A variance constraining Kalman filter for data assimilation

Georg Gottwald, Lewis Mitchell, Sebastian Reich*

University of Sydney, *Universität Potsdam

Cambridge, October 12 2010
Data assimilation tries to find the “best estimate” of a current state given noisy observations and an untrustworthy model (model error, chaos).

Analysis:
At time $t_i$, the analysis $z = z_a$ is obtained by incorporating the information of observations and of the background $z_f$.

Forecast:
The analysis $z = z_a$ and its associated variance $P_a$ are then subsequently propagated to the next observation time $t_{i+1}$ to yield $z_f$ and $P_f$ at the next time step. Then a new analysis step can be performed.
Data assimilation produces an *analysis*, i.e. the best estimate, as well as the *covariance of the analysis*, i.e. how much we can trust the analysis. There are several analysis schemes to solve this stochastic optimization problem, differing by how they combine observations and forecasts to produce the analysis:

- Optimal interpolation
- 3D-VAR, 4-DVAR
- Kalman filter

We work within the framework of *ensemble Kalman filters* (Evensen, 1996), which have the following properties:

- optimal filter provided the process is Gaussian like all Kalman filters
- propagate the forecast covariance using the full dynamics (rather than estimating it using a Gaussianity assumption)
Assume an $N$-dimensional dynamical system whose dynamics is given by

$$\dot{z} = f(z)$$

with the state variable $z \in \mathbb{R}^N$ (no model error).

Take observations $z_{\text{obs}}$ at equally spaced discrete observation times $t_n = n\Delta t_{\text{obs}}$ with

- observation operator $H : \mathbb{R}^N \rightarrow \mathbb{R}^p$ with $p \ll N$
- $z_{\text{obs}}(t_i) = Hz(t_i) + r_{\text{obs}}(t_i)$ with observational noise $r_{\text{obs}} \in \mathbb{R}^p$
- We assume $r_{\text{obs}} \sim \mathcal{N}(0, R_{\text{obs}})$ with observational error covariance matrix $R_{\text{obs}} \in \mathbb{R}^{p \times p}$
The Ensemble Transform Kalman Filter (ETKF)

An ensemble (Evensen, 1996) with \( k \) members \( z_k \)

\[
Z = [z_1, z_2, \ldots, z_k] \in \mathbb{R}^{N \times k}
\]

is propagated by the full nonlinear dynamics

\[
\dot{Z} = f(Z), \quad Z(0) = Z_b.
\]

The ensemble is split into its mean

\[
\bar{z} = \frac{1}{k} \sum_{i=1}^{k} z_i = Zw \quad \text{with} \quad w = \frac{1}{k} e \in \mathbb{R}^k \quad \text{and} \quad e = [1, \ldots, 1]^T \in \mathbb{R}^k
\]

and its ensemble deviation matrix

\[
Z' = Z - \bar{z}e^T = ZT \quad \text{with} \quad T = I - we^T \in \mathbb{R}^{k \times k}
\]

Minimise the cost function

\[
S(z) = \frac{1}{2} (z - z_f)^T P_f^{-1} (z - z_f) + \frac{1}{2} (z_{\text{obs}} - Hz)^T R_{obs}^{-1} (z_{\text{obs}} - Hz)
\]
Step 1: Forecast step

\[ Z_f = F(Z_b) \]

\[ P_f = \frac{1}{k-1} Z_f'(t)[Z_f'(t)]^T \]

Remark: \( P_f(t) \) is rank-deficient for \( k < N \) (\( N \sim 10^9 \) and \( k \sim 100 \))
Step 1: Forecast step

\[ Z_f = F(Z_b) \]
\[ P_f = \frac{1}{k-1} Z'_f(t)[Z'_f(t)]^T \]

Remark: \( P_f(t) \) is rank-deficient for \( k < N \) (\( N \sim 10^9 \) and \( k \sim 100 \))

Step 2: Analysis step

\[ \bar{z}_a = \bar{z}_f + K_{obs}(z_{obs} - H\bar{z}_f) \]

where
\[ K_{obs} = P_a H^T R_{obs}^{-1} \]
\[ P_a = \left( P_f^{-1} + H^T R_{obs}^{-1} H \right)^{-1} \]
The Ensemble Transform Kalman Filter (ETKF)

Step 3: Update of the ensemble

The ensemble needs to be consistent with

\[ P_a = \frac{1}{k-1} Z_a' [Z_a']^T \]

Method of ensemble square root filters:

- Ensemble transform Kalman filter (EnTKF) \((Tippett et al 2003)\):
  \[ Z_a' = Z_f' S \] with \( S \in \mathbb{R}^{k \times k} \)

- Ensemble adjustment Kalman filter (EnAKF) \((Anderson 2001)\):
  \[ Z_a' = A Z_f' \] with \( A \in \mathbb{R}^{N \times N} \)
The Ensemble Transform Kalman Filter (ETKF)

Step 3: Update of the ensemble

The ensemble needs to be consistent with

\[ P_a = \frac{1}{k-1} Z'_a [Z'_a]^T \]

Method of ensemble square root filters:

- Ensemble transform Kalman filter (EnTKF) \((Tippett et al 2003)\):
  \( Z'_a = Z'_f S \) with \( S \in \mathbb{R}^{k \times k} \)

- Ensemble adjustment Kalman filter (EnAKF) \((Anderson 2001)\):
  \( Z'_a = A Z'_f \) with \( A \in \mathbb{R}^{N \times N} \)

Step 4: Update of the forecast

Set \( Z_b = Z_a \) to propagate the ensemble forward again with the full dynamics to the next observation time.
A variance constraining Kalman filter

Cambridge, October 12th 2010
A variance constraining Kalman filter

Cambridge, October 12th 2010
A variance constraining Kalman filter
A variance constraining Kalman filter

Cambridge, October 12th 2010
Problems in using ETKF: sparse observational grids

There are many aspects of atmospheric flows which complicate data assimilation. One big problem is that only a very small number of observations – $O(10^3)$ – are available compared to the number of gridpoints for the model – $O(10^8)$ – → large unobserved regions!
Problems in using ETKF: sparse observational grids

There are many aspects of atmospheric flows which complicate data assimilation. One big problem is that only a very small number of observations – $O(10^3)$ – are available compared to the number of gridpoints for the model – $O(10^8)$ – → large unobserved regions!
Problems in using ETKF: sparse observational grids

A variance constraining Kalman filter

Cambridge, October 12th 2010
Problems in using ETKF: sparse observational grids

Our particular perspective here:

Proper (noisy) observations are available for some variables (observables) but not for other unresolved variables, for which only their statistical climatic behaviour such as their variance and their mean is available (pseudo-observables).

Applications are
- sparse observational networks
- when direct observations are not available (mesosphere)
- slow-fast systems

Question:
How can the statistical information available for some data which are otherwise not observable, be effectively incorporated into data assimilation?
Assume that the state space is decomposable according to $\mathbf{z} = (\mathbf{x}, \mathbf{y})$ with $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$ and $n + m = N$.

$x$: observables

$y$: pseudo-observables
Setting

Assume that the state space is decomposable according to $z = (x, y)$ with $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ and $n + m = N$.

$x$: observables

$y$: pseudo-observables

Assume climatic knowledge about the pseudo-observables $y$ (mean $a_{\text{target}}$ and variance $A_{\text{target}}$)

- pseudo-observation operator $h : \mathbb{R}^N \rightarrow \mathbb{R}^l$ with $l \leq m$ ($hh' = I_l$)
- $R_w$ is the unknown error covariance matrix associated with the pseudo-observables
Setting

Assume that the state space is decomposable according to $\mathbf{z} = (\mathbf{x}, \mathbf{y})$ with $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$ and $n + m = N$.

- $\mathbf{x}$: observables
- $\mathbf{y}$: pseudo-observables

Assume climatic knowledge about the pseudo-observables $\mathbf{y}$ (mean $\mathbf{a}_{\text{target}}$ and variance $\mathbf{A}_{\text{target}}$)

- pseudo-observation operator $\mathbf{h} : \mathbb{R}^N \to \mathbb{R}^l$ with $l \leq m$ ($\mathbf{hh}' = \mathbf{I}_l$)
- $\mathbf{R}_w$ is the unknown error covariance matrix associated with the pseudo-observables

Question:

How do we choose/find the error covariance matrix $\mathbf{R}_w$?
Assume that the state space is decomposable according to $z = (x, y)$ with $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ and $n + m = N$.

**Observables** $x$

**Pseudo-observables** $y$

Assume climatic knowledge about the pseudo-observables $y$ (mean $a_{\text{target}}$ and variance $A_{\text{target}}$)

- Pseudo-observation operator $h : \mathbb{R}^N \rightarrow \mathbb{R}^l$ with $l \leq m$ ($hh' = I_l$)
- $R_w$ is the unknown error covariance matrix associated with the pseudo-observables

**Question:**

How do we choose/find the error covariance matrix $R_w$?

**Preliminary answer:**

The naive first guess $R_w = A_{\text{target}}$ is wrong
Equivalence of Kalman filtering and 3DVAR (Lorenc 1986) which aims at minimizing a cost function

\[
S(z) = \frac{1}{2}(z - z_f)^TP_f^{-1}(z - z_f) + \frac{1}{2}(z_{obs} - Hz)^TR_{obs}^{-1}(z_{obs} - Hz) \\
+ \frac{1}{2}(a_{target} - hz)^TR_w^{-1}(a_{target} - hz)
\]

where \( z \) is the state variable at one observation time \( t_i = i\Delta t_{obs} \), \( z_f \) a given background, and \( P_f \) the associated forecast error covariance matrix.

Remark: Compare with the notion of weak constraints in variational data assimilation.
The equation for the critical point with $\nabla_z S(z) = 0$ is readily evaluated to be

$$\left( P_f^{-1} + H^T R_{\text{obs}}^{-1} H + h^T R_w^{-1} h \right) \bar{z}_a = P_f^{-1} \bar{z}_f + H^T R_{\text{obs}}^{-1} z_{\text{obs}} + h^T R_w^{-1} a_{\text{target}}$$

which can be written as

$$\bar{z}_a = \bar{z}_f - K_{\text{obs}} [H \bar{z}_f - z_{\text{obs}}] - K_w [h \bar{z}_f - a_{\text{target}}]$$

where

$$K_{\text{obs}} = P_a H^T R_{\text{obs}}^{-1}$$
$$K_w = P_a h^T R_w^{-1}$$

with the covariance of the analysis

$$P_a = \left( P_f^{-1} + H^T R_{\text{obs}}^{-1} H + h^T R_w^{-1} h \right)^{-1}$$
Constraining the variance of the pseudo-observable $\mathbf{hz}$ is done by requiring

$$h\mathbf{P}_a h^T = A_{\text{target}}$$

Introducing

$$\tilde{\mathbf{B}}^{-1} = \mathbf{P}_f^{-1} + \mathbf{H}^T \mathbf{R}_{\text{obs}}^{-1} \mathbf{H}$$

we obtain by multiplying the constraint with $\mathbf{R}_w^{-1}$ from the right to obtain

$$h \left( \tilde{\mathbf{B}}^{-1} + h^T \mathbf{R}_w^{-1} h \right)^{-1} h^T \mathbf{R}_w^{-1} = A_{\text{target}} \mathbf{R}_w^{-1}$$

which upon using the Sherman-Morrison-Woodbury formula yields the desired equation for $\mathbf{R}_w$

$$\mathbf{R}_w^{-1} = A_{\text{target}}^{-1} - \left( h\tilde{\mathbf{B}} h^T \right)^{-1}$$

Note the typical reciprocal addition formula for variances.
The naive expectation $R_w = A_{\text{target}}$ is true only for $|\{R_{\text{obs}}, P_f\}| \gg |A_{\text{target}}| = |\sigma_{\text{clim}}^2|$
\[ R_w^{-1} = A_{\text{target}}^{-1} - \left( h\tilde{\mathbf{B}}h^T \right)^{-1} \]

The naive expectation \( R_w = A_{\text{target}} \) is true only for
\[ |\{R_{\text{obs}}, P_f\}| \gg |A_{\text{target}}| = |\sigma^2_{\text{clim}}| \]

For sufficiently small background error covariance \( P_f \), the error covariance \( R_w \) is not positive definite ("switch"): 
- Diagonalise \( R_w^{-1} = V\Lambda V^T \)
- Set \( \bar{\Lambda}_{i,i} = \Lambda_{i,i} \) if \( \Lambda_{i,i} > 0 \)
- Set \( \bar{\Lambda}_{i,i} = 0 \) if \( \Lambda_{i,i} < 0 \)
- Use \( \bar{R}_w^{-1} = V\bar{\Lambda}V^T \)
# Summary of VCKF

## Step 1: Forecast step

\[ Z_f = F(Z_b) \]
\[ P_f = \frac{1}{k-1}Z'_f(t)[Z'_f(t)]^T \]

## Step 2: Analysis step

\[ \bar{z}_a = \bar{z}_f + K_{obs}(z_{obs} - H\bar{z}_f) + K_w(a_{target} - h\bar{z}_f) \]

\[ K_{obs} = \frac{P_aH^T R_{obs}^{-1}}{R_{obs}^{-1}} \]
\[ K_w = \frac{P_a h^T R_w^{-1}}{R_w^{-1}} \]
\[ P_a = \left( \frac{P_f^{-1} + H^T R_{obs}^{-1} H + h^T R_w^{-1} h}{R_w^{-1}} \right)^{-1} \]

\[ R_w^{-1} = A_{target}^{-1} - (hB h^T)^{-1} \]

## Step 3: Update of the ensemble

The ensemble needs to be consistent with

\[ P_a = \frac{1}{k-1}Z'_a [Z'_a]^T \]

## Step 4: Update of the forecast

Set \( Z_b = Z_a \) to propagate the ensemble forward again with the full dynamics to the next observation time.
Analytical linear toy model

Consider the system of coupled linear oscillators $z = (x, y)$ with $x \in \mathbb{R}^2$, $y \in \mathbb{R}^2$

$$dz = Mz\,dt - \Gamma z\,dt + S\,dW_t + Cz\,dt$$

with

$$M = \begin{pmatrix} \omega_x J & 0 \\ 0 & \omega_y J \end{pmatrix}, \quad \Gamma = \begin{pmatrix} \gamma_x I & 0 \\ 0 & \gamma_y I \end{pmatrix},$$

$$S = \begin{pmatrix} \sigma_x I & 0 \\ 0 & \sigma_y I \end{pmatrix}, \quad C = \begin{pmatrix} 0 & \lambda J \\ 0 & 0 \end{pmatrix}, \quad J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Introducing the propagator $L(t) = \exp((M - \Gamma + C)t)$, the solution can be obtained using Itô’s formula

$$z(t) = L(t)z_0 + S \int_0^t L(t - s)\,dW_s$$
Analytical linear toy model

We can calculate the mean

\[ m(t) = L(t)z_0 , \]

and covariance

\[ \Sigma(t) = S (2\Gamma - C)^{-1} (I - \exp(-(2\Gamma - C)t)) S^T , \]

where

\[ C = \begin{pmatrix} 0 & \lambda J \\ -\lambda J & 0 \end{pmatrix} . \]

The climatic mean \( m_{\text{clim}} \in \mathbb{R}^4 \) and covariance matrix \( \Sigma_{\text{clim}} \in \mathbb{R}^{4 \times 4} \) are then obtained in the limit of \( t \to \infty \) as

\[ m_{\text{clim}} = \lim_{t \to \infty} m(t) = 0 , \]

and

\[ \Sigma_{\text{clim}} = \lim_{t \to \infty} \Sigma(t) = S (2\Gamma - C)^{-1} S^T . \]

Remark: The coupling has to be sufficiently small with \( \lambda^2 \leq 4\gamma_x\gamma_y \).
Analytical linear toy model

We will investigate the variance constrained Kalman filter for this toy model:

- Under what conditions is $R_w$ positive definite and the variance constraint will be switched on?
- When does the VCKF yield skill improvement compared to the standard ETKF?

1. For an ensemble the covariance of the forecast is calculated by averaging over the ensemble and over realizations of the Brownian motion

$$P_f(t_{i+1}) = L(\Delta t_{\text{obs}}) P_a(t_i) L^T(\Delta t_{\text{obs}}) + \Sigma(\Delta t_{\text{obs}})$$

- Restrict to small observation times $\Delta t_{\text{obs}} \ll 1$
- Then $P_a(t_i) \approx P_f(t_{i+1})$
- Introduce filter inflation $\delta \geq 1$

$$\Delta t_{\text{obs}}(\delta) > \frac{\delta \lambda^2 + 4\gamma_x \gamma_y (1 - \delta)}{2\gamma_x (1 + \gamma_y^2)}$$
Analytical linear toy model

\[ \Delta t_{\text{obs}}(\delta) > \frac{\delta(\lambda^2 - 4\gamma_x \gamma_y) + 4\gamma_x \gamma_y}{2\gamma_x(1 + \gamma_y^2)} \]

Remarks:

- For \( \delta > 1 \) we can have \( \Delta t_{\text{obs}} < 0 \), indicating that the variance constraint will be switched on for all values of \( \Delta t_{\text{obs}} \).

- Because \( 4\gamma_x \gamma_y - \lambda^2 > 0 \) \( \Delta t_{\text{obs}}(\delta = 1) > \Delta t_{\text{obs}}(\delta > 1) \). This is intuitive, because the variance inflation will increase instances with \( |\mathbf{hP}_a \mathbf{h}^T| > \sigma_{\text{clim}}^2 \).

- For \( \lambda \gg 1 \) or \( \gamma_x \ll 1 \), our expression for \( \Delta t_{\text{obs}}(\delta) \) may not be consistent with the assumption of small observation times \( \Delta t_{\text{obs}} \ll 1 \).

- \( \partial \mathbf{R}_w^{-1} / \partial \Delta t_{\text{obs}} > 0 \) at \( \Delta t_{\text{obs}} = 0 \).

- For \( \Delta t_{\text{obs}} \to \infty \) and \( \mathbf{R}_{\text{obs}} \to \infty \), we have \( \mathbf{P}_f \to \Sigma_{\text{clim}} \) and the variance constraint should not be switched on, but in numerical simulations it is? This is a finite ensemble size effect.
Analytical linear toy model

2. When does the VCKF yield skill improvement when compared to the standard ETKF?

Filter skill

\[ \mathcal{E} = \mathbb{E}^t, dW \| \tilde{z}_a(t_i) - z_{\text{truth}}(t_i) \|^2_G \]

\( \mathbb{E}^t \) denotes temporal average over analyzes cycles, and averaging over Brownian paths. The norm \( \| ab \|_G = a^T G b \) can be employed with \( G = I \) for overall skill, \( G = H^T H \) for the observables only, and \( G = h^T h \) for the pseudo-observables only.

\[
\mathcal{E}^{\text{ETKF}} = \mathbb{E}^t \| (I - K_{\text{obs}} H)L(\Delta t_{\text{obs}}) \xi_{t_i-1} \|^2_G + \mathbb{E}^t \| (I - K_{\text{obs}} H)\eta_{t_i} \|^2_G + \mathbb{E}^t \| K_{\text{obs}} r_{\text{obs}} \|^2_G
\]

\[
\mathcal{E}^{\text{VCKF}} = \mathbb{E}^t \| (I - \tilde{K}_{\text{obs}} H)L(\Delta t_{\text{obs}}) \tilde{\xi}_{t_i-1} \|^2_G + \mathbb{E}^t \| (I - \tilde{K}_{\text{obs}} H)\eta_{t_i} \|^2_G + \mathbb{E}^t \| \tilde{K}_{\text{obs}} r_{\text{obs}} \|^2_G
\]

with the mutually independent, normally distributed random variables

\[
\xi_{t_i} = \tilde{z}_a(t_i) - z_{\text{truth}}(t_i) \sim \mathcal{N}(0, P_a(t_i))
\]

\[
\eta_{t_i} = S \int_{t_i-1}^{t_i} L(\Delta t_{\text{obs}} - s) dW_s \sim \mathcal{N}(0, \Sigma(\Delta t_{\text{obs}}))
\]

\[
r_{\text{obs}} \sim \mathcal{N}(0, R_{\text{obs}})
\]
Analytical linear toy model

Skill improvement for the pseudo-observables

\[ S = \mathcal{E}^{ETKF} / \mathcal{E}^{VCKF} \]

Remarks:

- \( S > 1 \) for either \( \gamma_y \to \infty \) or \( \gamma_x \to 0 \)
- Suggests that the skill is controlled by the ratio of the time scales of the observed and the unobserved variables (climate: ✓, regional forecast: ×)
- \( \partial S / \partial R_{obs} > 0 \) at \( R_{obs} = 0 \) (effective slowed down relaxation towards equilibrium of the observed variables)
- \( \partial S / \partial \delta > 0 \)
Consider $\mathbf{z} \in \mathbb{R}^N$ (typically $N = 40$):

$$\frac{dz_i}{dt} = z_{i-1}(z_{i+1} - z_{i-2}) - z_i + F \quad i = 1, \ldots, N$$

$$Z_{i \pm N} = z_i$$

This is a paradigmatic model for the midlatitude atmosphere:

- has forcing $F$
- has linear damping
- non-linear terms conserve the energy $\frac{1}{2} \sum_i ||z_i||^2$
Lorenz-96 model

Consider a latitudinal ring in the midlatitudes with a circumference of roughly 30,000 km. At those latitudes the doubling time is roughly 2.1–2.2 days:

For $N = 40$ and $F = 8$:
- 1 time unit $\approx 5$ days
- distance between observation stations:
  $750 \text{ km} \approx \frac{30,000}{40} \text{ km}$
Lorenz-96 model

Instead of 40 observations
Lorenz-96 model

Observe only every $N_{\text{obs}} = 5$ component

A variance constraining Kalman filter
Cambridge, October 12th 2010
Lorenz-96 model

The pseudo-observables contain the prior climatic knowledge:
\[ \mathbf{a}_{\text{target}} = \mu_{\text{clim}} \quad \text{and} \quad \mathbf{A}_{\text{target}} = \sigma_{\text{clim}}^2 \mathbf{I} \]
with \( \mu_{\text{clim}} = 2.34 \) and \( \sigma_{\text{clim}} = 3.6 \)
measured in a long time trajectory
Lorenz-96 model

The pseudo-observables contain the prior climatic knowledge: 
\[ a_{\text{target}} = \mu_{\text{clim}} \] and \[ A_{\text{target}} = \sigma^2_{\text{clim}} I \] with \[ \mu_{\text{clim}} = 2.34 \] and \[ \sigma_{\text{clim}} = 3.6 \] measured in a long time trajectory.

A variance constraining Kalman filter
Cambridge, October 12th 2010
Lorenz-96 model

Quantify the skill improvement by the r.m.s error

$$\mathcal{E} = \sqrt{\left\langle \frac{1}{TD} \sum_{l=1}^{T} \| \bar{z}_a(l\Delta t_{obs}) - z_{\text{truth}}(l\Delta t_{obs}) \|^2 \right\rangle}$$

$$R_{obs} = (0.25\sigma_{clim})^2 I$$

Best performance of VCKF over ETKF for:

- small $\Delta t_{obs}$
- $N_{obs} = 4$
Lorenz-96 model

How is the skill improvement distributed over observable pseudo-observables

A variance constraining Kalman filter  Cambridge, October 12th 2010
Lorenz-96 model

Dependency on observational noise level $R_{\text{obs}} = (\eta \sigma_{\text{clim}})^2 I$, $N_{\text{obs}} = 4$

\[ \Delta t_{\text{obs}} = 0.025 \quad (1 \text{ hour}) \]
\[ \Delta t_{\text{obs}} = 0.05 \quad (2 \text{ hours}) \]
\[ \Delta t_{\text{obs}} = 0.25 \quad (5 \text{ hours}) \]
Lorenz-96 model

There is an order of magnitude difference between the RMS errors for the observables and the pseudo-observables for large $N_{\text{obs}}$. This suggests that the information of the observed variables does not travel too far away from the observational sites.

Total RMS error for each site $i$, $i = 1, 2, \cdots, 40$ if only site $i^\star = 21$ is observed.

Remark: The advective time scale of the Lorenz-96 system is much smaller than $\Delta t_{\text{obs}}$ which explains why the skill is not equally distributed over the sites, and why, especially for large values of $N_{\text{obs}}$ we observe a big difference between the site-averaged skills of the observed and unobserved variables.
Lorenz-96 model: Filter divergence and blow-up (*Ott et al.* (2004); *Harlim & Majda* (2009)) for small $R_{\text{obs}}$

<table>
<thead>
<tr>
<th>$N_{\text{obs}}$</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.14</td>
<td>0.02</td>
<td>0.04</td>
<td>0.0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.40</td>
<td>0.22</td>
<td>0.03</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.67</td>
<td>0.29</td>
<td>0.04</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.98</td>
<td>0.49</td>
<td>0.11</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.96</td>
<td>0.64</td>
<td>0.15</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.76</td>
<td>0.77</td>
<td>0.44</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.32</td>
<td>0.83</td>
<td>0.58</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.89</td>
<td>0.67</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>0.49</td>
<td>0.82</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.01</td>
<td>0.58</td>
<td>0.67</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>0.82</td>
<td>0.82</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\Delta \tau_{\text{obs}}$</th>
<th>3 h</th>
<th>6 h</th>
<th>9 h</th>
<th>12 h</th>
<th>15 h</th>
<th>18 h</th>
<th>21 h</th>
<th>24 h</th>
<th>27 h</th>
<th>30 h</th>
</tr>
</thead>
<tbody>
<tr>
<td>ETKF</td>
<td>0.025</td>
<td>0.05</td>
<td>0.075</td>
<td>0.1</td>
<td>0.125</td>
<td>0.15</td>
<td>0.175</td>
<td>0.2</td>
<td>0.225</td>
<td>0.25</td>
</tr>
<tr>
<td>VCKF</td>
<td>0.01</td>
<td>0.24</td>
<td>0.11</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.42</td>
<td>0.36</td>
<td>0.10</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.11</td>
<td>0.22</td>
<td>0.02</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.01</td>
<td>0.06</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.02</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

A variance constraining Kalman filter  
Cambridge, October 12th 2010
Chaotic slow-fast system

Skew product system of a chaotically forced bistable system (Givon et al., Nonlinearity 17 (2004))

\[
\frac{dx}{dt} = x - x^3 + \frac{4}{90\varepsilon} y_2
\]

\[
\begin{align*}
\frac{dy_1}{dt} &= \frac{10}{\varepsilon^2} (y_2 - y_1) \\
\frac{dy_2}{dt} &= \frac{1}{\varepsilon^2} (28y_1 - y_2 - y_1y_3) \\
\frac{dy_3}{dt} &= \frac{1}{\varepsilon^2} (y_1y_2 - \frac{8}{3} y_3)
\end{align*}
\]

Homogenization (Kurths, 1973)

\[
dx = (x - x^3) dt + \sigma dW
\]
Estimating drift and diffusion coefficients

- Create a long trajectory
Estimating drift and diffusion coefficients

- Create a long trajectory
  - Coarse grain the dynamics with sample time $\Delta t$. We found that roughly 4 fast oscillations suffices
Create a long trajectory

Coarse grain the dynamics with sample time $\Delta t$. We found that roughly 4 fast oscillations suffices.

Project the dynamics onto the $q_1$-$q_2$ subspace
Estimating drift and diffusion coefficients

- Create a long trajectory
  Coarse grain the dynamics with sample time $\Delta t$. We found that roughly 4 fast oscillations suffices

- Project the dynamics onto the $q_1$-$q_2$ subspace
  Coarse grain the phase space by partitioning the $q_1$-$q_2$ in $Q_1$-$Q_2$-bins of side length $\Delta Q$
Estimating drift and diffusion coefficients

The conditional expectation values allow us to determine the parameters of a stochastic differential equation of the form

\[ dq = G(q) \Delta t + \sigma dW \]

according to

\[
D_i(Q) = \frac{1}{\Delta t} \mathbb{E} \left[ (q_i^{n+1} - q_i^n) \right]_{q_i^n \in (Q_i, Q_i+\Delta b)} \\
\approx \frac{1}{\Delta t} \mathbb{E} \left[ dq_i \right]_{q_i^n \in (Q_i, Q_i+\Delta b)} \\
\approx G_i(Q)
\]

\[
S_{ij}(Q) = \frac{1}{\Delta t} \mathbb{E} \left[ (q_i^{n+1} - q_i^n)(q_j^{n+1} - q_j^n) \right]_{q_i^n, j^n \in (Q_i, Q_i+\Delta b)} \\
\approx \frac{1}{\Delta t} \mathbb{E} \left[ dq_i dq_j \right]_{q_i^n, j^n \in (Q_i, Q_i+\Delta b)} \\
\approx (\sigma \sigma^T)_{ij} + \mathcal{O}(\Delta t)
\]
Chaotic slow-fast system

A variance constraining Kalman filter

Cambridge, October 12th 2010
Chaotic slow-fast system

\[ \tau_{\text{autocorr}} \approx 200 \]
\[ \tau_{\text{FSLyap}} \approx 120 \]
\[ \tau_{\text{exit}} = 151 \]
\[ \tau_{\text{fast}} = 0.01 \]
Summary and outlook

We have here

- derived a variance constraining Kalman filter
- applied this filter to a sparse observational grid
- shown that the filter
  - has better skill than ETKF for small ($\leq 6h$) observation times
  - has better skill for observed and pseudo-observables
  - is stabilizing and avoids filter divergencies such as blow-up
  - is robust to incomplete knowledge of the climatic mean and variance

We will

- investigate blow-up further