm{T}} \right)^{-1}$$

EKF

$$\mathbf{x}_k^a = \mathbf{x}_k^b + \mathbf{K}_k \left(\mathbf{y}_k - \mathbf{H}_k \mathbf{x}_k^b \right)$$

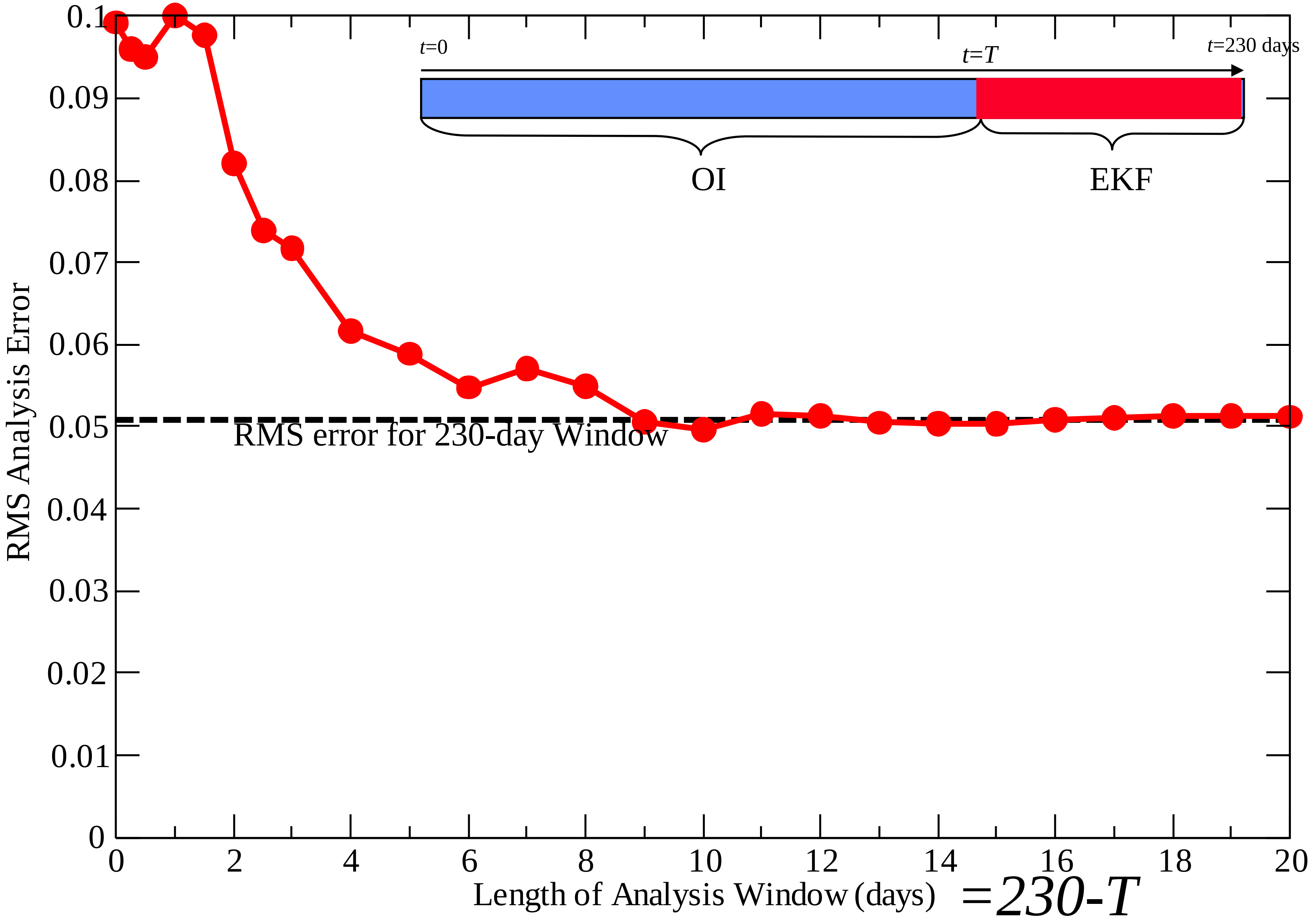
$$\mathbf{x}_{k+1}^b = \mathcal{M}_{t_k \to t_{k+1}} \left(\mathbf{x}_k^a \right)$$

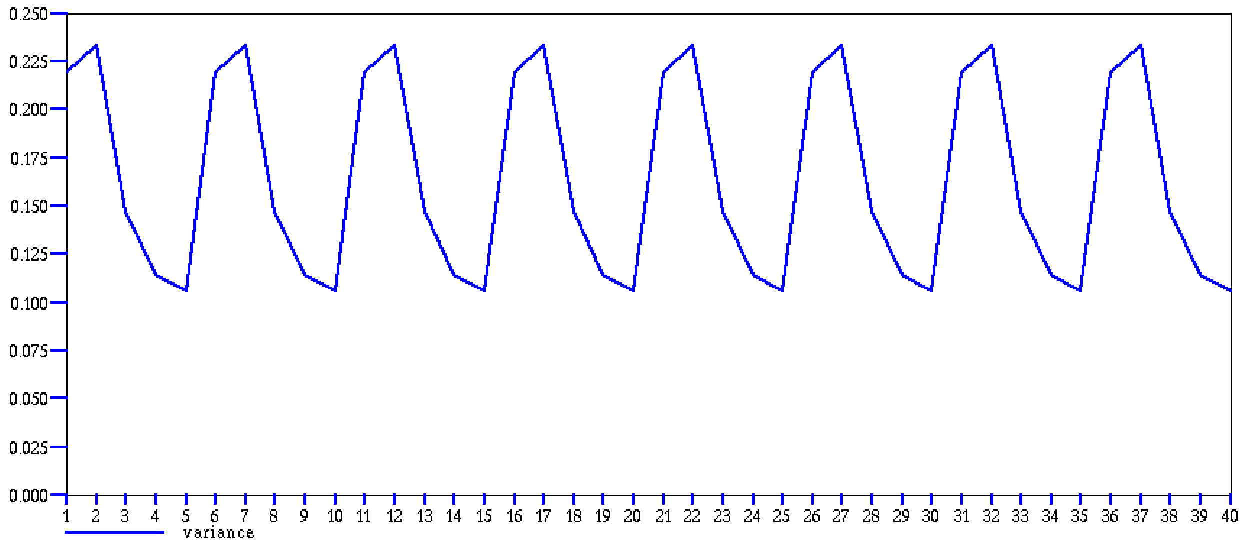
$$\mathbf{K}_k = \mathbf{P}_k^b \mathbf{H}_k^T \left(\mathbf{R}_k + \mathbf{H}_k \mathbf{P}_k^b \mathbf{H}_k^{\mathrm{T}} \right)^{-1}$$

$$\mathbf{P}_k^a = \left[\mathbf{R}_k^{-1} + \left(\mathbf{P}_k^b \right)^{-1} \right]^{-1}$$

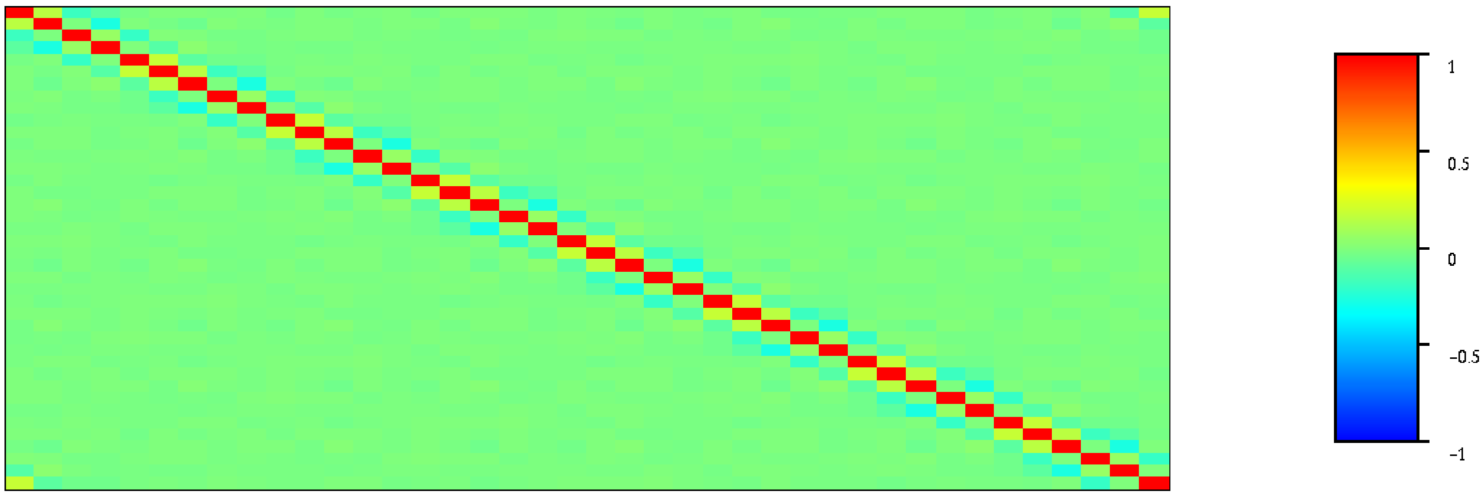
$$\mathbf{P}_{k+1}^f = \mathbf{M}_{t_k \to t_{k+1}} \mathbf{P}_k^a \mathbf{M}_{t_k \to t_{k+1}}^{\mathrm{T}} + \mathbf{Q}_{k+1}$$

RMS Error of Final (day 230) Analysis

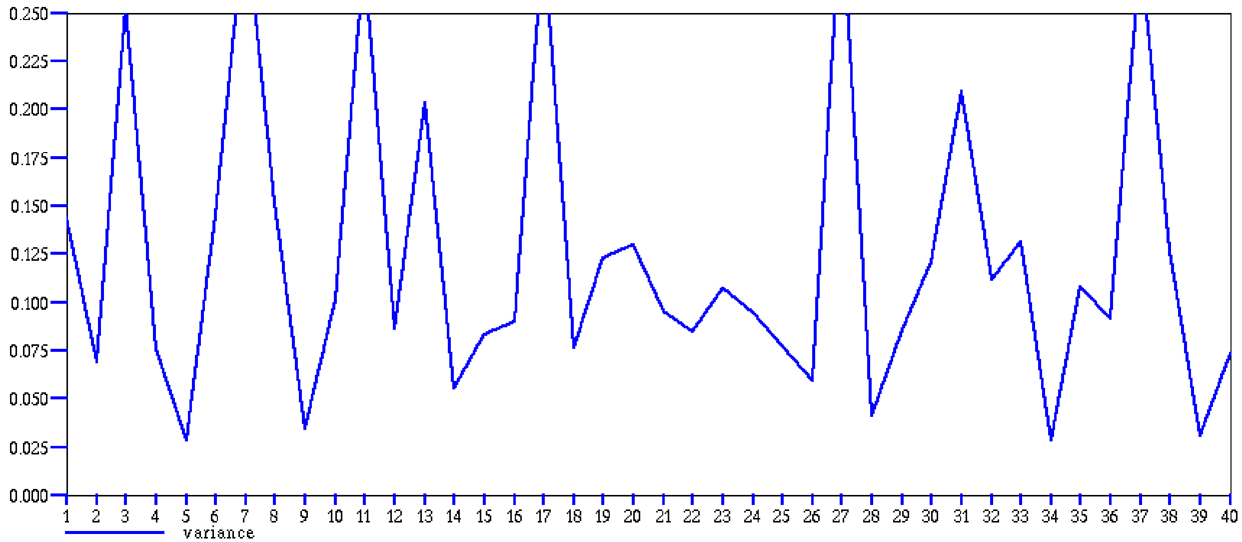




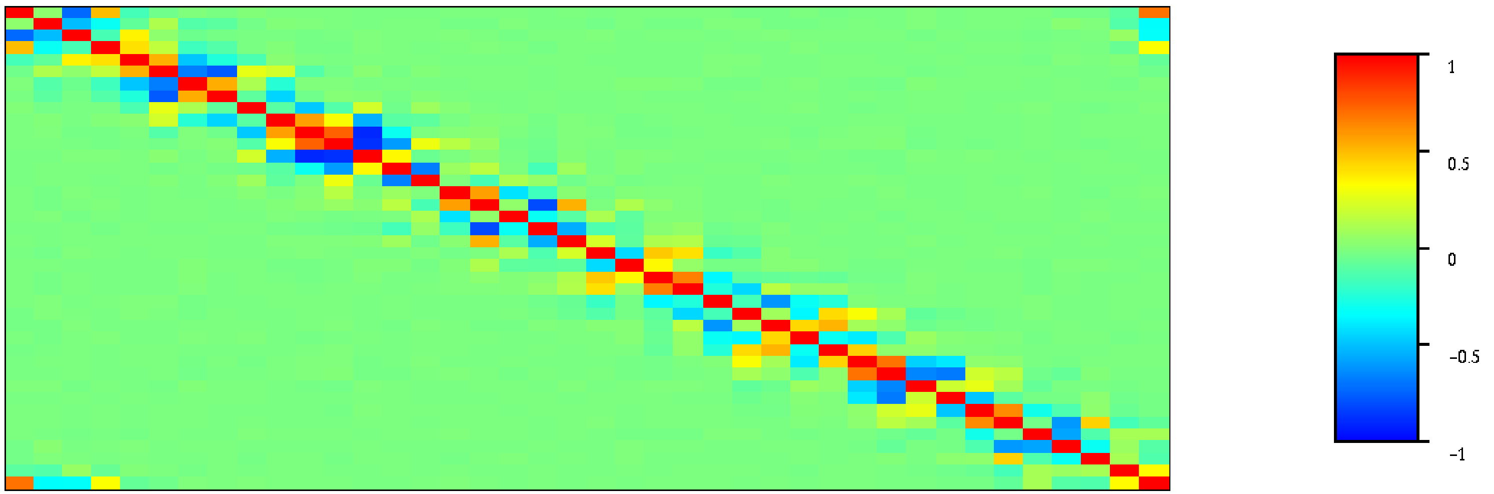
Window = 0 days



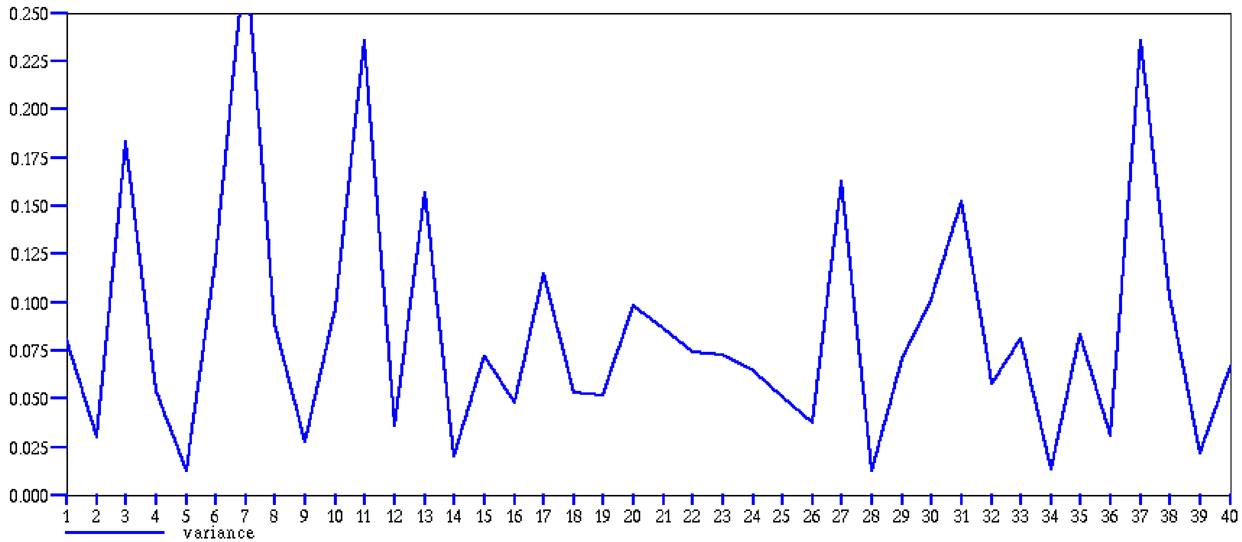
ECMWF



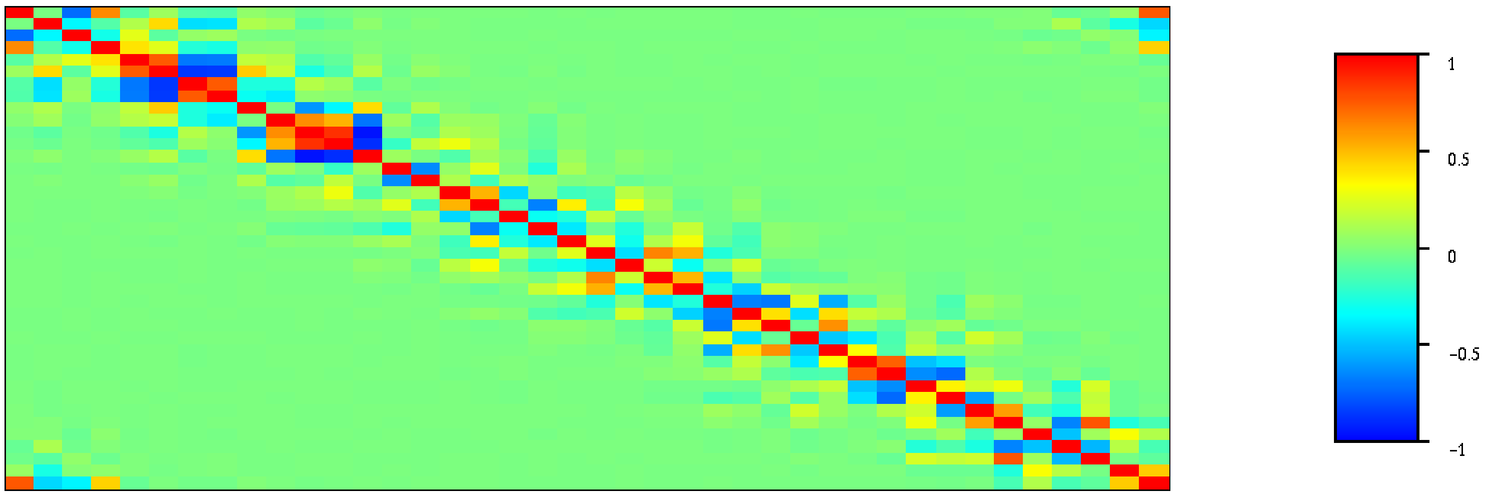
Window = 1 days



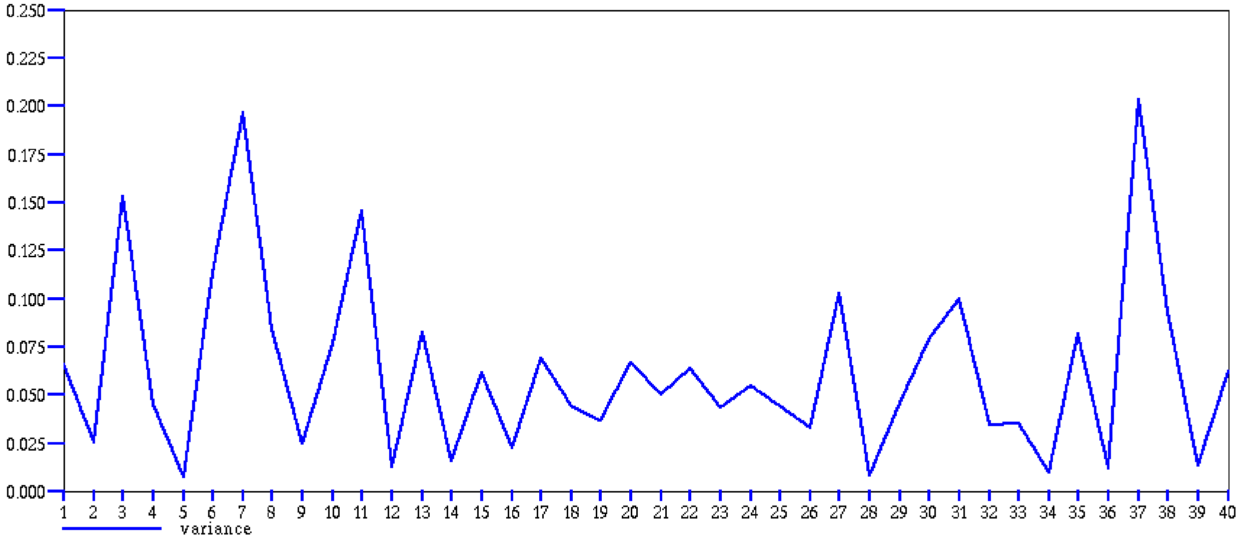
ECMWF



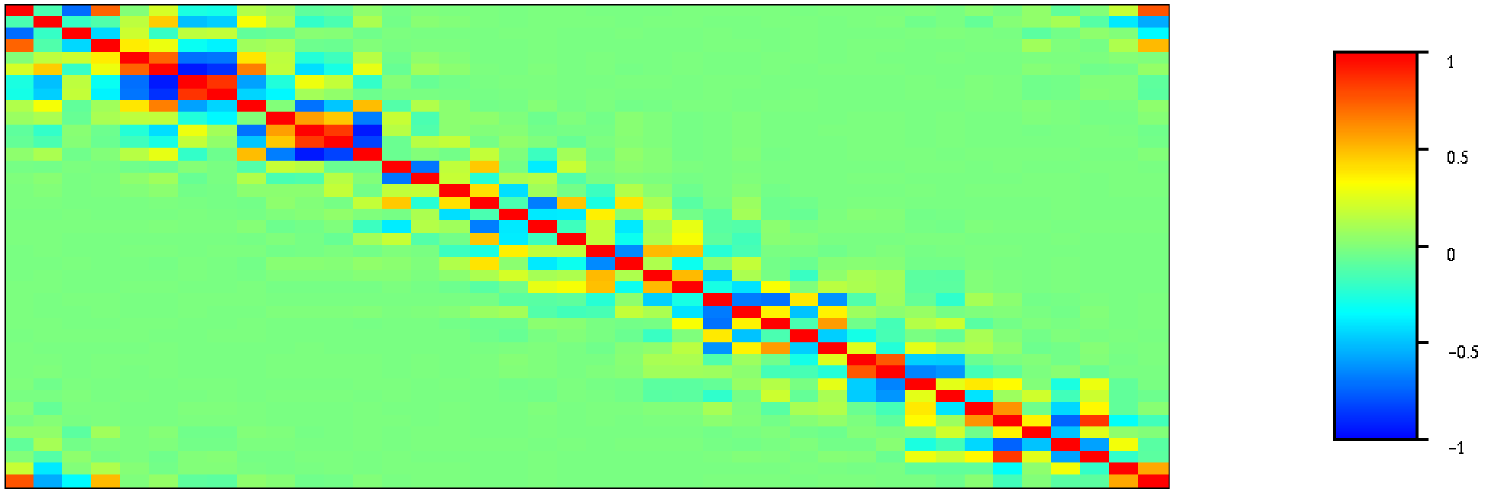
Window = 2 days



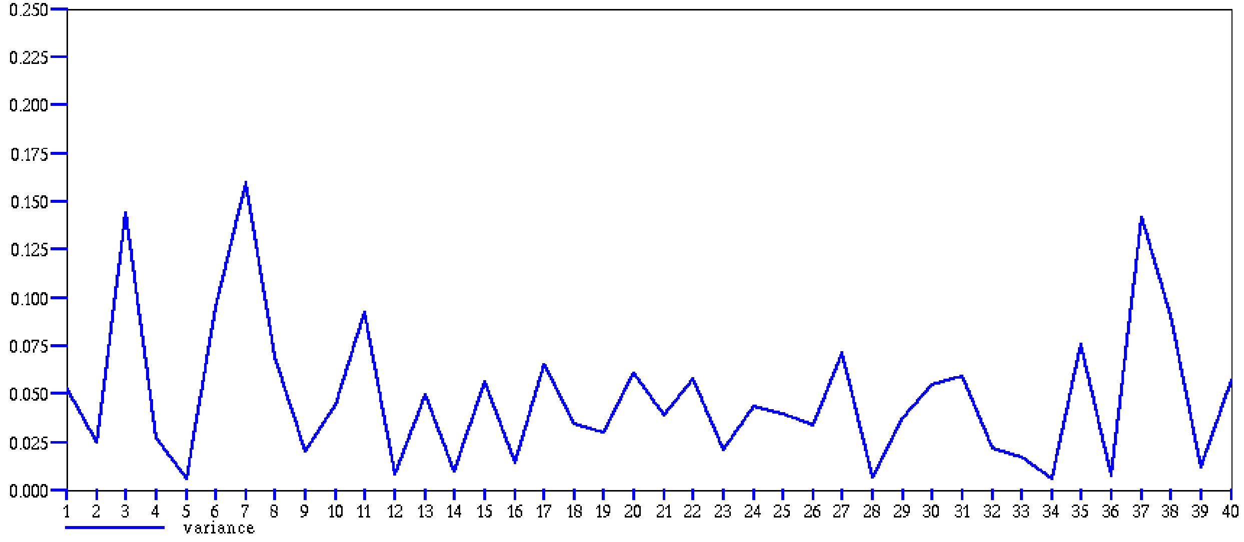
ECMWF



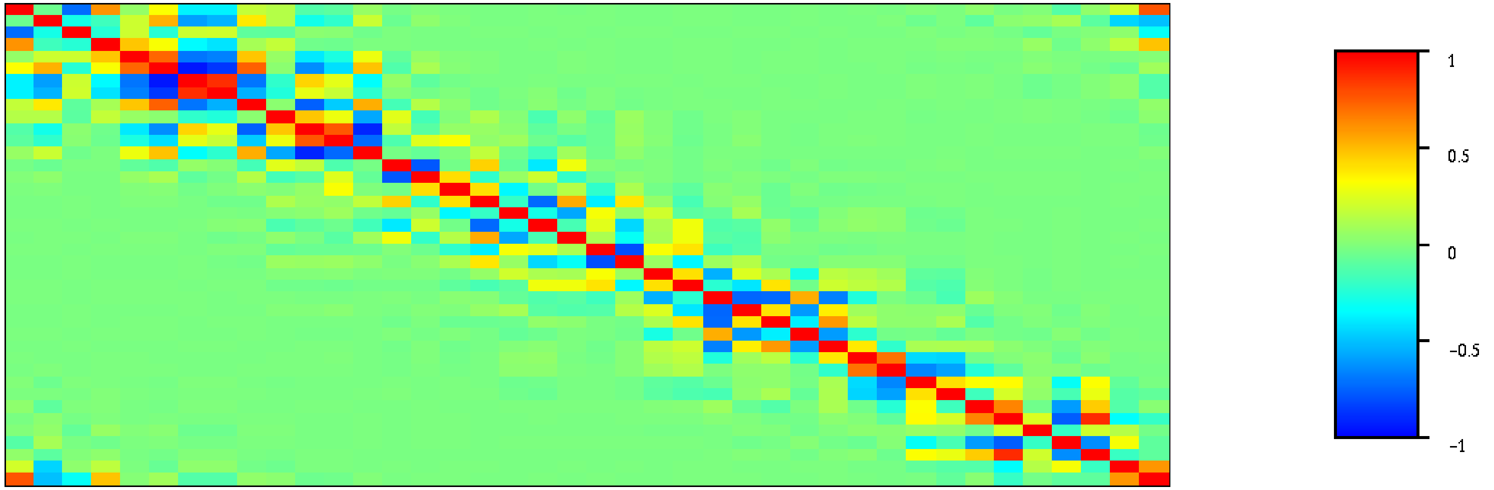
Window = 3 days



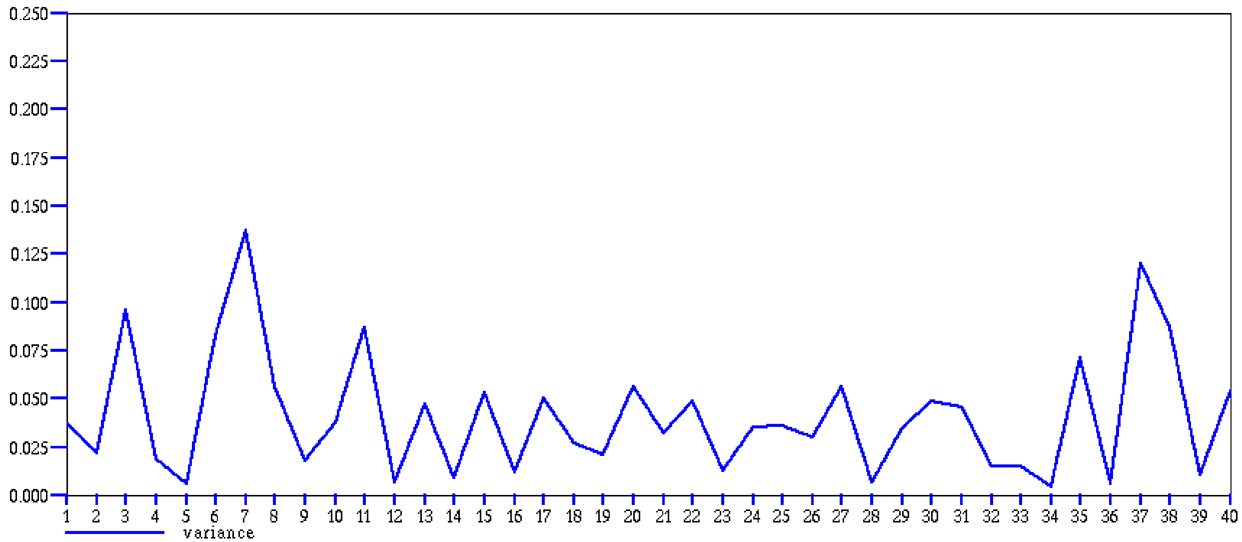
ECMWF



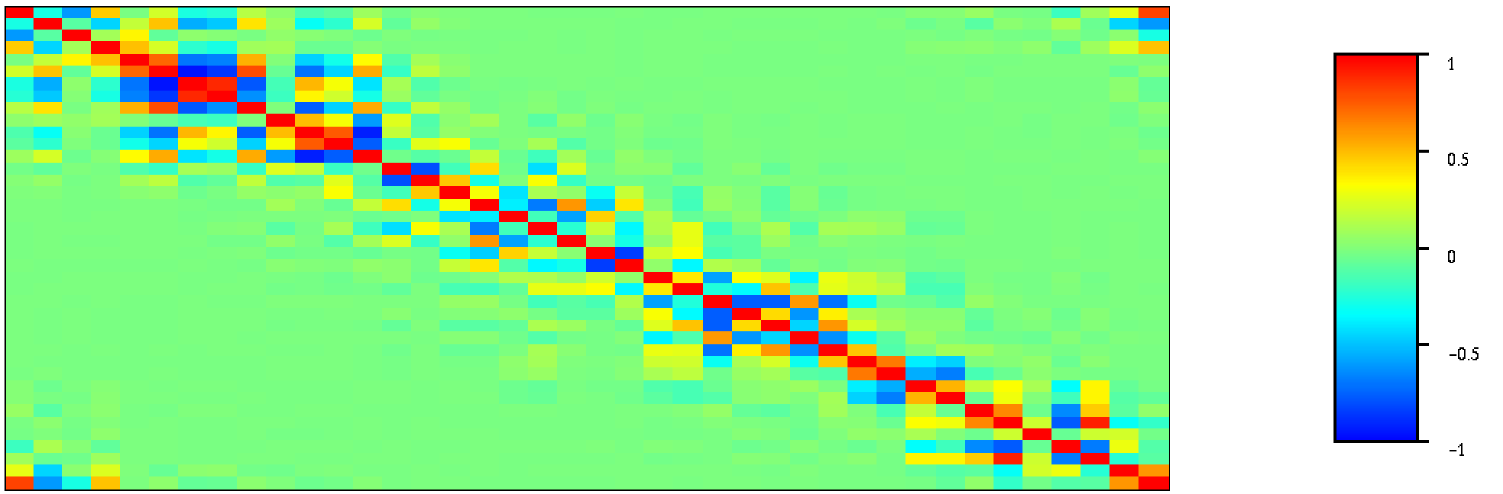
Window = 4 days

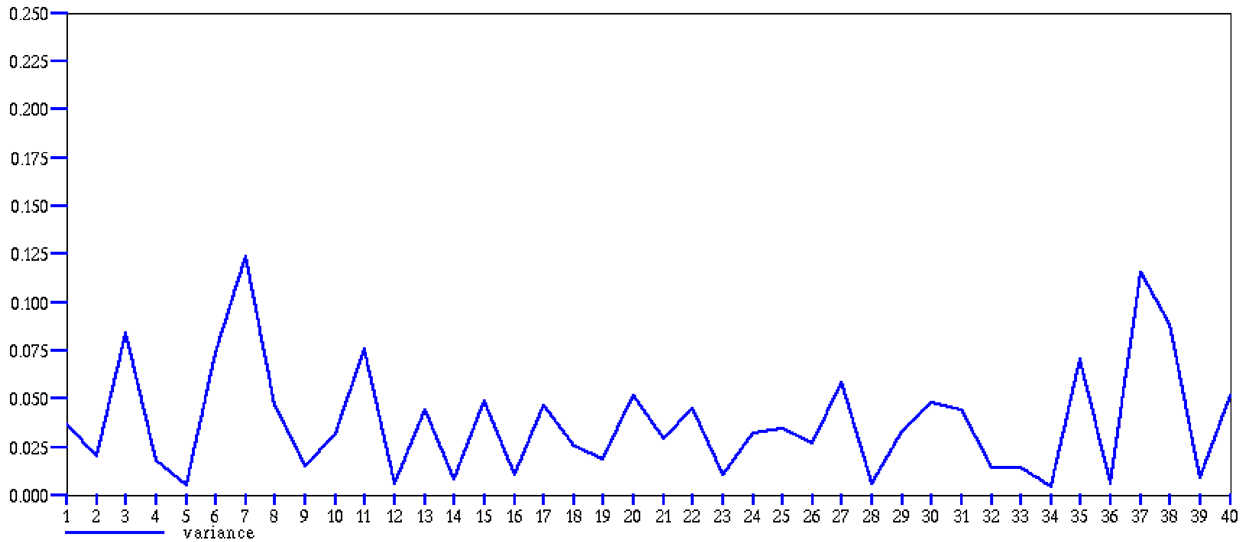


ECMWF

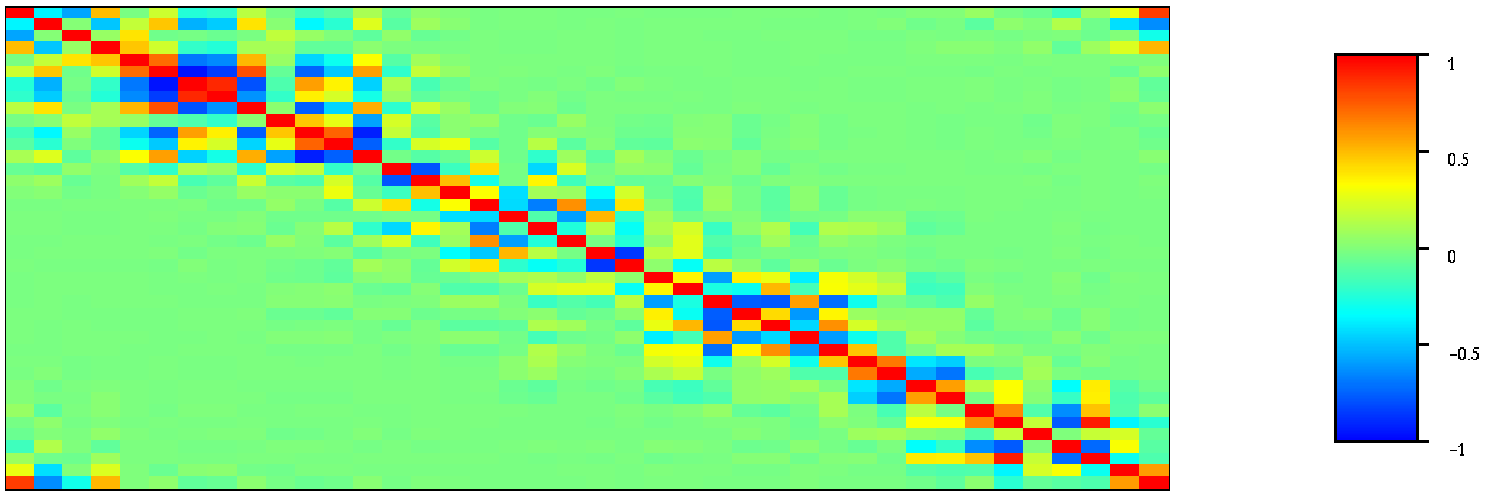


Window = 5 days

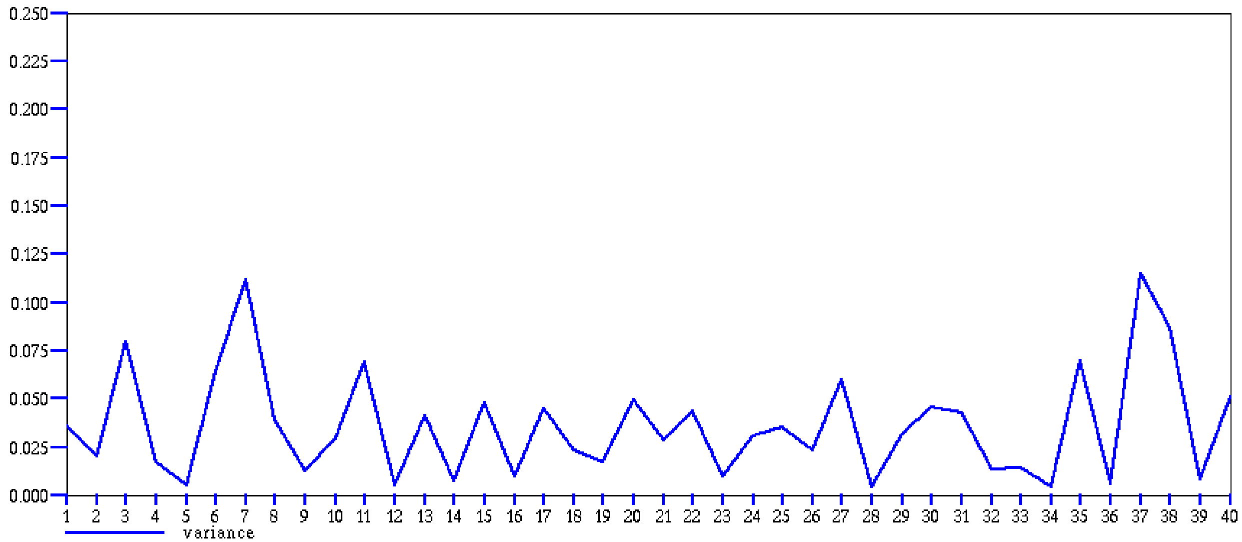




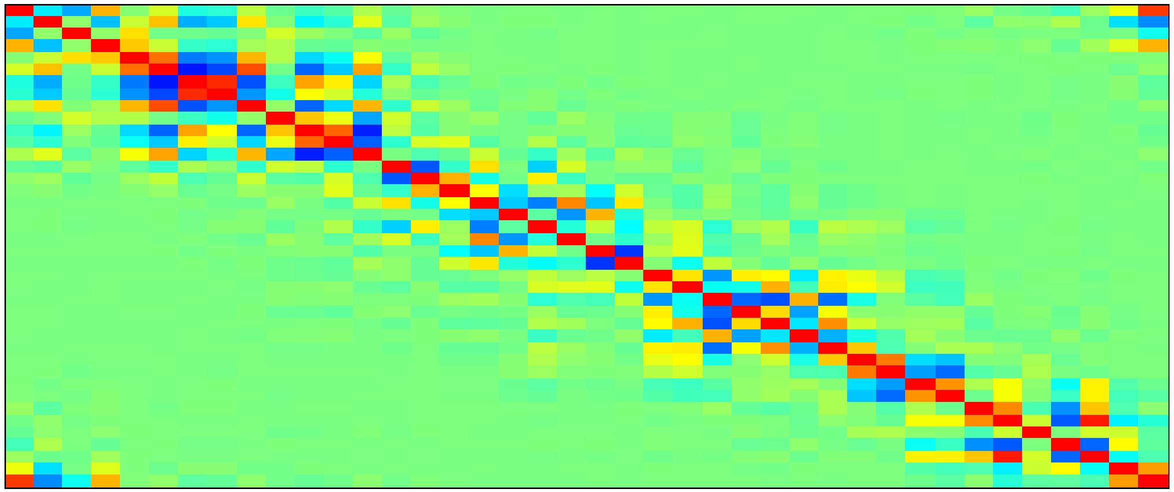
Window = 6 days

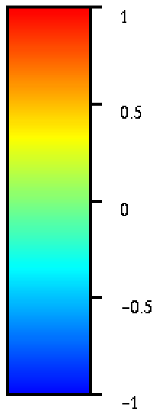


ECMWF

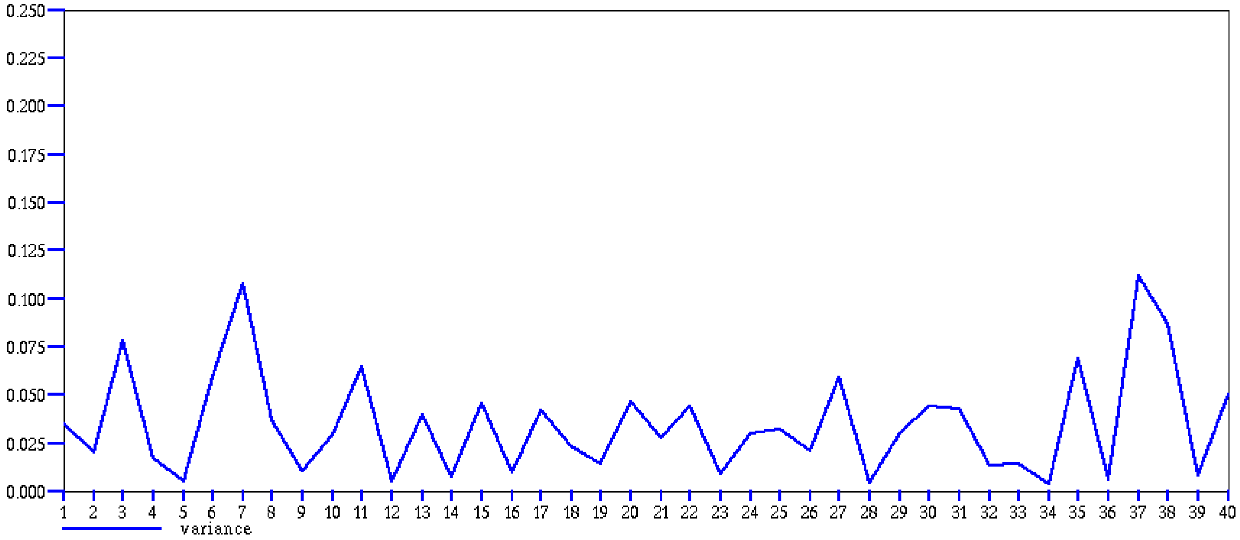


Window = 7 days

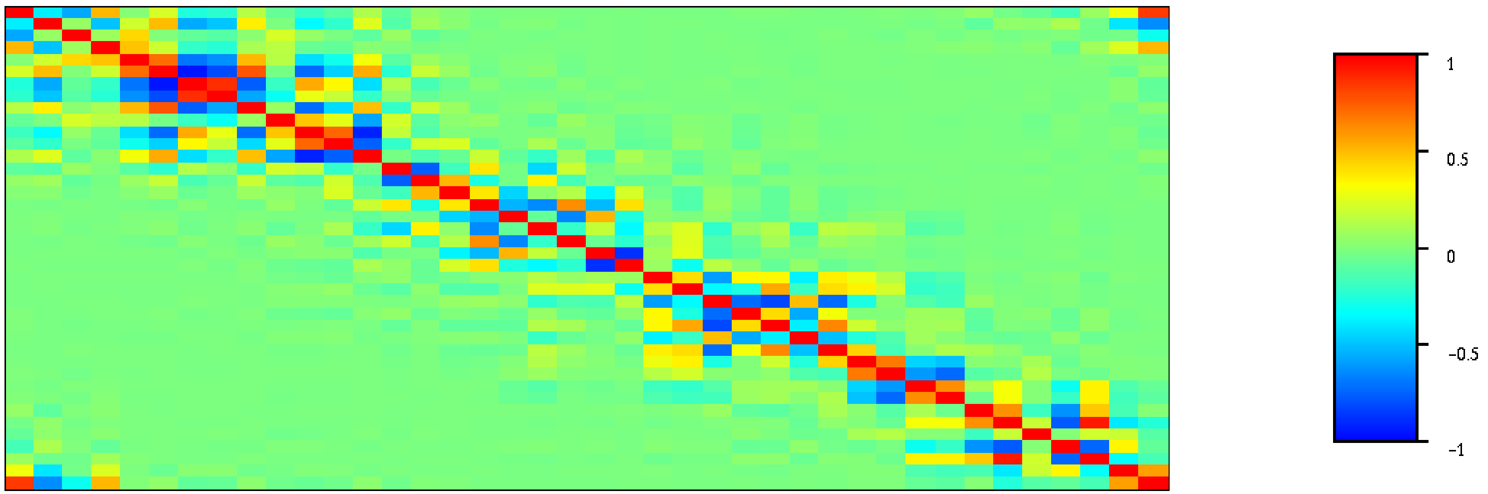


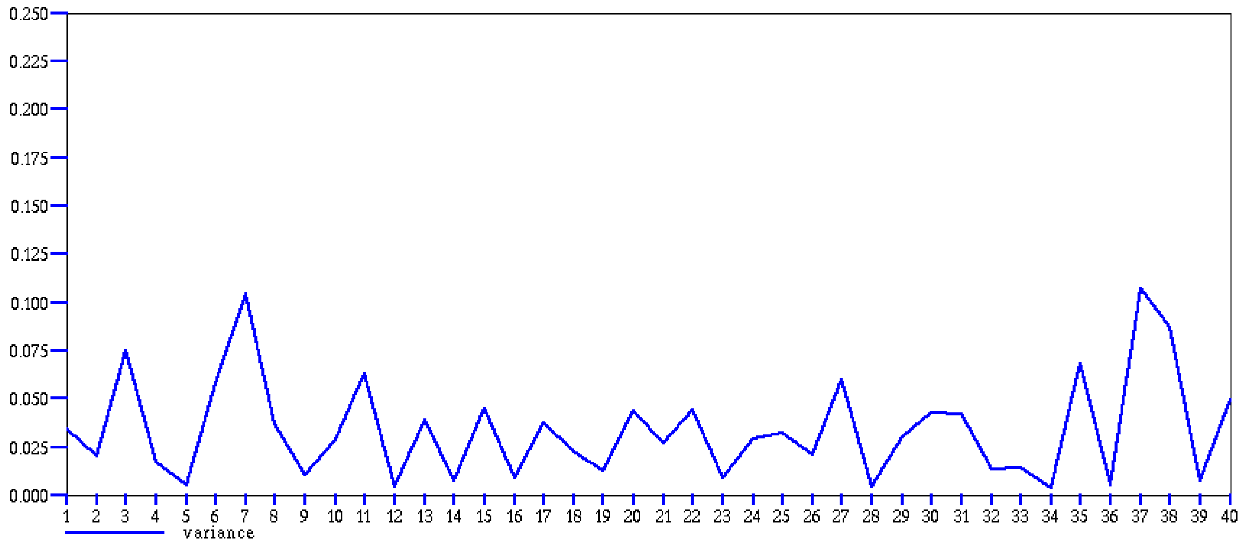


ECMWF

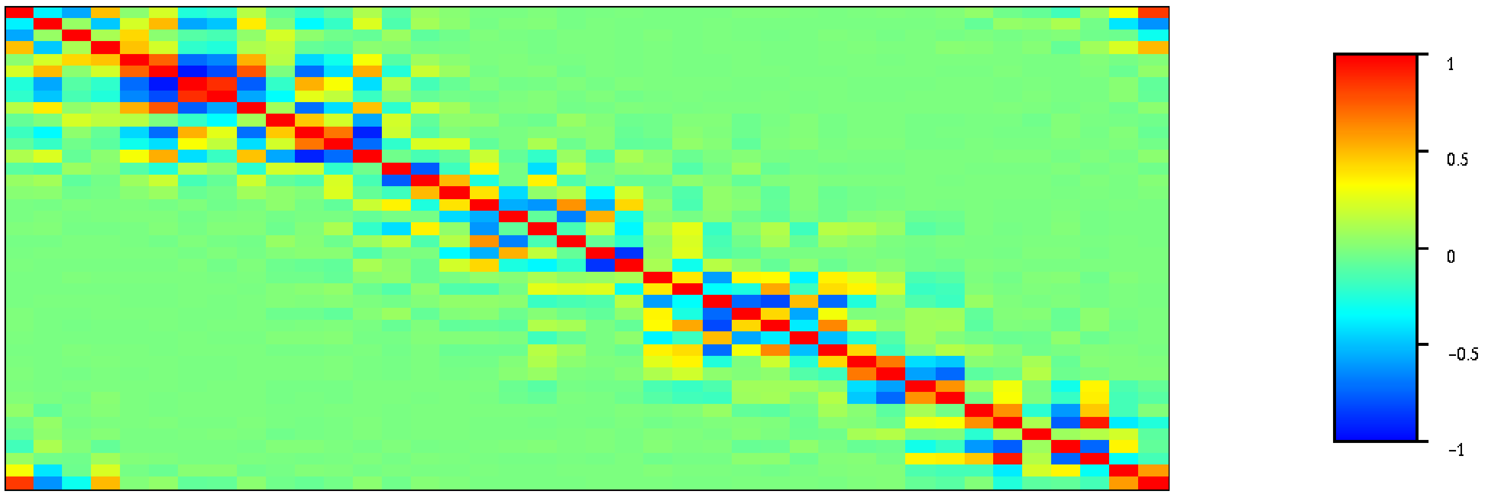


Window = 8 days

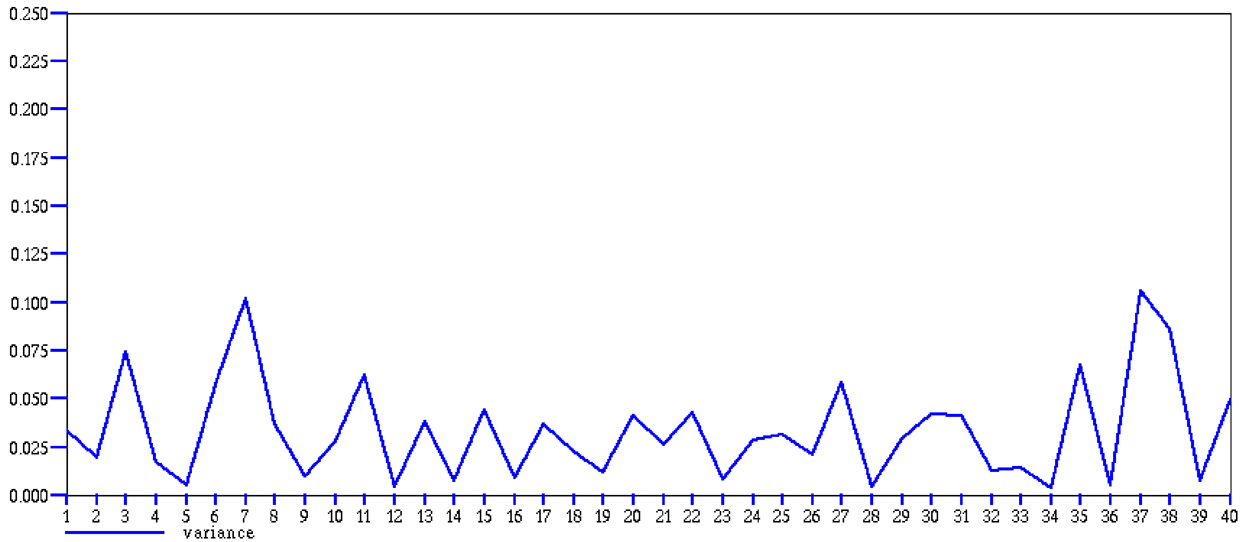




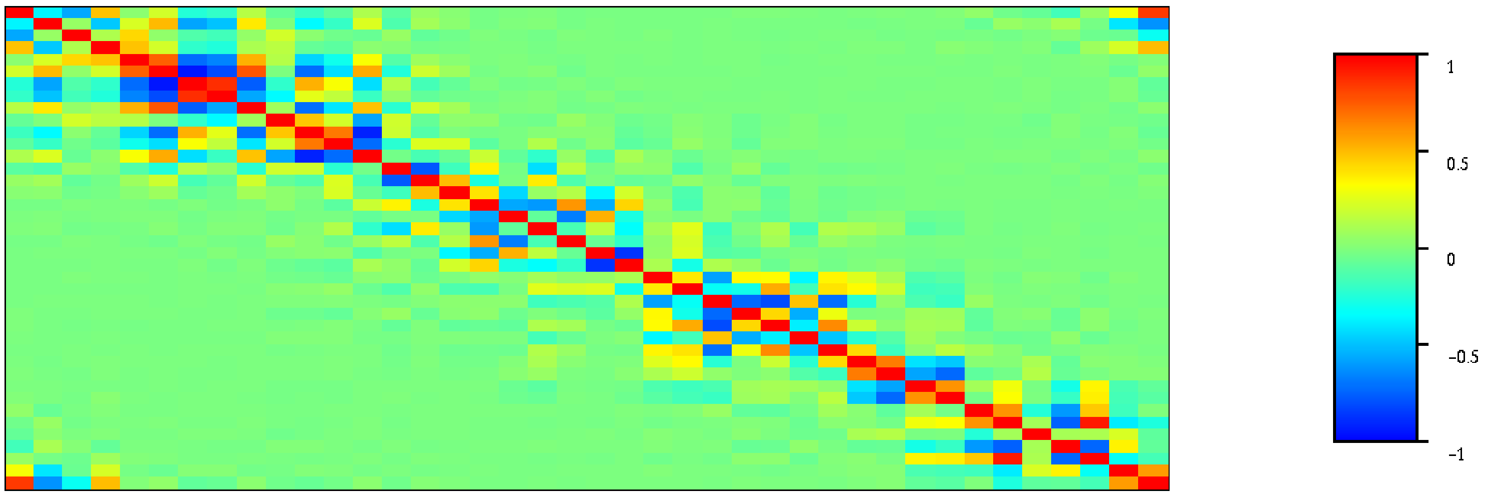
Window = 9 days



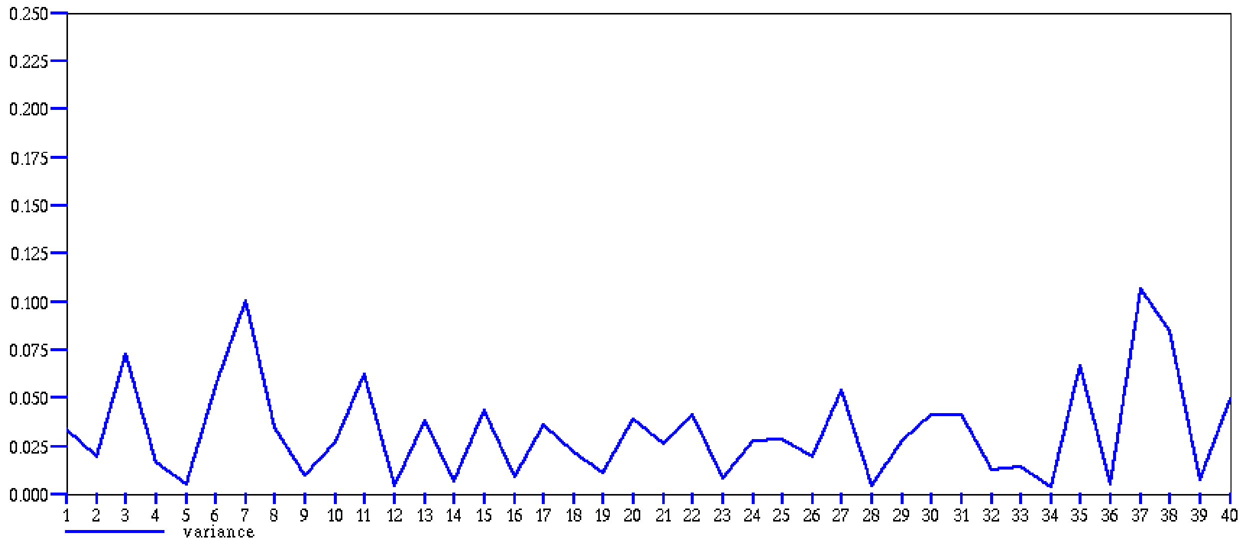
ECMWF



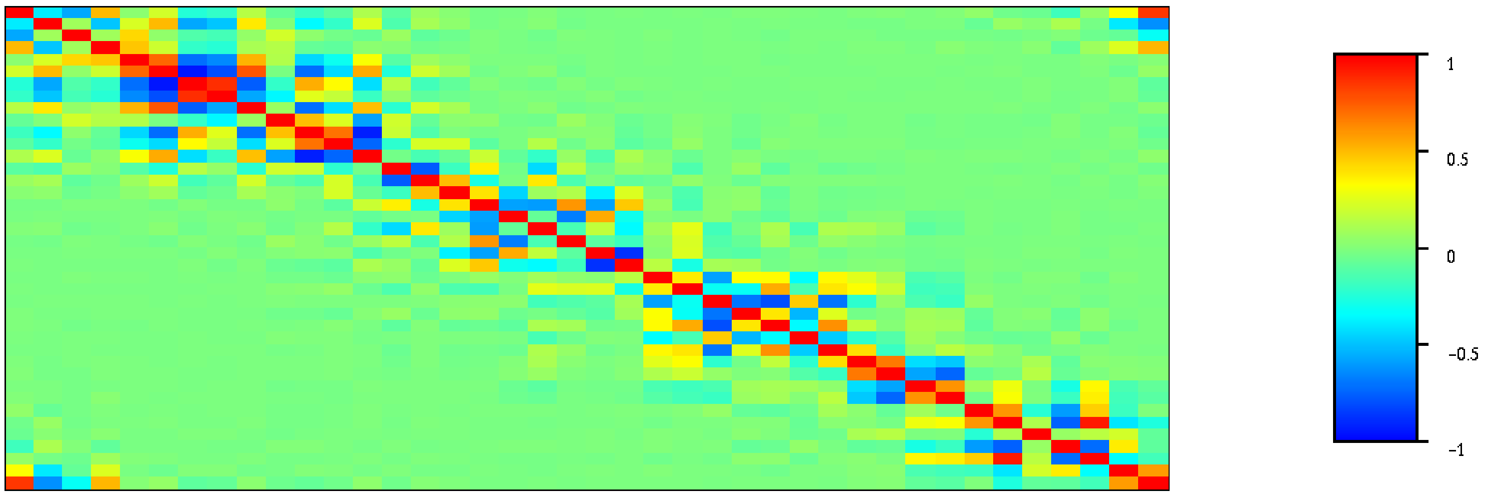
Window = 10 days



ECMWF



Window = 230 days



ECMWF

The Variational Approach to Kalman Smoothing

- **Ideally, we would like to run a Kalman Smoother over an interval of several days centred on each analysis time.**
- **The Kalman Smoother consists of forward and backward passes through the data.**
  - **The forward pass runs a Kalman Filter over the data to produce preliminary analyses at times t_k (k=0…K) that are optimal with respect to observations in the interval $[t_0 \ldots t_k]$.**
  - **The backward pass modifies the preliminary analyses to take into account observations in the interval $(t_k \ldots t_K]$.**
- **Both passes require the manipulation (and propagation) of covariance matrices of dimension $N \times N$, where N is the dimension of the state vector.**
- **The correct handling of these matrices is <u>crucial</u> to the optimality of the analysis.**

The Variational Approach to Kalman Smoothing

- **But, the Kalman Smoother is not a practical algorithm for very large (*N*~10^6) systems:**
  - **Covariance matrices contain ~10^{12} elements (and we have to invert them)!**
  - **In the Kalman Filter pass, propagating the covariance matrix requires ~10^6 model integrations!**
- **Many approximations have been suggested (e.g. Todling, Cohn and Sivakumaran, 1998), based on propagating a low-rank approximation to the covariance matrix.**
- **<u>Given the critical role of the covariance matrix in the Kalman Smoother, it is likely that these approximations have a significant impact on the optimality of the analysis.</u>**
- **NB: There is little evidence to suggest that current approximate Kalman filters are significantly more optimal than e.g. 4D-Var, at least for NWP.**

The Variational Approach to Kalman Smoothing

- **However, it is possible to implement a Kalman smoother for a large-scale system without approximating the covariance matrix.**

- **The argument relies on the algebraic equivalence between weak-constraint 4dVar and the Kalman smoother.**

- **The resulting analysis system (long-window 4dVar) is suitable for both NWP and re-analysis.**

- **Specifically (e.g. Ménard+Daley 1996):**
  - **The sequence of states ($x_0,\ldots,x_K$) generated by a fixed-interval Kalman Smoother for the interval [$t_0\ldots t_K$], with initial state x^b and initial covariance matrix $\mathbf{P}_0^b$, minimizes the cost function:**

$$J(\mathbf{x}_0,\ldots,\mathbf{x}_K) = \frac{1}{2}\left(\mathbf{x}_0 - \mathbf{x}^b\right)^{\mathrm{T}} \left(\mathbf{P}_0^b\right)^{-1} \left(\mathbf{x}_0 - \mathbf{x}^b\right) + \frac{1}{2}\sum_{k=0}^{K}\left(\mathbf{H}_k\mathbf{x}_k - \mathbf{y}_k\right)^{\mathrm{T}} \mathbf{R}_k^{-1}\left(\mathbf{H}_k\mathbf{x}_k - \mathbf{y}_k\right) + \frac{1}{2}\sum_{k=1}^{K}\left(\mathbf{x}_k - \mathbf{M}_{t_{k-1}\to t_k}\mathbf{x}_{k-1}\right)^{\mathrm{T}} \mathbf{Q}_k^{-1}\left(\mathbf{x}_k - \mathbf{M}_{t_{k-1}\to t_k}\mathbf{x}_{k-1}\right)$$

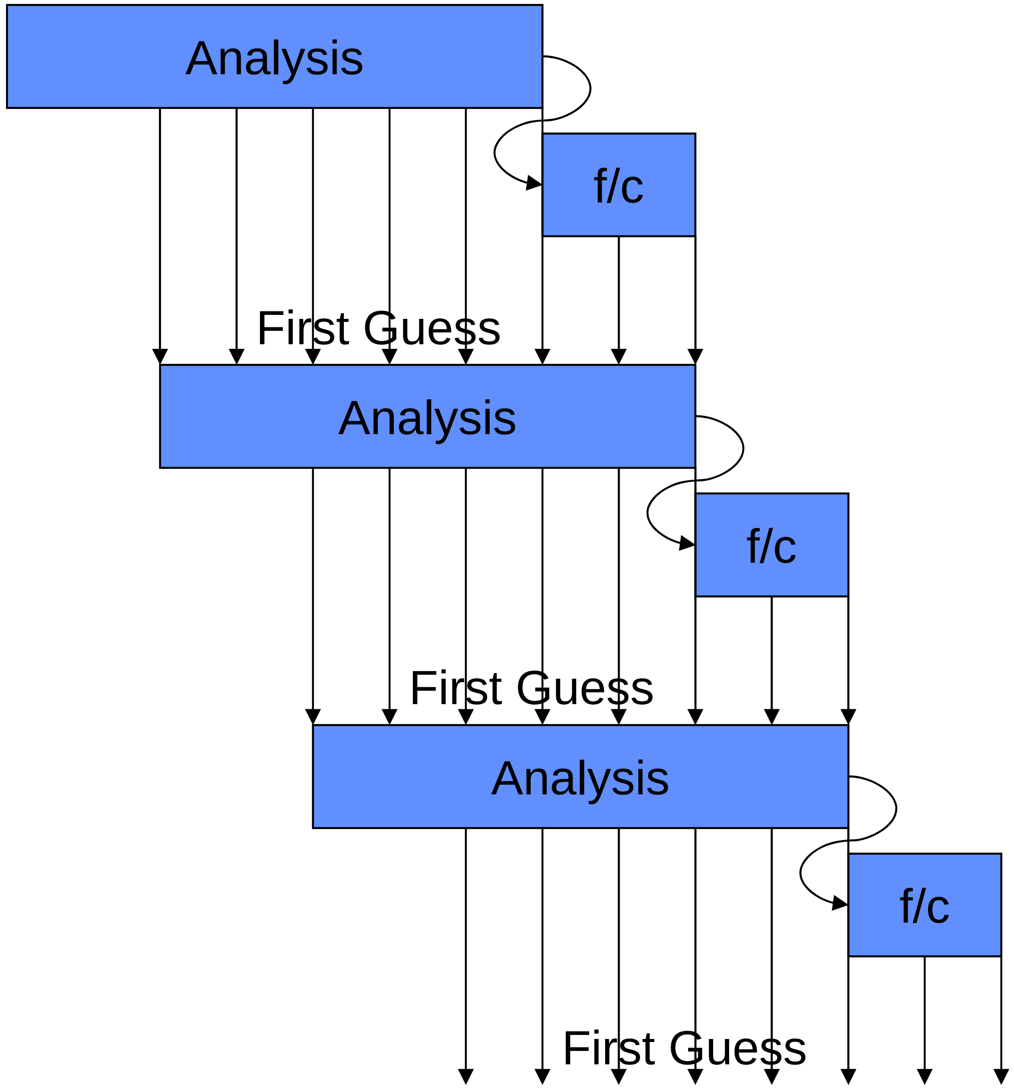
- **In fact, the variational version is more general than the Kalman Smoother. It can handle time-correlated model errors, non-gaussian observation errors, nonlinear models, etc.) .**

Demonstration of the Variational Approach for the L95 Toy Problem

- **Weak-constraint 4dVar was run for 230 days, with window lengths of 1-20 days.**
- **One cycle of analysis performed every 6 hours.**
  - **NB: Analysis windows overlap.**
- **First guess was constructed from the overlapping part of the preceding cycle, plus a 6-hour forecast:**
- **Quadratic cost function, despite nonlinear system!**
- **No background term in the cost function!**

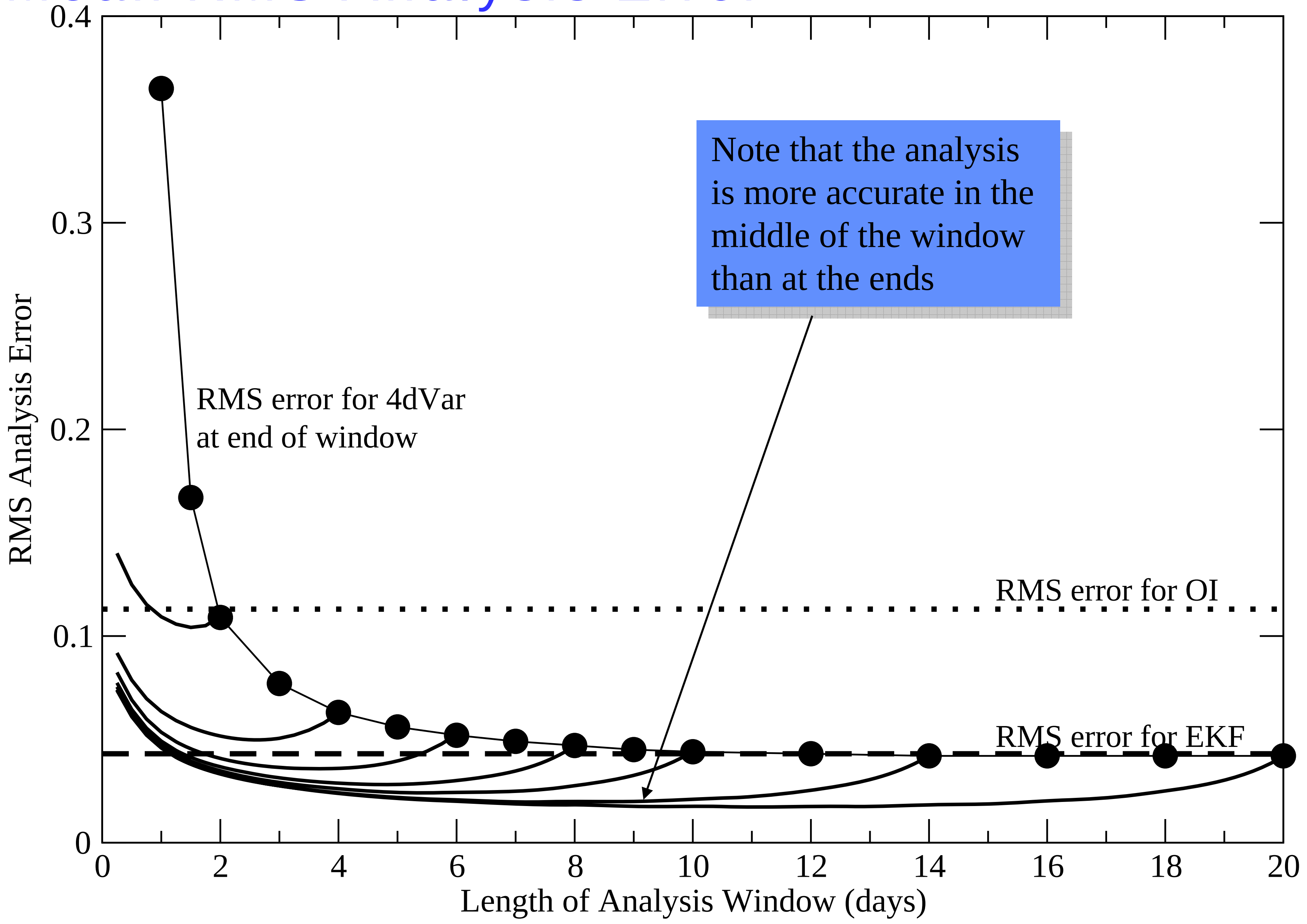
$$J(\mathbf{x}_0,\ldots,\mathbf{x}_K) = \frac{1}{2}\sum_{k=0}^{K}\left(\mathbf{H}_k\mathbf{x}_k - \mathbf{y}_k\right)^{\mathrm{T}}\mathbf{R}_k^{-1}\left(\mathbf{H}_k\mathbf{x}_k - \mathbf{y}_k\right) + \frac{1}{2}\sum_{k=1}^{K}\boldsymbol{\eta}_k^{\mathrm{T}}\mathbf{Q}_k^{-1}\boldsymbol{\eta}_k$$

$$\text{where: } \boldsymbol{\eta}_k = \mathbf{x}_k - \mathcal{M}_{t_k \to t_{k+1}}(\tilde{\mathbf{x}}_{k-1}) - \mathbf{M}_{t_{k-1}\to t_k}(\mathbf{x}_{k-1} - \tilde{\mathbf{x}}_{k-1})$$


Analysis
f/c
First Guess
Analysis
f/c
First Guess
Analysis
f/c
First Guess

ECMWF

Mean RMS Analysis Error



What about Multiple Minima?

- **Example: strong-constraint 4D-Var for the Lorenz three-variable model:**

from: Roulston, 1999

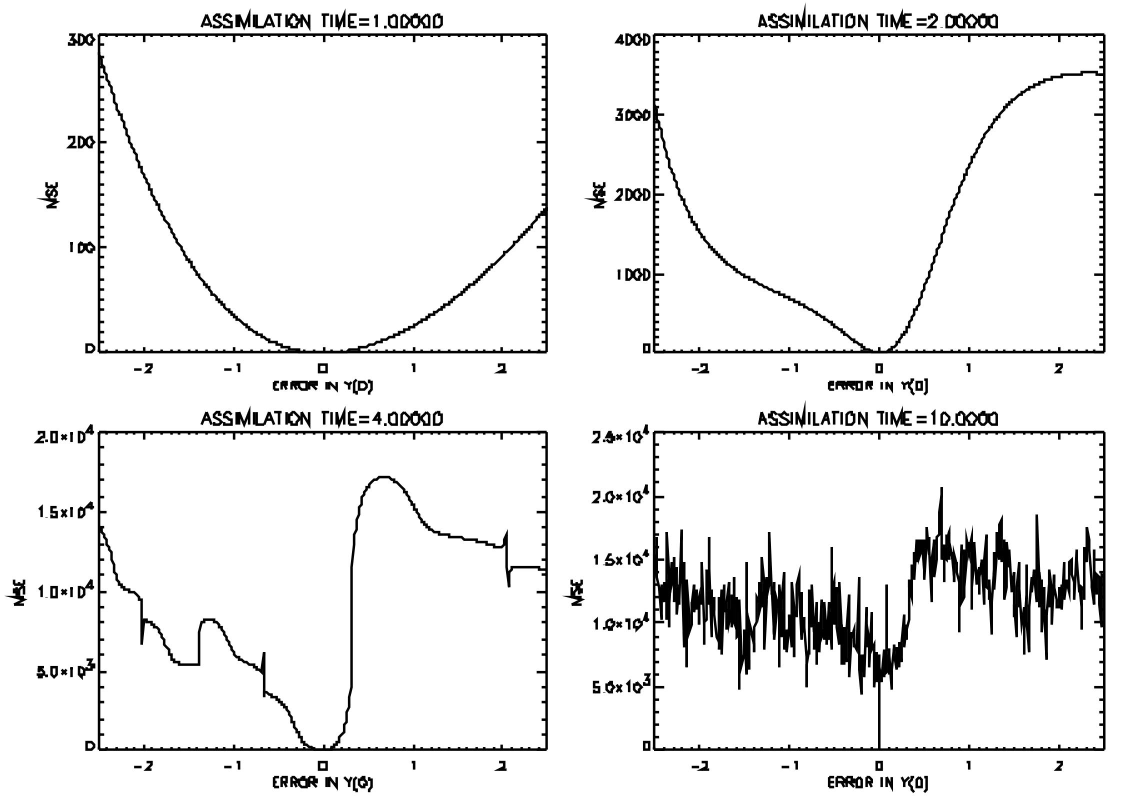
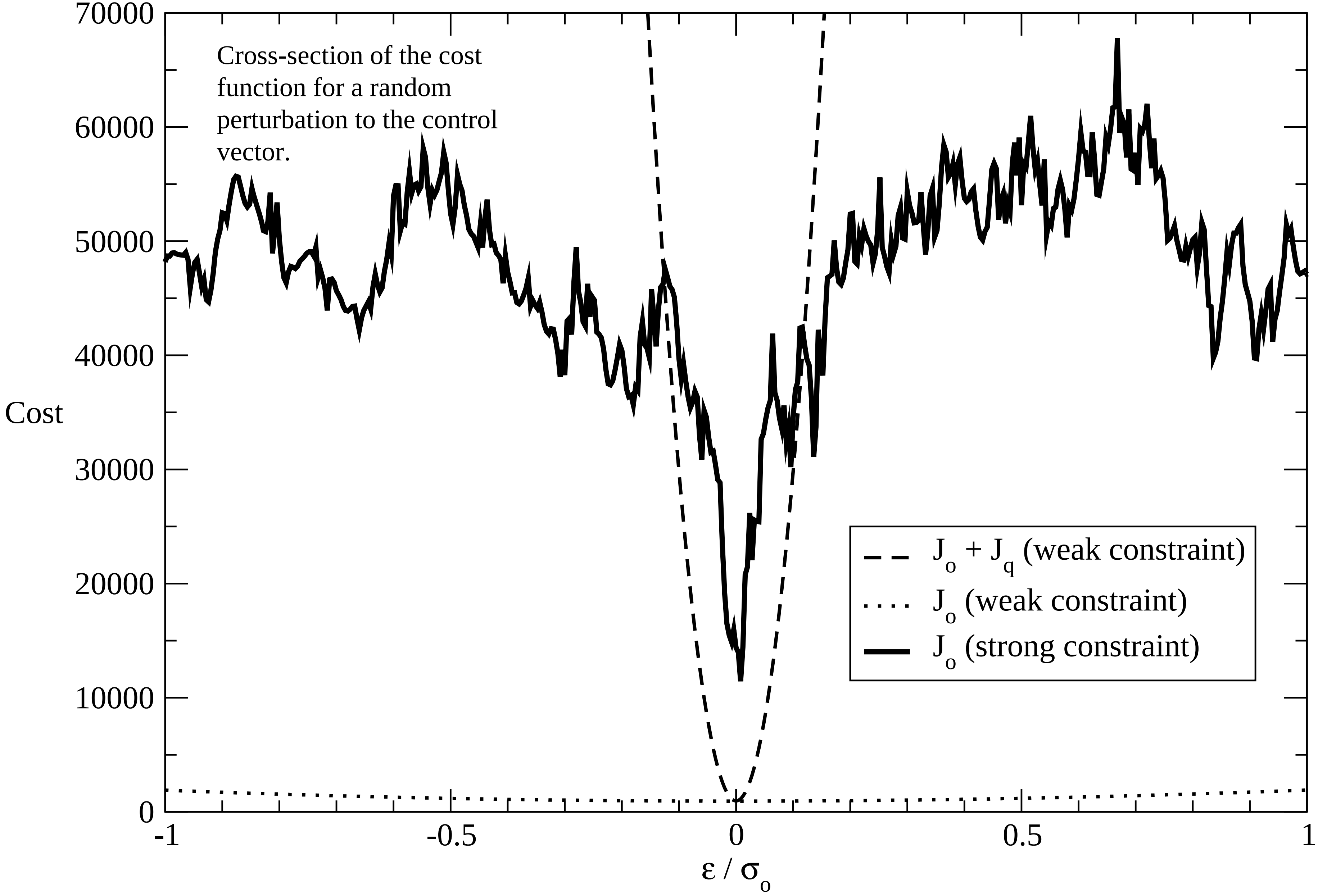
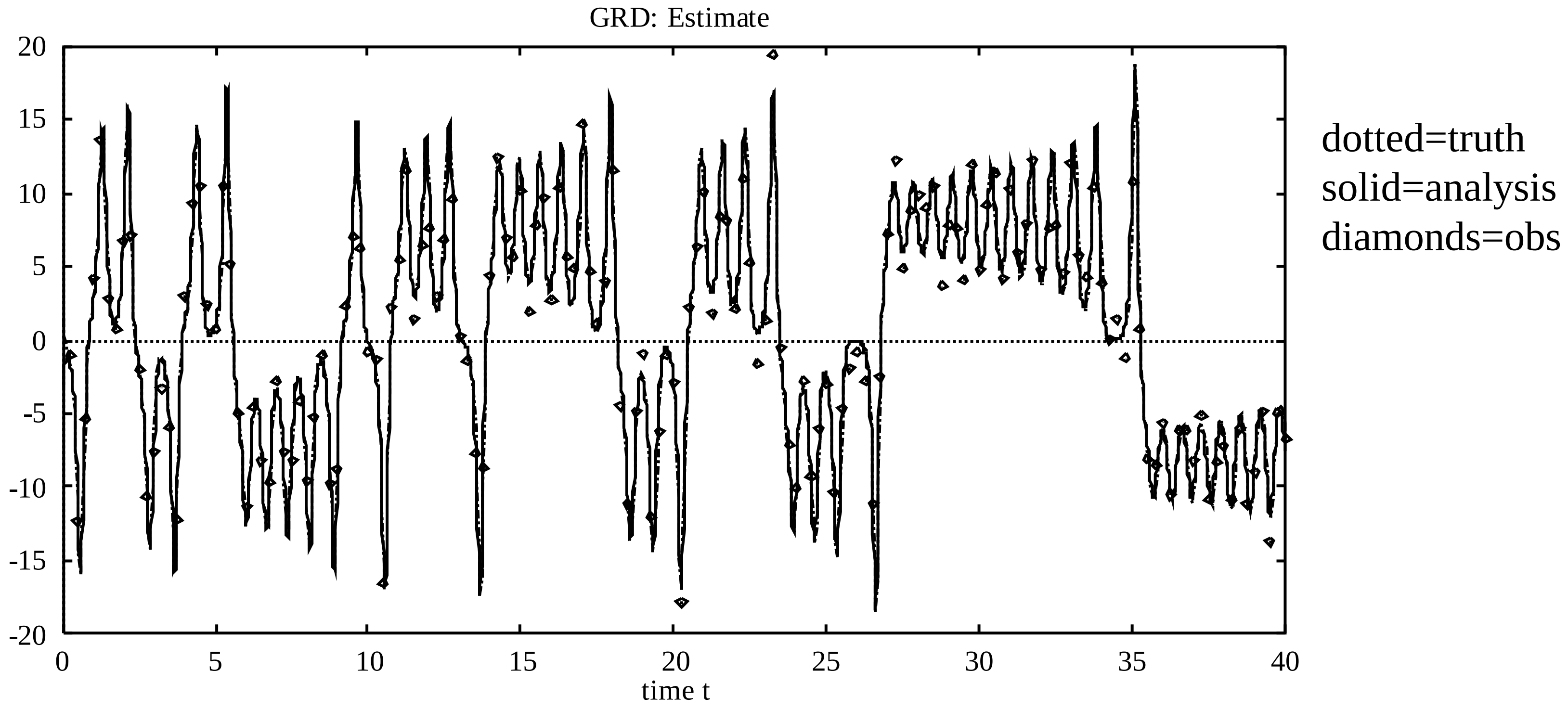


Figure 1: The MSE cost function in the Lorenz model as a function of error in the initial Value of the Y coordinate. The function becomes increasingly pathological as the assimilation period is increased.



Cross-section of the cost function for a random perturbation to the control vector.
$J_o + J_q$ (weak constraint)
J_o (weak constraint)
J_o (strong constraint)
Cost
70000
60000
50000
40000
30000
20000
10000
0
-1
-0.5
0
0.5
1
ε / σ_o

What About... Multiple Minima?



- **From: Evensen (1997).**
  - **Weak constraint 4dVar for the Lorenz 3-variable system.**
  - **~50 orbits of the lobes of the attractor, and 15 lobe transitions.**

Summary

- **Re-Analysis is a <u>smoothing</u> problem, not a filtering <u>problem</u>.**
- **Using observations <u>in the future</u> with respect to the analysis time can reduce analysis error <u>below</u> what is achievable in NWP.**
- **Overlapping the analysis windows improves time-consistency.**
- **Long-window, weak-constraint 4dVar is an efficient algorithm for solving the <u>un-approximated</u> Kalman smoothing equations for large-dimensional systems.**
- **No rank-reduction or other mangling of the covariance matrix is required!**
- **A window length of ~10 days is required for full optimality (but a suboptimal smoother can beat an optimal filter).**
- **Weak constraint 4dVar is suitable for both NWP and re-analysis.**