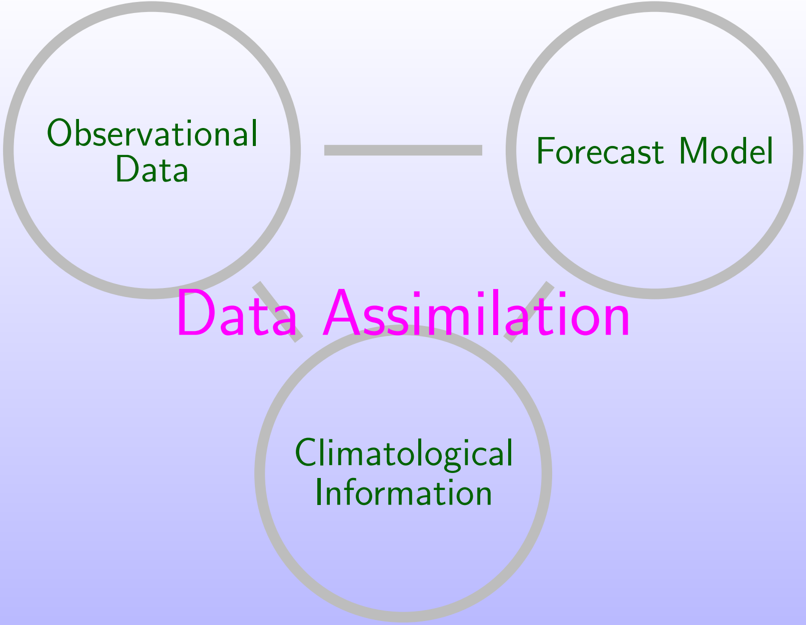
A variance limiting Kalman filter for data assimilation: I. Sparse observational grids II. Model error

Georg Gottwald, Lewis Mitchell, Sebastian Reich*

University of Sydney, *Universität Potsdam

Durham, August 6th, 2011

Data Assimilation

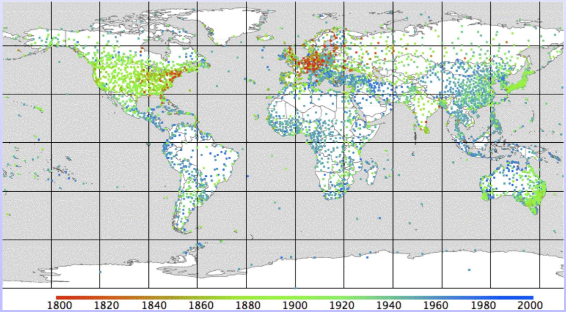

Observational Data
Forecast Model
Data Assimilation
Climatological Information

I. Sparse observational grids

Typically only a very small number of observations – $\mathcal{O}(10^4)/\mathcal{O}(10^7)$ – are available compared to the number of gridpoints for the model – $\mathcal{O}(10^9)$ –
$\longrightarrow$ large unobserved regions !

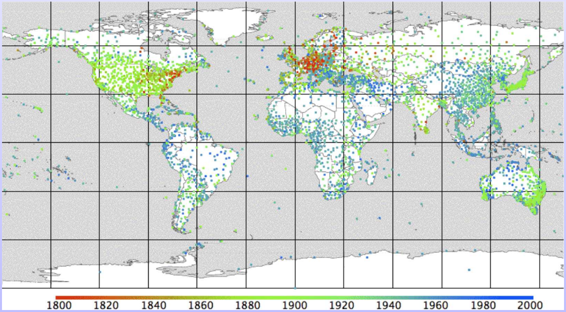
I. Sparse observational grids

Typically only a very small number of observations – $\mathcal{O}(10^4)/\mathcal{O}(10^7)$ – are available compared to the number of gridpoints for the model – $\mathcal{O}(10^9)$ – $\longrightarrow$ large unobserved regions !



I. Sparse observational grids

Typically only a very small number of observations – $\mathcal{O}(10^4)/\mathcal{O}(10^7)$ – are available compared to the number of gridpoints for the model – $\mathcal{O}(10^9)$ – $\longrightarrow$ large unobserved regions !



Sparse observational grids

What is the effect of the sparsity of observations?

- The obvious: We don't have much information
- Overestimation of error covariances (exacerbated by finite ensemble sizes) (*Whitaker et al. 2009*)
- The subtle: We create spurious correlations and unbalanced flow (*Kepert 2009*)

Sparse observational grids

What is the effect of the sparsity of observations?

- The obvious: We don't have much information
- Overestimation of error covariances (exacerbated by finite ensemble sizes) (*Whitaker et al. 2009*)
- The subtle: We create spurious correlations and unbalanced flow (*Kepert 2009*)

Applications are

- sparse observational networks
- re-analysis of climate
- when direct observations are not available (mesosphere)
- slow-fast systems

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).

Question:

How can the statistical information available for some data which are otherwise not observable, be effectively incorporated into data assimilation to control overestimation?

Setting

Assume an N-dimensional dynamical system whose dynamics is given by

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

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

Setting

Assume an N-dimensional dynamical system whose dynamics is given by

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

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

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

observables

Observations $\mathbf{x}_{\rm obs}$ at observation times $t_n = n\Delta t_{\rm obs}$

- observation operator $\mathbf{H} : \mathbb{R}^N \to \mathbb{R}^n$
- $\mathbf{x}_{\rm obs}(t_i) = \mathbf{H}\mathbf{z}(t_i) + \mathbf{r}_{\rm obs}(t_i)$ with observational noise $\mathbf{r}_{\rm obs}$
- $\mathbf{r}_{\rm obs} \sim \mathcal{N}(0, \mathbf{R}_{\rm obs})$ with error covariance matrix $\mathbf{R}_{\rm obs}$

observables

Observations $\mathbf{x}_{\text{obs}}$ at observation times $t_n = n\Delta t_{\text{obs}}$

- observation operator $\mathbf{H} : \mathbb{R}^N \rightarrow \mathbb{R}^n$
- $\mathbf{x}_{\text{obs}}(t_i) = \mathbf{Hz}(t_i) + \mathbf{r}_{\text{obs}}(t_i)$ with observational noise $\mathbf{r}_{\text{obs}}$
- $\mathbf{r}_{\text{obs}} \sim \mathcal{N}(0, \mathbf{R}_{\text{obs}})$ with error covariance matrix $\mathbf{R}_{\text{obs}}$

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 $\mathbf{h} : \mathbb{R}^N \rightarrow \mathbb{R}^m$
- $\mathbf{R}_{\text{w}}$ is the unknown error covariance matrix associated with the pseudo-observables

observables

Observations $\mathbf{x}_{\text{obs}}$ at observation times $t_n = n\Delta t_{\text{obs}}$

- observation operator $\mathbf{H} : \mathbb{R}^N \to \mathbb{R}^n$
- $\mathbf{x}_{\text{obs}}(t_i) = \mathbf{Hz}(t_i) + \mathbf{r}_{\text{obs}}(t_i)$ with observational noise $\mathbf{r}_{\text{obs}}$
- $\mathbf{r}_{\text{obs}} \sim \mathcal{N}(0, \mathbf{R}_{\text{obs}})$ with error covariance matrix $\mathbf{R}_{\text{obs}}$

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 $\mathbf{h} : \mathbb{R}^N \to \mathbb{R}^m$
- $\mathbf{R}_{\text{w}}$ is the unknown error covariance matrix associated with the pseudo-observables

Question:

How do we choose/find the error covariance matrix $\mathbf{R}_{\text{w}}$?

observables

Observations $\mathbf{x}_{\text{obs}}$ at observation times $t_n = n\Delta t_{\text{obs}}$

- observation operator $\mathbf{H} : \mathbb{R}^N \to \mathbb{R}^n$
- $\mathbf{x}_{\text{obs}}(t_i) = \mathbf{Hz}(t_i) + \mathbf{r}_{\text{obs}}(t_i)$ with observational noise $\mathbf{r}_{\text{obs}}$
- $\mathbf{r}_{\text{obs}} \sim \mathcal{N}(0, \mathbf{R}_{\text{obs}})$ with error covariance matrix $\mathbf{R}_{\text{obs}}$

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 $\mathbf{h} : \mathbb{R}^N \to \mathbb{R}^m$
- $\mathbf{R}_{\text{w}}$ is the unknown error covariance matrix associated with the pseudo-observables

Question:

How do we choose/find the error covariance matrix $\mathbf{R}_{\text{w}}$?
(The naive first guess $\mathbf{R}_{\text{w}} = \mathbf{A}_{\text{target}}$ is wrong)

The Variance Limiting Kalman Filter (VLKF)

An ensemble (Evensen, 1996) with k members $\mathbf{z}_k$

$$\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_k] \in \mathbb{R}^{N \times k}$$

is propagated by the full nonlinear dynamics

$$\dot{\mathbf{Z}} = F(\mathbf{Z}), \qquad \mathbf{Z}(0) = \mathbf{Z}_b \ .$$

The ensemble is split into its mean $\bar{\mathbf{z}}$ and its ensemble deviation matrix $\mathbf{Z}'$

Step 1: Forecast step

$$\mathbf{Z}_f = F(\mathbf{Z}_b)$$

$$\mathbf{P}_f = \frac{1}{k-1} \mathbf{Z}'_f(t) [\mathbf{Z}'_f(t)]^T$$

Remark: $\mathbf{P}_f(t)$ is rank-deficient for $k < N$ ($N \sim 10^9$ and $k \sim 100$)

The Variance Limiting Kalman Filter (VLKF)

Step 2: Analysis step:

Minimise the cost function

$$S(\mathbf{z}) = \frac{1}{2}(\mathbf{z}-\mathbf{z}_f)^T \mathbf{P}_f^{-1}(\mathbf{z}-\mathbf{z}_f) + \frac{1}{2}(\mathbf{x}_{\text{obs}} - \mathbf{H}\mathbf{z})^T \mathbf{R}_{\text{obs}}^{-1}(\mathbf{x}_{\text{obs}} - \mathbf{H}\mathbf{z})$$
$$+ \ \frac{1}{2}(\mathbf{a}_{\text{target}} - \mathbf{h}\mathbf{z})^T \mathbf{R}_{\text{w}}^{-1}(\mathbf{a}_{\text{target}} - \mathbf{h}\mathbf{z})$$

The Variance Limiting Kalman Filter (VLKF)

Step 2: Analysis step:

Minimise the cost function

$$S(\mathbf{z}) = \frac{1}{2}(\mathbf{z}-\mathbf{z}_f)^T \mathbf{P}_f^{-1}(\mathbf{z}-\mathbf{z}_f) + \frac{1}{2}(\mathbf{x}_{\rm obs}-\mathbf{H}\mathbf{z})^T \mathbf{R}_{\rm obs}^{-1}(\mathbf{x}_{\rm obs}-\mathbf{H}\mathbf{z}) + \frac{1}{2}(\mathbf{a}_{\rm target}-\mathbf{h}\mathbf{z})^T \mathbf{R}_{\rm w}^{-1}(\mathbf{a}_{\rm target}-\mathbf{h}\mathbf{z})$$

$$\bar{\mathbf{z}}_a = \bar{\mathbf{z}}_f - \mathbf{K}_{\rm obs}\left[\mathbf{H}\bar{\mathbf{z}}_f - \mathbf{x}_{\rm obs}\right] - \mathbf{K}_{\rm w}\left[\mathbf{h}\bar{\mathbf{z}}_f - \mathbf{a}_{\rm target}\right]$$

where $\mathbf{K}_{\rm obs} = \mathbf{P}_a\mathbf{H}^T\mathbf{R}_{\rm obs}^{-1}$

$$\mathbf{K}_{\rm w} = \mathbf{P}_a\mathbf{h}^T\mathbf{R}_{\rm w}^{-1}$$

with the covariance of the analysis

$$\mathbf{P}_a = \left(\mathbf{P}_f^{-1} + \mathbf{H}^T\mathbf{R}_{\rm obs}^{-1}\mathbf{H} + \mathbf{h}^T\mathbf{R}_{\rm w}^{-1}\mathbf{h}\right)^{-1}$$

The Variance Limiting Kalman Filter (VLKF)

Step 2: Analysis step

Constraining the variance of the pseudo-observable $\mathbf{hz}$ is done by requiring

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

Introducing ${\mathcal{P}_{\mathbf{a}}}^{-1} = \mathbf{P}_f^{-1} + \mathbf{H}^T\mathbf{R}_{\text{obs}}^{-1}\mathbf{H}$, we obtain

$$\mathbf{R}_{\text{w}}^{-1} = \mathbf{A}_{\text{target}}^{-1} - \left(\mathbf{h}\mathcal{P}_{\mathbf{a}}\mathbf{h}^T\right)^{-1}$$

The Variance Limiting Kalman Filter (VLKF)

Step 2: Analysis step

Constraining the variance of the pseudo-observable $\mathbf{hz}$ is done by requiring

$$\mathbf{h}\mathbf{P}_a\mathbf{h}^T = \mathbf{A}_{\mathrm{target}}$$

Introducing ${\mathcal{P}_{\mathbf{a}}}^{-1} = \mathbf{P}_f^{-1} + \mathbf{H}^T\mathbf{R}_{\mathrm{obs}}^{-1}\mathbf{H}$, we obtain

$$\mathbf{R}_{\mathrm{w}}^{-1} = \mathbf{A}_{\mathrm{target}}^{-1} - \left(\mathbf{h}\mathcal{P}_{\mathbf{a}}\mathbf{h}^T\right)^{-1}$$

- The naive expectation $\mathbf{R}_{\mathrm{w}} = \mathbf{A}_{\mathrm{target}}$ is true only for $|\{\mathbf{R}_{\mathrm{obs}}, \mathbf{P}_f\}| \gg |\mathbf{A}_{\mathrm{target}}|$

The Variance Limiting Kalman Filter (VLKF)

Step 2: Analysis step

Constraining the variance of the pseudo-observable $\mathbf{hz}$ is done by requiring

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

Introducing ${\mathcal{P}_{\mathbf{a}}}^{-1} = \mathbf{P}_f^{-1} + \mathbf{H}^T\mathbf{R}_{\text{obs}}^{-1}\mathbf{H}$, we obtain

$$\mathbf{R}_{\text{w}}^{-1} = \mathbf{A}_{\text{target}}^{-1} - \left(\mathbf{h}\mathcal{P}_{\mathbf{a}}\mathbf{h}^T\right)^{-1}$$

- The naive expectation $\mathbf{R}_{\text{w}} = \mathbf{A}_{\text{target}}$ is true only for $|\{\mathbf{R}_{\text{obs}}, \mathbf{P}_f\}| \gg |\mathbf{A}_{\text{target}}|$
- For sufficiently small background error covariance $\mathbf{P}_f$, the error covariance $\mathbf{R}_{\text{w}}$ is not positive definite ("switch"): Update only *overestimated* eigendirections with $|\mathbf{h}\mathcal{P}_{\mathbf{a}}\mathbf{h}^T| > |\mathbf{A}_{\text{target}}|$

The Variance Limiting Kalman Filter (VLKF)

Step 3: Update of the ensemble

The ensemble needs to be consistent with

$$\mathbf{P}_a = \frac{1}{k-1}\mathbf{Z}'_a\left[\mathbf{Z}'_a\right]^T$$

Method of ensemble square root filters:

- Ensemble transform Kalman filter (EnTKF) (*Tippett et al 2003*): $\mathbf{Z}'_a = \mathbf{Z}'_f\mathbf{S}$ with $\mathbf{S} \in \mathbb{R}^{k\times k}$
- Ensemble adjustment Kalman filter (EnAKF) (*Anderson 2001*): $\mathbf{Z}'_a = \mathbf{A}\mathbf{Z}'_f$ with $\mathbf{A} \in \mathbb{R}^{N\times N}$

The Variance Limiting Kalman Filter (VLKF)

Step 3: Update of the ensemble

The ensemble needs to be consistent with

$$\mathbf{P}_a = \frac{1}{k-1}\mathbf{Z}'_a \left[\mathbf{Z}'_a\right]^T$$

Method of ensemble square root filters:

- Ensemble transform Kalman filter (EnTKF) (*Tippett et al 2003*): $\mathbf{Z}'_a = \mathbf{Z}'_f \mathbf{S}$ with $\mathbf{S} \in \mathbb{R}^{k\times k}$
- Ensemble adjustment Kalman filter (EnAKF) (*Anderson 2001*): $\mathbf{Z}'_a = \mathbf{A}\mathbf{Z}'_f$ with $\mathbf{A} \in \mathbb{R}^{N\times N}$

Step 4: Update of the forecast

Set $\mathbf{Z}_b = \mathbf{Z}_a$ to propagate the ensemble forward again with the full dynamics to the next observation time

Summary of VLKF

Step 1: Forecast step

$$\mathbf{Z}_f = F(\mathbf{Z}_b)$$

$$\mathbf{P}_f = \frac{1}{k-1}\mathbf{Z}'_f(t)[\mathbf{Z}'_f(t)]^T$$

Step 2: Analysis step

$$\bar{\mathbf{z}}_a = \bar{\mathbf{z}}_f + \mathbf{K}_{\text{obs}}(\mathbf{x}_{\text{obs}} - \mathbf{H}\bar{\mathbf{z}}_f) + \mathbf{K}_{\text{w}}(\mathbf{a}_{\text{target}} - \mathbf{h}\bar{\mathbf{z}}_f)$$

$$\mathbf{K}_{\text{obs}} = \mathbf{P}_a\mathbf{H}^T\mathbf{R}_{\text{obs}}^{-1}\,, \quad \mathbf{K}_{\text{w}} = \mathbf{P}_a\mathbf{h}^T\mathbf{R}_{\text{w}}^{-1}\,, \quad \mathbf{P}_a = \left(\mathbf{P}_f^{-1} + \mathbf{H}^T\mathbf{R}_{\text{obs}}^{-1}\mathbf{H} + \mathbf{h}^T\mathbf{R}_{\text{w}}^{-1}\mathbf{h}\right)^{-1}$$

$$\mathbf{R}_{\text{w}}^{-1} = \mathbf{A}_{\text{target}}^{-1} - (\mathbf{h}\mathcal{P}_{\mathbf{a}}\mathbf{h}^T)^{-1}$$

Step 3: Update of the ensemble

The ensemble needs to be consistent with

$$\mathbf{P}_a = \frac{1}{k-1}\mathbf{Z}'_a\left[\mathbf{Z}'_a\right]^T$$

Step 4: Update of the forecast

Set $\mathbf{Z}_b = \mathbf{Z}_a$ 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 $\mathbf{z} = (\mathbf{x}, \mathbf{y})$ with $\mathbf{x} \in \mathbb{R}^2$, $\mathbf{y} \in \mathbb{R}^2$

$$\boxed{d\mathbf{z} = \boldsymbol{M}\mathbf{z}\,dt - \boldsymbol{\Gamma}\mathbf{z}\,dt + \boldsymbol{S}\,d\boldsymbol{W}_t + \boldsymbol{C}\mathbf{z}\,dt}$$

with

$$\boldsymbol{M} = \begin{pmatrix} \omega_x \boldsymbol{J} & \mathbf{0} \\ \mathbf{0} & \omega_y \boldsymbol{J} \end{pmatrix} \qquad \boldsymbol{\Gamma} = \begin{pmatrix} \gamma_x \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \gamma_y \mathbf{I} \end{pmatrix}$$

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

Introducing the propagator $\boldsymbol{L}(t) = \exp\left((\boldsymbol{M} - \boldsymbol{\Gamma} + \boldsymbol{C})t\right)$, the solution can be obtained using Itô's formula

$$\mathbf{z}(t) = \boldsymbol{L}(t)\mathbf{z}_0 + \boldsymbol{S}\int_0^t \boldsymbol{L}(t-s)\,d\boldsymbol{W}_s$$

Analytical linear toy model

We can calculate the mean

$$\boldsymbol{m}(t) = \boldsymbol{L}(t)\mathbf{z}_0 \ ,$$

and covariance

$$\boldsymbol{\Sigma}(t) = \boldsymbol{S}\left(2\boldsymbol{\Gamma} - \boldsymbol{\mathcal{C}}\right)^{-1}\left(\mathbf{I} - \exp\left(-\left(2\boldsymbol{\Gamma} - \boldsymbol{\mathcal{C}}\right)t\right)\right)\boldsymbol{S}^T \ ,$$

where

$$\boldsymbol{\mathcal{C}} = \begin{pmatrix} \mathbf{0} & \lambda \boldsymbol{J} \\ -\lambda \boldsymbol{J} & \mathbf{0} \end{pmatrix} .$$

The climatic mean $\boldsymbol{m}_{\rm clim} \in \mathbb{R}^4$ and covariance matrix $\boldsymbol{\Sigma}_{\rm clim} \in \mathbb{R}^{4\times 4}$ are then obtained in the limit of $t \to \infty$ as

$$\boldsymbol{m}_{\rm clim} = \lim_{t\to\infty} \boldsymbol{m}(t) = 0 \ ,$$

and

$$\boldsymbol{\Sigma}_{\rm clim} = \lim_{t\to\infty} \boldsymbol{\Sigma}(t) = \boldsymbol{S}\left(2\boldsymbol{\Gamma} - \boldsymbol{\mathcal{C}}\right)^{-1}\boldsymbol{S}^T \ .$$

Remark: The coupling has to be sufficiently small with $\lambda^2 < 4\gamma_x\gamma_y$.

Analytical linear toy model

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

- Under what conditions is $\mathbf{R}_{\mathrm{w}}$ positive definite and the variance constraint will be switched on?
- When does the VLKF yield skill improvement compared to the standard ETKF?

Analytical linear toy model

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

- Under what conditions is $\mathbf{R}_{\rm w}$ positive definite and the variance constraint will be switched on?
- When does the VLKF yield skill improvement compared to the standard ETKF?

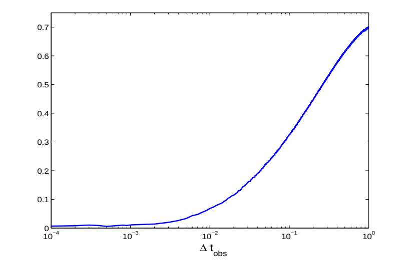
1. For an ensemble the covariance of the forecast is calculated by averaging over realizations of the Brownian motion and over the ensemble

$$\mathbf{P}_f(t_{i+1}) = \boldsymbol{L}(\Delta t_{\rm obs})\,\mathbf{P}_a(t_i)\,\boldsymbol{L}^T(\Delta t_{\rm obs}) + \boldsymbol{\Sigma}(\Delta t_{\rm obs})$$

- Introduce *filter inflation* $\delta \geq 1$
- Restrict to small observation times $\Delta t_{\rm obs} \ll 1$
- Then $\mathbf{P}_a(t_i) \approx \mathbf{P}_f(t_{i+1})$

Analytical linear toy model

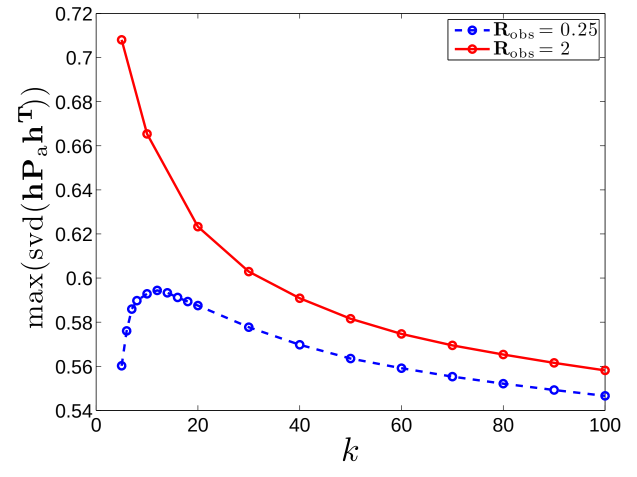
$$\Delta t_{\mathrm{obs}}(\delta) > \frac{\delta(\lambda^2 - 4\gamma_x\gamma_y) + 4\gamma_x\gamma_y}{2\gamma_x(1+\gamma_y^2)}$$



- For $\delta > 1$ we can have $\Delta t_{\mathrm{obs}} < 0 \quad (4\gamma_x\gamma_y - \lambda^2 > 0)$
- $\Delta t_{\mathrm{obs}}(\delta = 1) > \Delta t_{\mathrm{obs}}(\delta > 1)$.
- $\partial \mathbf{R}_{\mathrm{w}}^{-1}/\partial \Delta t_{\mathrm{obs}} > 0$ at $\Delta t_{\mathrm{obs}} = 0$.
- For $\lambda \gg 1$ or $\gamma_x \ll 1$, our expression for $\Delta t_{\mathrm{obs}}(\delta)$ may not be consistent with the assumption of small observation times $\Delta t_{\mathrm{obs}} \ll 1$
- For $\Delta t_{\mathrm{obs}} \to \infty$ and large $\mathbf{R}_{\mathrm{obs}}$, we have $\mathbf{P}_f \to \mathbf{\Sigma}_{\mathrm{clim}}$ and the variance constraint should not be switched on, but in numerical simulations it is?

Analytical linear toy model

Finite size effect (no inflation):



Analytical linear toy model

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

Filter skill

$$\mathcal{E} = \mathbb{E}^{t,dW} \|\bar{\mathbf{z}}_a(t_i) - \mathbf{z}_{\text{truth}}(t_i)\|^2_{\mathbf{G}}$$

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

$$\begin{aligned}
\mathcal{E}^{\text{ETKF}} &= \mathbb{E}^t \|(\mathbf{I} - \mathbf{K}_{\text{obs}}\mathbf{H})\boldsymbol{L}(\Delta t_{\text{obs}})\, \boldsymbol{\xi}_{t_{i-1}}\|^2_{\mathbf{G}} + \mathbb{E}^t \|(\mathbf{I} - \mathbf{K}_{\text{obs}}\mathbf{H})\boldsymbol{\eta}_{t_i}\|^2_{\mathbf{G}} + \mathbb{E}^t \|\mathbf{K}_{\text{obs}}\mathbf{r}_{\text{obs}}\|^2_{\mathbf{G}} \\
\mathcal{E}^{\text{VLKF}} &= \mathbb{E}^t \|(\mathbf{I} - \tilde{\mathbf{K}}_{\text{obs}}\mathbf{H})\boldsymbol{L}(\Delta t_{\text{obs}})\, \tilde{\boldsymbol{\xi}}_{t_{i-1}}\|^2_{\mathbf{G}} + \mathbb{E}^t \|(\mathbf{I} - \tilde{\mathbf{K}}_{\text{obs}}\mathbf{H})\boldsymbol{\eta}_{t_i}\|^2_{\mathbf{G}} + \mathbb{E}^t \|\tilde{\mathbf{K}}_{\text{obs}}\mathbf{r}_{\text{obs}}\|^2_{\mathbf{G}}
\end{aligned}$$

with the mutually independent, normally distributed random variables

$$\begin{aligned}
\boldsymbol{\xi}_{t_i} &= \bar{\mathbf{z}}_a(t_i) - \mathbf{z}_{\text{truth}}(t_i) \sim \mathcal{N}(0, \mathbf{P}_a(t_i)) \\
\boldsymbol{\eta}_{t_i} &= \boldsymbol{S} \int_{t_{i-1}}^{t_i} \boldsymbol{L}(\Delta t_{\text{obs}} - s)\, d\boldsymbol{W}_s \sim \mathcal{N}(0, \boldsymbol{\Sigma}(\Delta t_{\text{obs}})) \\
\mathbf{r}_{\text{obs}} &\sim \mathcal{N}(0, \mathbf{R}_{\text{obs}})
\end{aligned}$$

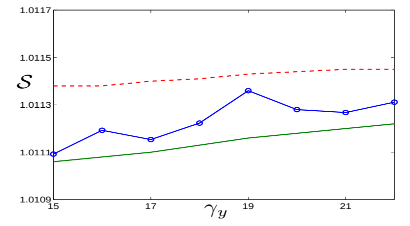
Analytical linear toy model

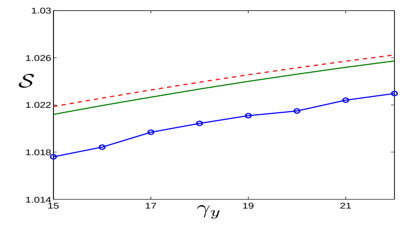
Skill improvement for the pseudo-observables

$$\mathcal{S} = \mathcal{E}^{\mathrm{ETKF}} / \mathcal{E}^{\mathrm{VLKF}}$$

Remarks:

- $\mathcal{S} > 1$ 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
- $\partial \mathcal{S} / \partial R_{\mathrm{obs}} > 0$ at $R_{\mathrm{obs}} = 0$ (effective slowed down relaxation towards equilibrium of the observed variables)
- $\partial \mathcal{S} / \partial \delta > 0$





Lorenz-96 model

Consider $\mathbf{z} \in \mathbb{R}^N$ (typically $N = 40$):

$$\boxed{\begin{aligned} &\frac{dz_i}{dt} = z_{i-1}\left(z_{i+1} - z_{i-2}\right) - z_i + F \qquad i = 1, \cdots, N \\ &z_{i\pm N} = z_i \end{aligned}}$$

This is a paradigmatic model for the midlatitude atmosphere:

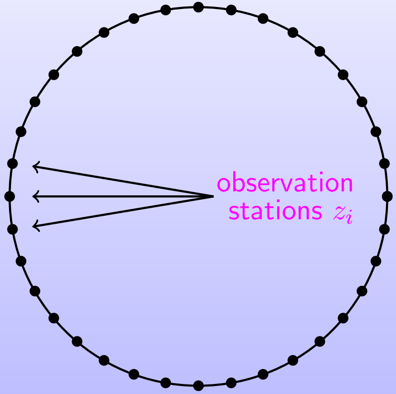
- has forcing F
- has linear damping
- non-linear terms conserve the energy $\frac{1}{2}\sum_i ||z_i||^2$

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:

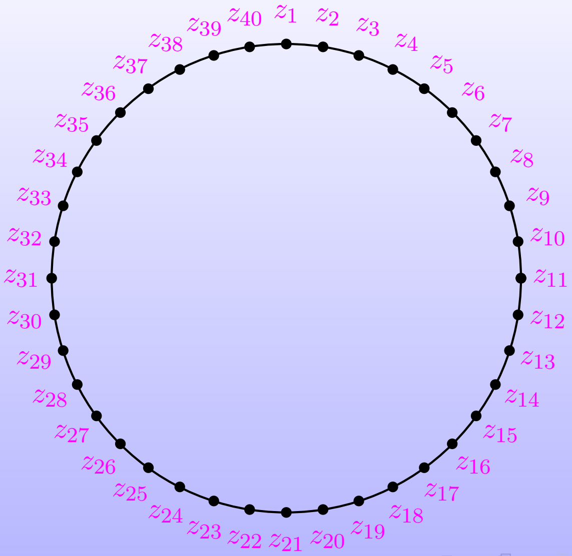
For $N = 40$ and $F = 8$:

- 1 time unit $\approx$ 5 days
- distance between observation stations: 750 km $\approx 30,000/40$ km



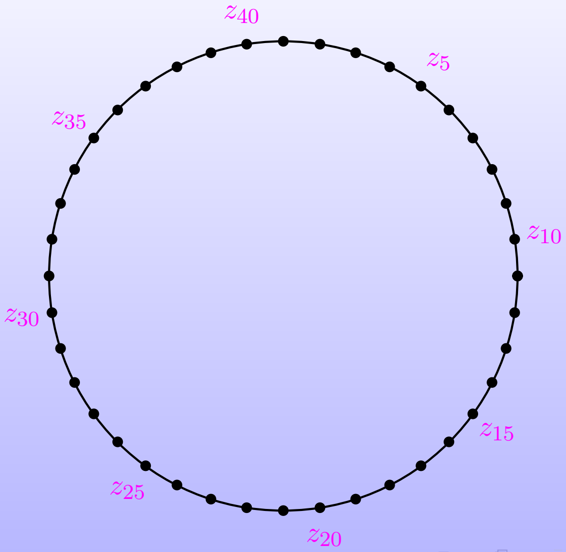
Lorenz-96 model

Instead of 40 observations



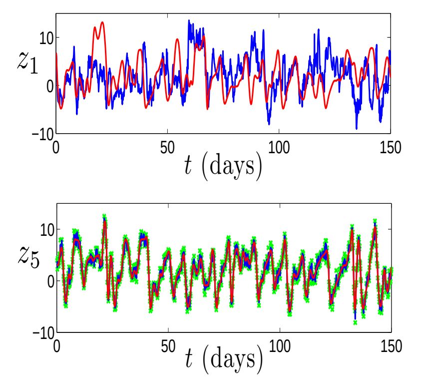
Lorenz-96 model

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



Lorenz-96 model

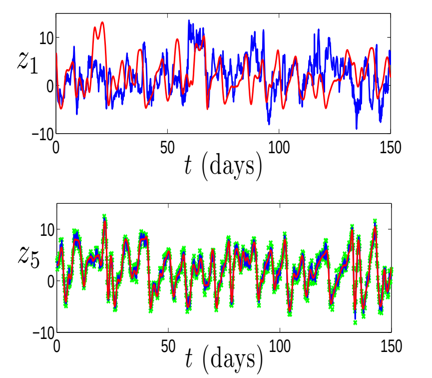
The pseudo-observables contain the prior climatic knowledge:
$\mathbf{a}_{\text{target}} = \mu_{\text{clim}}$ and $\mathbf{A}_{\text{target}} = \sigma_{\text{clim}}^2 \mathbf{I}$ with $\mu_{\text{clim}} = 2.34$ and $\sigma_{\text{clim}} = 3.6$ measured from a long time trajectory
$N_{\text{obs}} = 6$, $\Delta t_{\text{obs}} = 4$ hours, $\mathbf{R}_{\text{obs}} = (0.25\sigma_{\text{clim}})^2 \mathbf{I}$



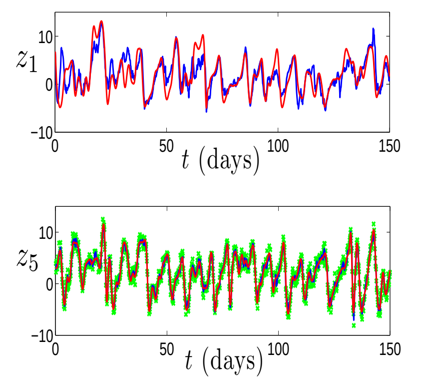
ETKF

Lorenz-96 model

The pseudo-observables contain the prior climatic knowledge:
$\mathbf{a}_{\mathrm{target}} = \mu_{\mathrm{clim}}$ and $\mathbf{A}_{\mathrm{target}} = \sigma_{\mathrm{clim}}^2 \mathbf{I}$ with $\mu_{\mathrm{clim}} = 2.34$ and $\sigma_{\mathrm{clim}} = 3.6$ measured from a long time trajectory
$N_{\mathrm{obs}} = 6$, $\Delta t_{\mathrm{obs}} = 4$ hours, $\mathbf{R}_{\mathrm{obs}} = (0.25\sigma_{\mathrm{clim}})^2 \mathbf{I}$



ETKF



VLKF

Lorenz-96 model

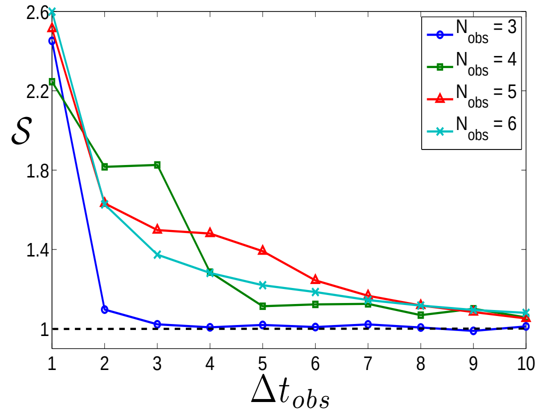
Quantify the skill improvement by the r.m.s error

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

$$\mathbf{R}_{\rm obs} = (0.25\sigma_{clim})^2\,\mathbf{I}$$

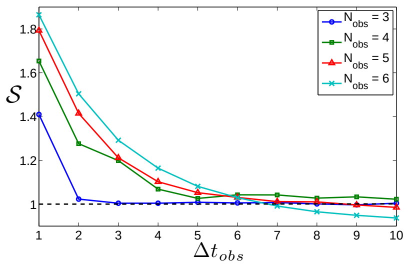
Best performance of VLKF over ETKF for:

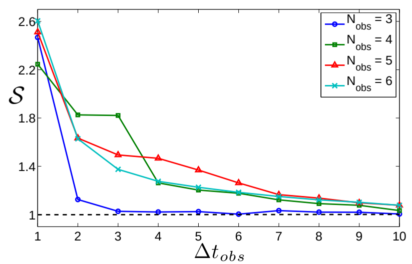
- small $\Delta t_{\rm obs}$
- $N_{\rm obs} = 4$



Lorenz-96 model

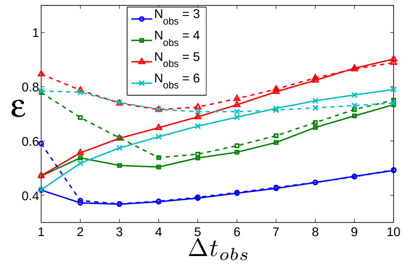
How is the skill distributed over

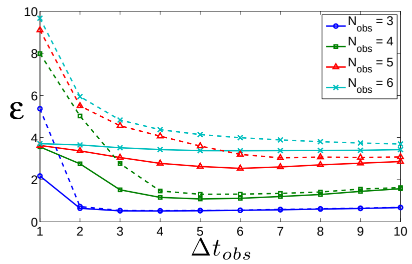




observables

pseudo-observables

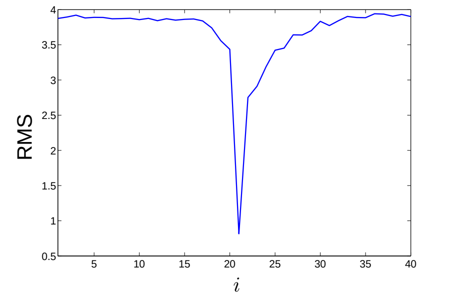




Lorenz-96 model

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

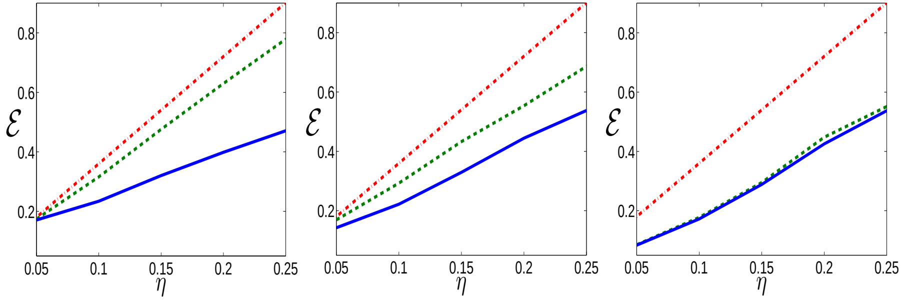
Total RMS error for each site i, $i = 1, 2, \cdots, 40$ if only site $i^\star = 21$ is observed.



Remark: The advective time scale of the Lorenz-96 system is much smaller than Δt_{obs} which explains why the skill is not equally distributed over the sites, and why, especially for large values of N_{obs} we observe a big difference between the site-averaged skills of the observed and unobserved variables.

Lorenz-96 model

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



$\Delta t_{\text{obs}} = 0.025$ (1 hour) $\Delta t_{\text{obs}} = 0.05$ (2 hours) $\Delta t_{\text{obs}} = 0.25$ (5 hours)

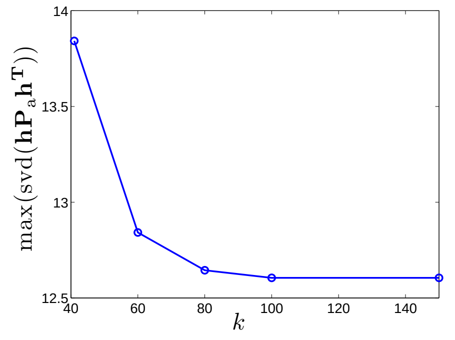
Red: Observations Green: ETKF Blue: VLKF

Lorenz-96 model

VLKF produces significant skill in sparse observational grids for

- small observation intervals (< 6 hours)
- the larger the observational noise the better

As before, the increased skill is a finite size effect:



Lorenz-96 model: Filter divergence and blow-up (*Harlim & Majda (2010)*) for small $\mathrm{R}_{\mathrm{obs}}$

ETKF

N_{obs}	6	0.14	×	×	0.98	0.96	0.76	0.32	0.05	0.02	0.01
	5	0.02	0.40	0.67	0.73	0.84	0.89	0.94	0.82	0.49	0.19
	4	0	0.04	0.22	0.29	0.49	0.64	0.77	0.83	0.89	0.82
	3	0	0	0	0.03	0.04	0.11	0.15	0.44	0.58	0.67
	2	0	0	0	0	0	0.01	0	0.01	0.05	0.15
	1	0	0	0	0	0	0	0	0	0	0
		0.025 3 h	0.05 6 h	0.075 9 h	0.1 12 h	0.125 15 h	0.15 18 h	0.175 21 h	0.2 24 h	0.225 27 h	0.25 30 h
		$\Delta\tau_{obs}$									

VLKF

N_{obs}	6	0.01	0.42	0.11	0.01	0	0	0	0	0	0
	5	0	0.24	0.36	0.10	0.01	0	0	0	0	0
	4	0	0.03	0.22	0.12	0.06	0.02	0	0	0	0
	3	0	0	0	0.02	0	0.01	0.01	0.01	0	0
	2	0	0	0	0	0	0	0	0	0	0.01
	1	0	0	0	0	0	0	0	0	0	0
		0.025 3 h	0.05 6 h	0.075 9 h	0.1 12 h	0.125 15 h	0.15 18 h	0.175 21 h	0.2 24 h	0.225 27 h	0.25 30 h
		$\Delta\tau_{obs}$									

Lorenz-96 model: Filter divergence and blow-up (*Harlim & Majda (2010)*) for small $\mathrm{R}_{\mathrm{obs}}$

ETKF

N_{obs}	6	0.14	×	×	0.98	0.96	0.76	0.32	0.05	0.02	0.01
	5	0.02	0.40	0.67	0.73	0.84	0.89	0.94	0.82	0.49	0.19
	4	0	0.04	0.22	0.29	0.49	0.64	0.77	0.83	0.89	0.82
	3	0	0	0	0.03	0.04	0.11	0.15	0.44	0.58	0.67
	2	0	0	0	0	0	0.01	0	0.01	0.05	0.15
	1	0	0	0	0	0	0	0	0	0	0
		0.025	0.05	0.075	0.1	0.125	0.15	0.175	0.2	0.225	0.25
		3 h	6 h	9 h	12 h	15 h	18 h	21 h	24 h	27 h	30 h
		$\Delta\tau_{obs}$									

VLKF

N_{obs}	6	0.01	0.42	0.11	0.01	0	0	0	0	0	0
	5	0	0.24	0.36	0.10	0.01	0	0	0	0	0
	4	0	0.03	0.22	0.12	0.06	0.02	0	0	0	0
	3	0	0	0	0.02	0	0.01	0.01	0.01	0	0
	2	0	0	0	0	0	0	0	0	0	0.01
	1	0	0	0	0	0	0	0	0	0	0
		0.025	0.05	0.075	0.1	0.125	0.15	0.175	0.2	0.225	0.25
		3 h	6 h	9 h	12 h	15 h	18 h	21 h	24 h	27 h	30 h
		$\Delta\tau_{obs}$									

(*GAG, Mitchell & Reich, MWR 2011, in press*)

II. Model Error

Numerical codes tend to overestimate "noise" at the grid resolution. This is typically controlled via artificial viscosity.
For example, to control unwanted gravity wave activity severe divergence damping is introduced to the equations of motion to stabilize the numerical scheme causing an underestimation of the error covariances (*Durran 1999*). Sometimes we want conservation properties (J. Thuburn 2008).

Can one use numerical forecast models which are not artificially damped and control the resulting overestimation of the forecast covariance within the data assimilation procedure?

$$\frac{dz_i}{dt} = z_{i-1}(z_{i+1} - z_{i-2}) - \gamma z_i + F$$

Truth: $\gamma = 1$

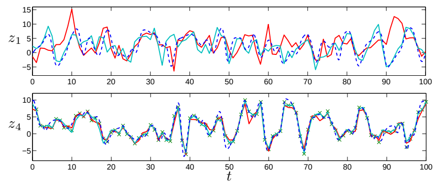
Forecast model: $\gamma < 1$

Now we will be interested in the case of $\Delta t_{\mathrm{obs}} \gg 1$

Model Error

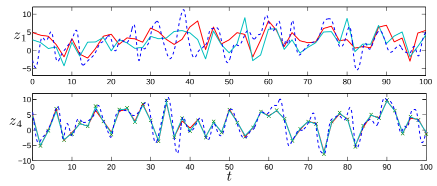
$\gamma = 0.5$:

$\Delta t_{\text{obs}} = 24$ hours



Reproducing the truth

$\Delta t_{\text{obs}} = 48$ hours

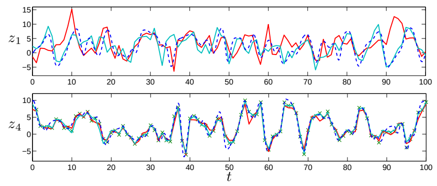


Reproducing the statistics

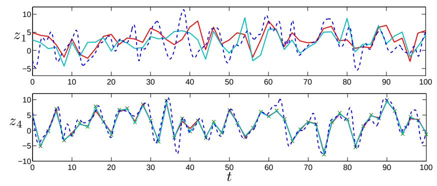
Model Error

$$\gamma = 0.5:$$

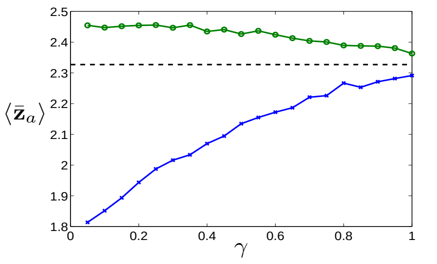
$\Delta t_{\mathrm{obs}} = 24$ hours



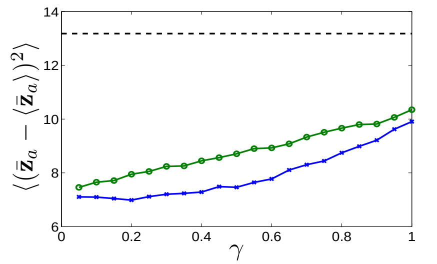
$\Delta t_{\mathrm{obs}} = 48$ hours



Reproducing the truth



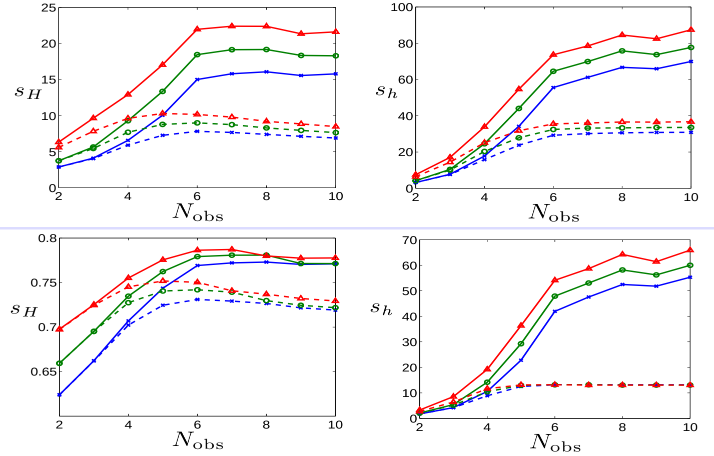
Reproducing the statistics



Model Error

$\mathbf{P}_f$:

$\mathbf{P}_a$:

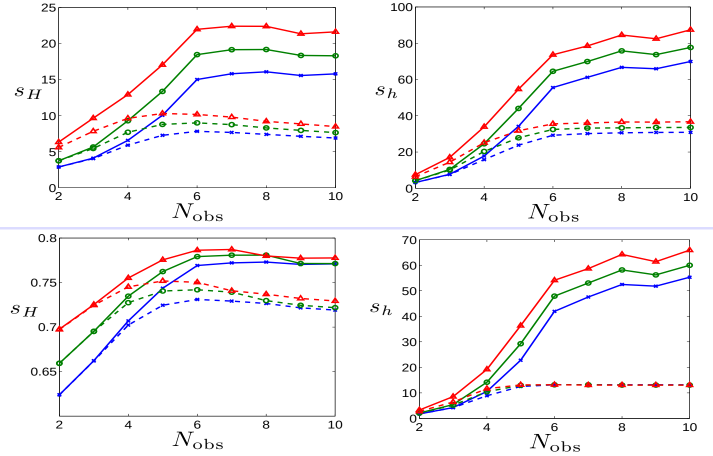


$\gamma = 1$ $\gamma = 0.9$ $\gamma = 0.8$ ($\Delta t_{\mathrm{obs}} = 48$hrs) continuous: ETKF dashed: VLKF

Model Error

$\mathbf{P}_f$:

$\mathbf{P}_a$:



$\gamma = 1$ $\quad \gamma = 0.9$ $\quad \gamma = 0.8$ $\quad (\Delta t_{\rm obs} = 48\text{hrs})$ continuous: ETKF dashed: VLKF

- increased sparsity and model error lead to overestimation
- covariances of observables are also limited for VLKF

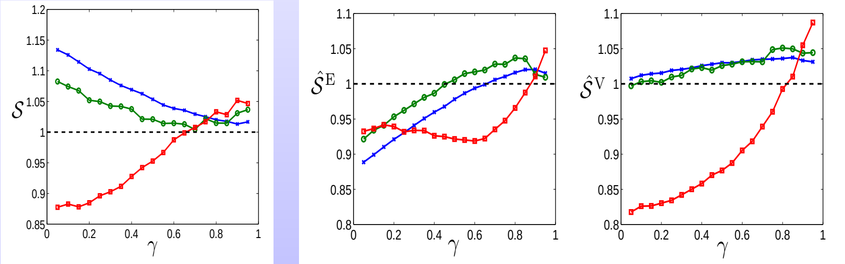
Model Error

We consider performance over standard ETKF

$$\mathcal{S} = \frac{\mathcal{E}^{\mathrm{E}}}{\mathcal{E}^{\mathrm{V}}} \ ,$$

and over using the "poor man's" analysis of observations and climatology

$$\hat{\mathcal{S}}^{E} = \frac{\hat{\mathcal{E}}}{\mathcal{E}^{\mathrm{E}}} \ , \qquad \hat{\mathcal{S}}^{V} = \frac{\hat{\mathcal{E}}}{\mathcal{E}^{\mathrm{V}}} \ .$$



$\Delta t_{\mathrm{obs}} = 24$hrs $\quad \Delta t_{\mathrm{obs}} = 36$hrs $\quad \Delta t_{\mathrm{obs}} = 48$hrs

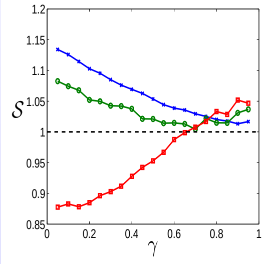
Model Error

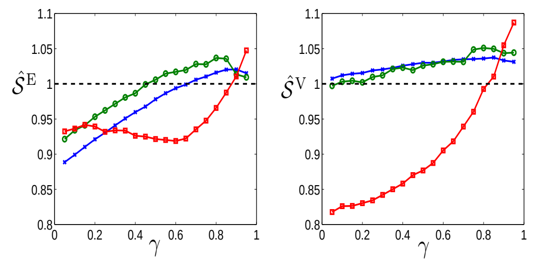
We consider performance over standard ETKF

$$\mathcal{S} = \frac{\mathcal{E}^{\rm E}}{\mathcal{E}^{\rm V}} \ ,$$

and over using the "poor man's" analysis of observations and climatology

$$\hat{\mathcal{S}}^{E} = \frac{\hat{\mathcal{E}}}{\mathcal{E}^{\rm E}} \ , \qquad \hat{\mathcal{S}}^{V} = \frac{\hat{\mathcal{E}}}{\mathcal{E}^{\rm V}} \ .$$



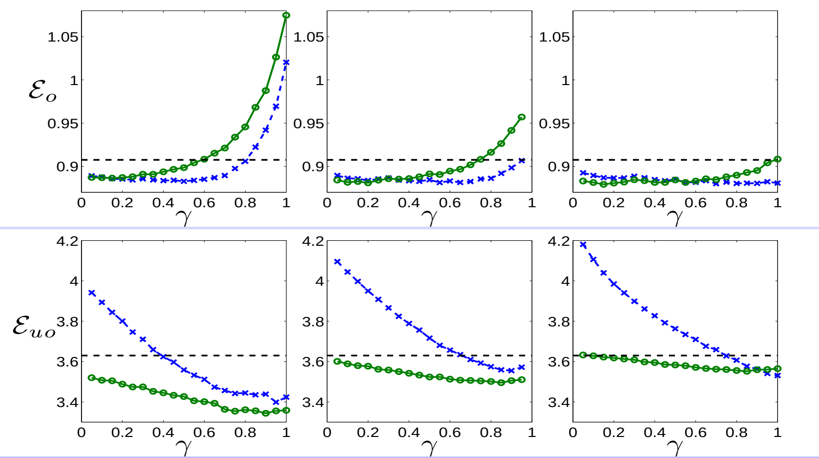


$\Delta t_{\rm obs} = 24$hrs $\quad \Delta t_{\rm obs} = 36$hrs $\quad \Delta t_{\rm obs} = 48$hrs

Trade-off: The smaller γ, the better skill over ETKF, but the less skill compared with "poor man's" analysis

Model Error

How is the RMS error distributed over the observables and the pseudo-observables ($\Delta t_{\text{obs}} = 48$hrs)?



$N_{\text{obs}} = 3$ $N_{\text{obs}} = 4$ $N_{\text{obs}} = 8$

(ETKF/VLKF)

Summary and outlook

We have here

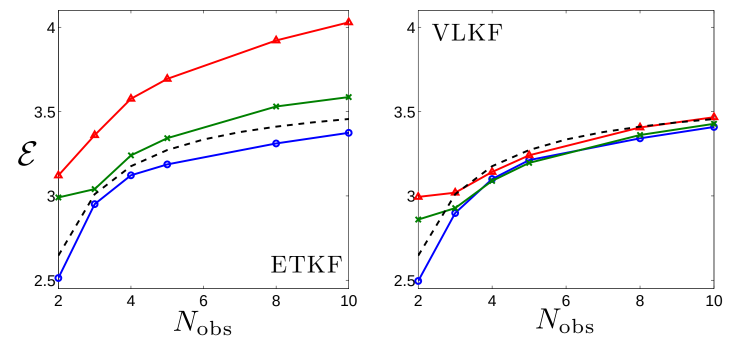
- derived a variance limiting Kalman filter (VLKF) which adaptively damps unrealistic excitation of ensemble spread in underresolved regions
- applied this filter to a sparse observational grid
  - has better skill than ETKF for small ($\leq 6h$) observation times
  - has better skill for observables and pseudo-observables
  - is stabilizing and avoids filter divergencies such as blow-up
  - is robust to incomplete knowledge of the climatic mean and variance
- applied this filter to model error (underdamping)
  - has better skill than ETKF for large observation intervals ($\geq 36h$)
  - has worse or equal skill for observed variables
  - trade-off between superior skill over ETKF and being better than observations/climatology

We will

- apply this filter to slow-fast systems where fast degrees of freedom do not need to be tracking
- investigate blow-up further

Model Error

Dependency of skill on sparsity N_{obs}



Model Error

Dependency of skill on noise error $\mathbf{R}_{\text{obs}} = (\eta\sigma_{\text{clim}})^2\mathbf{I}$