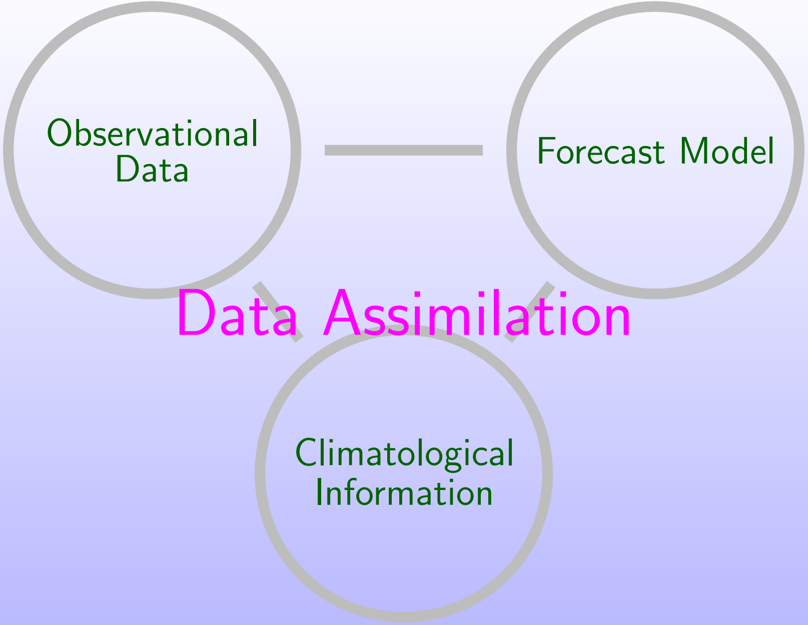
Constraining overestimation of error covariances in ensemble Kalman filters

Georg Gottwald, Andy Majda°, Lewis Mitchell*, Sebastian Reich†

University of Sydney, ○Courant, *University of Vermont, †Universität Potsdam

Toulouse, November 13th, 2012

Data Assimilation


Observational Data
Forecast Model
Data Assimilation
Climatological Information

Using climatology on 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*)

Using climatology on 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*)

Applications are

- sparse observational networks
- balance
- model error
- controlling catastrophic filter divergence

Using climatology on 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*)

Applications are

- sparse observational networks
- balance
- model error
- controlling catastrophic filter divergence

Further applications are

- re-analysis of climate
- when direct observations are not available (mesosphere)
- general 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?

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).

observables

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

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

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).

observables

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

- observation operator $\mathbf{H} : \mathbb{R}^N \to \mathbb{R}^n$
- $\mathbf{x}_{\text{obs}}(t_i) = \mathbf{H}\mathbf{z}(t_i) + \mathbf{r}_{\text{obs}}(t_i)$ 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

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).

observables

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

- observation operator $\mathbf{H} : \mathbb{R}^N \to \mathbb{R}^n$
- $\mathbf{x}_{\text{obs}}(t_i) = \mathbf{H}\mathbf{z}(t_i) + \mathbf{r}_{\text{obs}}(t_i)$ 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 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

$$J(\mathbf{z}) = \frac{1}{2}(\mathbf{z} - \mathbf{z}_f)^T \mathbf{P}_f^{-1}(\mathbf{z} - \mathbf{z}_f) + \frac{1}{2}(\boldsymbol{\mathcal{H}}\mathbf{z} - \boldsymbol{\mathcal{Y}})^T \boldsymbol{\mathcal{R}}^{-1}(\boldsymbol{\mathcal{H}}\mathbf{z} - \boldsymbol{\mathcal{Y}})$$

$$\boldsymbol{\mathcal{Y}} = \begin{pmatrix} \mathbf{x}_{\mathrm{obs}} \\ \mathbf{a}_{\mathrm{target}} \end{pmatrix}, \boldsymbol{\mathcal{H}} = \begin{pmatrix} \mathbf{H} \\ \mathbf{h} \end{pmatrix}, \boldsymbol{\mathcal{R}}^{-1} = \begin{pmatrix} \mathbf{R}_{\mathrm{obs}}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{\mathrm{w}}^{-1} \end{pmatrix}$$

The Variance Limiting Kalman Filter (VLKF)

Step 2: Analysis step

$$J(\mathbf{z}) = \frac{1}{2}(\mathbf{z}-\mathbf{z}_f)^T \mathbf{P}_f^{-1}(\mathbf{z}-\mathbf{z}_f) + \frac{1}{2}(\mathcal{H}\mathbf{z}-\mathcal{Y})^T \mathcal{R}^{-1}(\mathcal{H}\mathbf{z}-\mathcal{Y})$$

$$\mathcal{Y} = \begin{pmatrix} \mathbf{x}_{\rm obs} \\ \mathbf{a}_{\rm target} \end{pmatrix}, \mathcal{H} = \begin{pmatrix} \mathbf{H} \\ \mathbf{h} \end{pmatrix}, \mathcal{R}^{-1} = \begin{pmatrix} \mathbf{R}_{\rm obs}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{\rm w}^{-1} \end{pmatrix}$$

$$\bar{\mathbf{z}}_a = \bar{\mathbf{z}}_f - \mathcal{K}\left[\mathcal{H}\mathbf{z}_f - \mathcal{Y}\right]$$

where $\mathcal{K} = \mathbf{P}_f \mathcal{H}^T (\mathcal{H}\mathbf{P}_f\mathcal{H}^T + \mathcal{R})^{-1}$

with the covariance of the analysis

$$\mathbf{P}_a = \left[\mathbf{I} - \mathcal{K}\mathcal{H}\right]\mathbf{P}_f$$

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

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

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

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*): $\mathbf{Z}'_a = \mathbf{Z}'_f \boldsymbol{S}_{\mathrm{w}}$ with $\boldsymbol{S}_{\mathrm{w}} \in \mathbb{R}^{k \times k}$
- 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*):
  $\mathbf{Z}'_a = \mathbf{Z}'_f \boldsymbol{S}_\mathrm{w}$ with $\boldsymbol{S}_\mathrm{w} \in \mathbb{R}^{k\times k}$
- 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

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

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

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

Step 3: Update of the ensemble

The ensemble needs to be consistent with

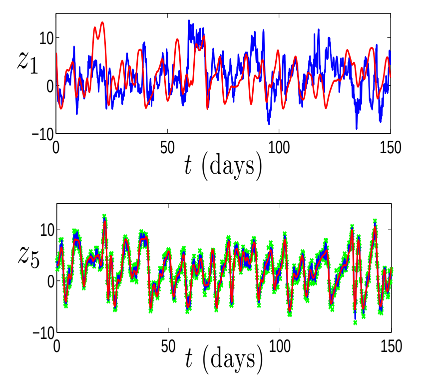
$$\mathbf{P}_a = [\mathbf{I} - \mathbf{K}_{\text{obs}}\mathbf{H} - \mathbf{K}_{\text{w}}\mathbf{h}]\ \mathbf{P}_f \ = \ \frac{1}{k-1}\mathbf{Z}'_a\ [\mathbf{Z}'_a]^T$$

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.

I. Lorenz-96 model: $\dot{z}_i = z_{i-1}(z_{i+1} - z_{i-2}) - z_i + F$

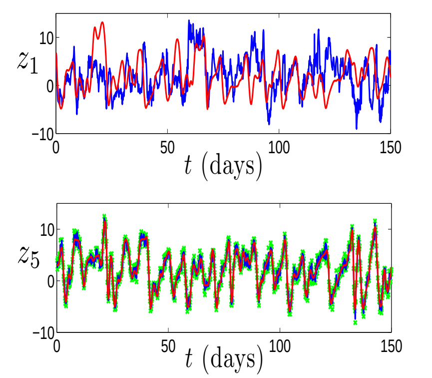
The pseudo-observables contain the prior climatic knowledge:
$\mathbf{a}_{\text{target}} = \mu_{\text{clim}}$ and $\mathbf{A}_{\text{target}} = \sigma^2_{\text{clim}}\mathbf{I}$ with $\mu_{\text{clim}} = 2.34$ and $\sigma_{\text{clim}} = 3.6$ measured from a long time trajectory
$N_{\text{obs}} = 5$, $\Delta t_{\text{obs}} = 4$ hours, $\mathbf{R}_{\text{obs}} = (0.25\sigma_{\text{clim}})^2\mathbf{I}$



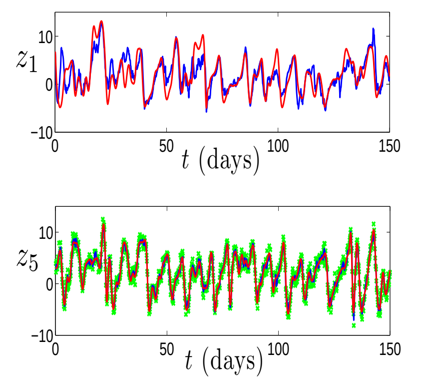
ETKF

I. Lorenz-96 model: $\dot{z}_i = z_{i-1}\left(z_{i+1} - z_{i-2}\right) - z_i + F$

The pseudo-observables contain the prior climatic knowledge:
$\mathbf{a}_{\text{target}} = \mu_{\text{clim}}$ and $\mathbf{A}_{\text{target}} = \sigma^2_{\text{clim}}\mathbf{I}$ with $\mu_{\text{clim}} = 2.34$ and $\sigma_{\text{clim}} = 3.6$ measured from a long time trajectory
$N_{\text{obs}} = 5$, $\Delta t_{\text{obs}} = 4$ hours, $\mathbf{R}_{\text{obs}} = (0.25\sigma_{\text{clim}})^2\mathbf{I}$



ETKF



VLKF

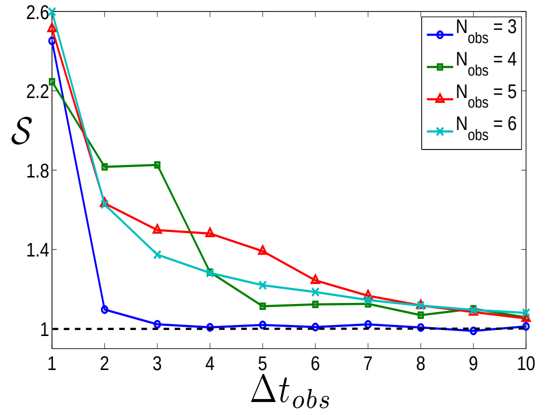
Quantify the skill improvement by the r.m.s error

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

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

Best performance of VLKF over ETKF for:

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



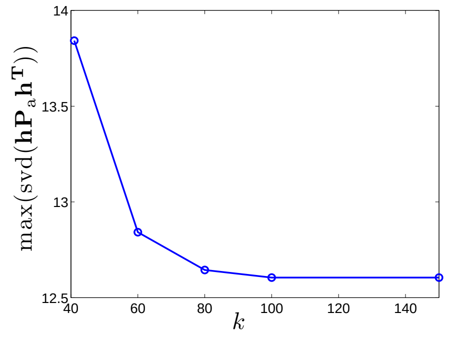
I. Lorenz-96 model

VLKF produces significant skill in sparse observational grids for

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

VLKF produces significant skill in sparse observational grids for

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



The overestimation of error covariances in sparse networks is a finite ensemble size effect

II. Filter divergence and blow-up with sparse observations

(traditional) filter divergence: Underestimation of error covariance leads to filter trusting its own forecast for sufficiently large $\mathbf{R}_{\rm obs}$ (cf. *Ng et al (2011)*)

II. Filter divergence and blow-up with sparse observations

(traditional) filter divergence: Underestimation of error covariance leads to filter trusting its own forecast for sufficiently large $\mathbf{R}_{\rm obs}$ (cf. *Ng et al (2011)*)

catastrophic filter divergence: Filter develops machine-infinity blow-up for sufficiently small $\mathbf{R}_{\rm obs}$ (*Harlim & Majda (2010); GAG, Mitchell, Reich (2011)*)

ETKF

N_{obs}	3 h	6 h	9 h	12 h	15 h	18 h	21 h	24 h	27 h	30 h
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
	$\Delta\tau_{obs}$									

VLKF

N_{obs}	3 h	6 h	9 h	12 h	15 h	18 h	21 h	24 h	27 h	30 h
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
	$\Delta\tau_{obs}$									

II. Genesis of blow-up

We study the 5D Lorenz-96 model

$$\dot{z}_i = z_{i-1}(z_{i+1} - z_{i-2}) - z_i + F \qquad i = 1, \cdots, 5$$

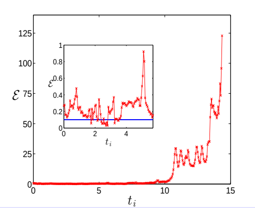
with negative forcing $F = -16$

Lyapunov exponents: $\lambda = (2.72, 0.09, -0.09, -1.83, -5.89)$

Attractor dimension: $D_{\mathrm{attr}} = 4.15$

Decay rate of the autocorrelation: $\tau_{\mathrm{corr}} \approx 0.14$

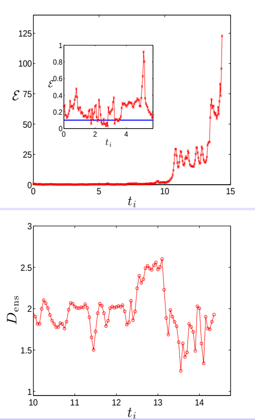
II. Genesis of blow-up



RMS error

- stable until $t_1 \approx 10$
- non-tracking episode until $t_2 \approx 13$
- blow-up for $t > 13$

II. Genesis of blow-up



RMS error

- stable until $t_1 \approx 10$
- non-tracking episode until $t_2 \approx 13$
- blow-up for $t > 13$

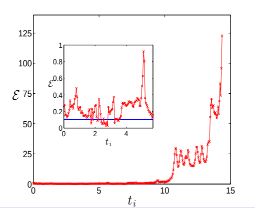
Finite size effects: Ensemble dimension (*Patil et al (2001)*)

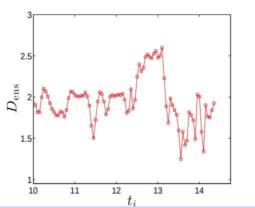
$$D_{\text{ens}} = \frac{\left(\sum_{i=1}^{k} \sqrt{\mu_i}\right)^2}{\sum_{i=1}^{k} \mu_i} \in (1, \min(k, D))$$

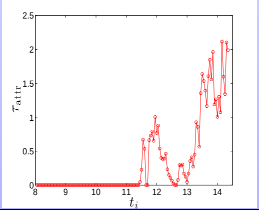
where μ_i are eigenvalues of the $k \times k$ covariance matrix

$$C = X_f^T X_f$$

II. Genesis of blow-up







RMS error

- stable until $t_1 \approx 10$
- non-tracking episode until $t_2 \approx 13$
- blow-up for $t > 13$

Finite size effects: Ensemble dimension (*Patil et al (2001)*)

$$D_{\text{ens}} = \frac{\left(\sum_{i=1}^{k} \sqrt{\mu_i}\right)^2}{\sum_{i=1}^{k} \mu_i} \in (1, \min(k, D))$$

where μ_i are eigenvalues of the $k \times k$ covariance matrix

$$C = X_f^T X_f$$

Filter pushes analysis off the attractor

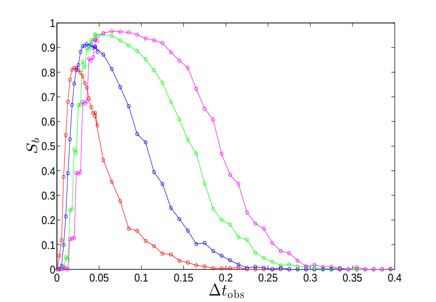
$$\bar{\mathbf{z}}_{a,i} = \bar{\mathbf{z}}_{f,i} - \frac{\mathbf{P}_{f\,i1}}{\mathbf{P}_{f\,11} + \mathbf{R}_{\text{obs}}} \left[\bar{\mathbf{z}}_{f,i} - \mathbf{x}_{\text{obs}}\right]$$

Lyapunov exponents

$$\lambda = (2.72, 0.09, -0.09, -1.83, -5.89)$$

II. Filter divergence and blow-up with sparse observations

Blow-up is caused by the forecast scheme (numerical instability):



Take home message:

Catastrophic blow up may be caused by the combination of finite size ensembles and fast attraction towards the attractor when

- there are sparse but accurate observations
- the underlying system has high variance

III. Controlling balance

Modified Lorenz-96 system (Bergemann & Reich (2010))

$$\begin{aligned}\dot{x}_j &= (1-\eta)\left(x_{j-1}(x_{j+1}-x_{j-2})\right) - x_j + F \\ &\quad + \eta\left(x_{j-1}h_{j+1} - x_{j-2}h_{j-1}\right) \\ \varepsilon^2 \ddot{h}_j &= -h_j + \alpha^2\left(h_{j-1} - 2h_j + h_{j+1}\right) + x_j\end{aligned}$$

- fast part is purely dispersive
- nonlinear terms conserve energy
  $H = \frac{\eta}{2}\sum_{j=1}^{D}\left(\frac{\eta-1}{\eta}x_j^2 + \epsilon^2\dot{h}_j^2 + h_j^2 + \alpha^2(h_{j+1}-h_{j-1})^2 - 2x_jh_j\right)$
- approximate slow manifold given by
  $\mathcal{B}_j(x_j, h_j) = x_j - \left(h_j - \alpha^2\left(h_{j-1} - 2h_j + h_{j+1}\right)\right) = 0$

III. Controlling balance

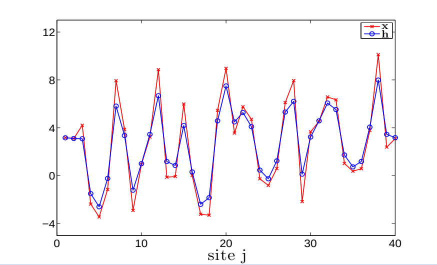
Initially balanced fields with

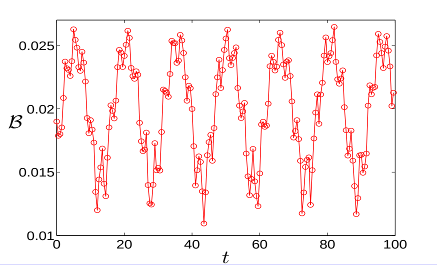
$$\mathcal{B}_j(x_j, h_j) = x_j - \left(h_j - \alpha^2 \left(h_{j-1} - 2h_j + h_{j+1}\right)\right) = 0$$

do not develop unbalanced motion on very long times:

The amount of unbalance (fast energy) can be measured by

$$\bar{\mathcal{B}}(t) = \sqrt{\frac{1}{D} \sum_{j=1}^{D} \left(x_j - h_j + \alpha^2 \left(h_{j-1} - 2h_j + h_{j+1}\right)\right)^2}$$

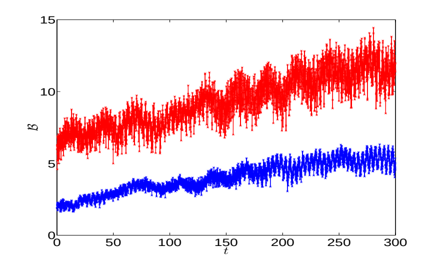




III. Controlling balance

The filtering procedure can severely disturb balance (cf. Lorenc (2003), Kepert (2009), Greybush et al (2011)) - this is the case with and without localization in sparse observational networks

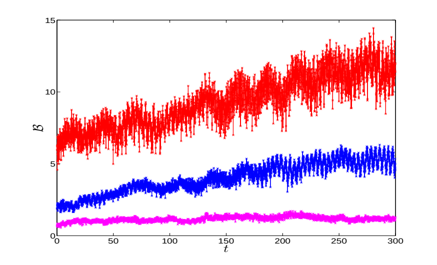
Here only x is observed with $N_{\rm obs} = 2$ with $\mathbf{h} = \mathcal{B}$ and $\mathbf{a}_{\rm target} = 0$

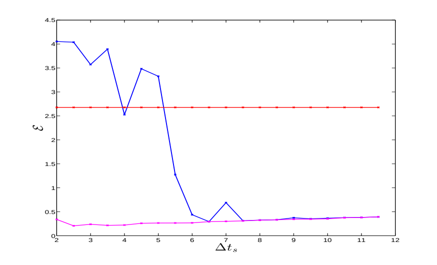


Red: ETKF Blue: DEnKF with localisation

III. Controlling balance

The filtering procedure can severely disturb balance (cf. Lorenc (2003), Kepert (2009), Greybush et al (2011))- this is the case with and without localization in sparse observational networks
Here only x is observed with $N_{\mathrm{obs}} = 2$ with $\mathbf{h} = \mathcal{B}$ and $\mathbf{a}_{\mathrm{target}} = 0$





Red: ETKF   Blue: DEnKF with localisation   Magenta: VLKF

IV. Model error and forecast bias

A major problem of ensemble filters is underdispersiveness (*Buizza et al 2005*).

This can be linked to

- *dynamical model error*: misrepresentation of unresolved subgrid scale processes (*Palmer 2001*)
- *numerical model error*: large errors produced at grid-scale (overestimation) which are controlled by artificial viscosity (underestimation)

IV. Model error and forecast bias

A major problem of ensemble filters is underdispersiveness (*Buizza et al 2005*).

This can be linked to

- *dynamical model error*: misrepresentation of unresolved subgrid scale processes (*Palmer 2001*)
- *numerical model error*: large errors produced at grid-scale (overestimation) which are controlled by artificial viscosity (underestimation) but
  - unrealistic drainage of energy out of the system (*Shutts 2005*)
  - frontogenesis (*Blumen 1990*)
  - large scale statistics depends on the numerically preserved conservation laws (*Thuburn 2008, Dubinkina & Frank 2007, 2010*)

IV. Model error and forecast bias

A major problem of ensemble filters is underdispersiveness (*Buizza et al 2005*).

This can be linked to

- *dynamical model error*: misrepresentation of unresolved subgrid scale processes (*Palmer 2001*)
- *numerical model error*: large errors produced at grid-scale (overestimation) which are controlled by artificial viscosity (underestimation) but
  - unrealistic drainage of energy out of the system (*Shutts 2005*)
  - frontogenesis (*Blumen 1990*)
  - large scale statistics depends on the numerically preserved conservation laws (*Thuburn 2008, Dubinkina & Frank 2007, 2010*)

Main question: Can one get away with underdamped forecast models, yet still control the resulting covariance overestimation within the data assimilation procedure?

IV. Model error and forecast bias

A major problem of ensemble filters is underdispersiveness (*Buizza et al 2005*).

This can be linked to

- *dynamical model error*: misrepresentation of unresolved subgrid scale processes (*Palmer 2001*)
- *numerical model error*: large errors produced at grid-scale (overestimation) which are controlled by artificial viscosity (underestimation) but
  - unrealistic drainage of energy out of the system (*Shutts 2005*)
  - frontogenesis (*Blumen 1990*)
  - large scale statistics depends on the numerically preserved conservation laws (*Thuburn 2008, Dubinkina & Frank 2007, 2010*)

Main question: Can one get away with underdamped forecast models, yet still control the resulting covariance overestimation within the data assimilation procedure?

We will use climatological information of the mean and variance

IV. Model error and forecast bias

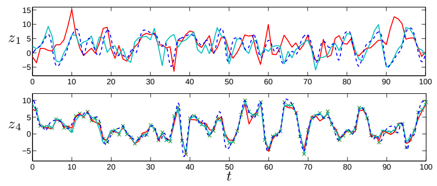
$$\frac{dz_i}{dt} = z_{i-1}(z_{i+1} - z_{i-2}) - \gamma z_i + F$$

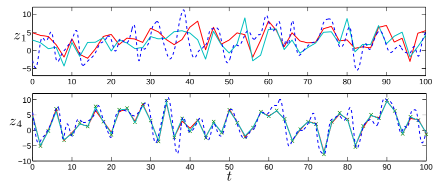
Truth: $\gamma = 1$
Forecast model: $\gamma < 1$

Perfect model case: $\mathcal{S} \approx 1$ for $\Delta t_{\mathrm{obs}} \gg 1$
Now we will be interested in the case of $\Delta t_{\mathrm{obs}} \gg 1$

IV. Model error and forecast bias

$$\gamma = 0.5$$





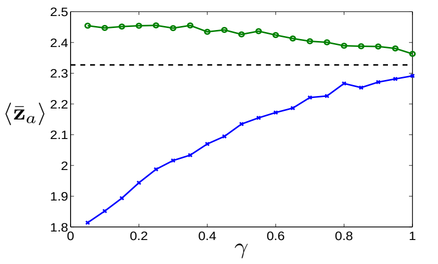
$\Delta t_{\text{obs}} = 24$ hours
Reproducing the truth

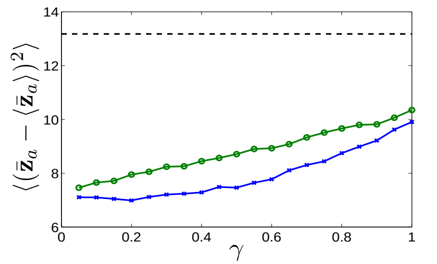
$\Delta t_{\text{obs}} = 48$ hours
Reproducing the statistics

(Truth VLKF ETKF)

IV. Model error and forecast bias

Reproducing the statistics





(ETKF VLKF truth)

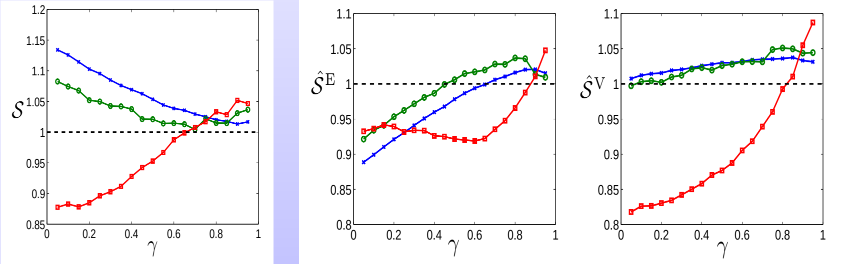
IV. Model error and forecast bias

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

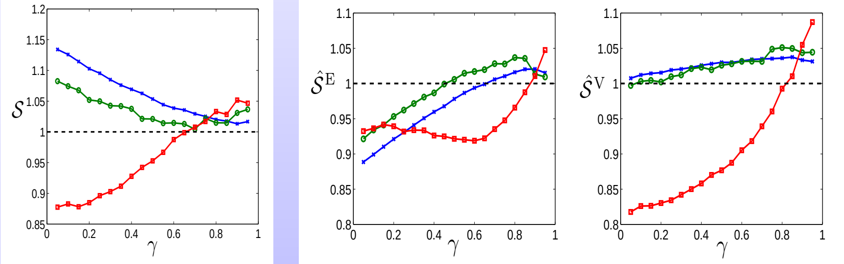
IV. Model error and forecast bias

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

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

Summary

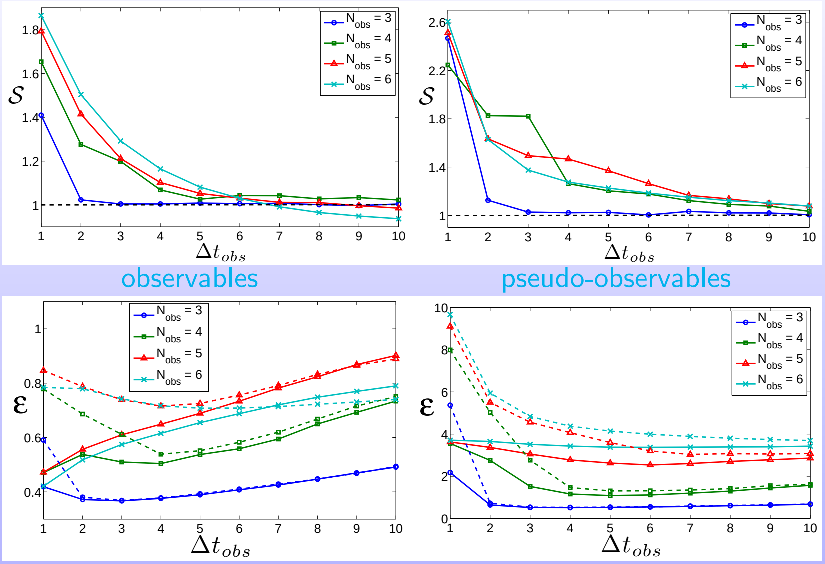
We have here

- 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
- proposed a mechanism for blow-up filter divergence
- applied this filter to control balance
  - has better skill than DEnKF and controls unbalance
- applied this filter to model error (underdamping)
  - has better skill than ETKF for large observation intervals ($\geq 36h$)
  - trade-off between superior skill over ETKF and being better than observations/climatology

(GAG, Mitchell & Reich, MWR 2011; Mitchell & GAG, QJRMS 2012, in press)

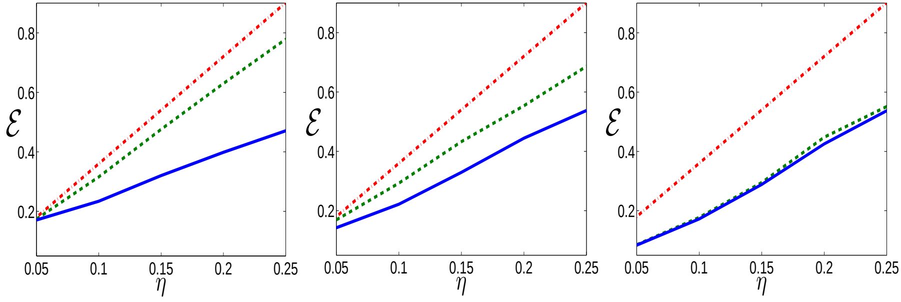
Lorenz-96 model

How is the skill distributed over



Lorenz-96 model

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



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

Red: Observations Green: ETKF Blue: VLKF

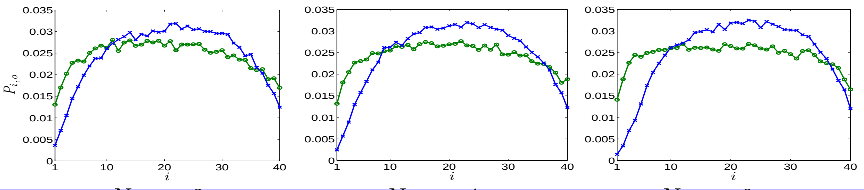
Constraining the covariances produces more reliable ensembles – even in the case when there is no skill improvement with $\mathcal{S} = 1$

Ranked probability histograms for the observables

- sort the forecast ensemble $\mathbf{X}_f = [x_{f,1}, x_{f,2}, ..., x_{f,k}]$ and create bins $(-\infty, x_{f,1}]$, $(x_{f,1}, x_{f,2}]$, ... , $(x_{f,k}, \infty)$ at each forecast step
- increment whichever bin the actual truth falls into at each forecast step

Convex histogram: underestimating ensemble
Concave histogram: overestimating ensemble
Flat histogram: reliable ensemble for which each ensemble member has equal probability of being nearest to the truth

$\gamma = 0.5$ and $\Delta t_{\rm obs} = 48$hrs



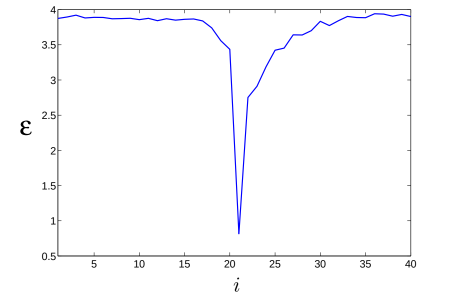
$N_{\rm obs} = 3$ $N_{\rm obs} = 4$ $N_{\rm obs} = 8$

(ETKF VLKF)

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.

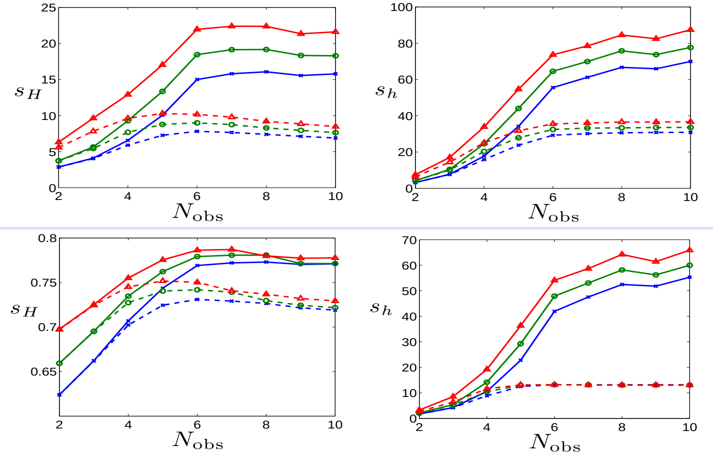


Remark: The advective time scale of the Lorenz-96 system is much smaller than Δt_{obs} which explains why the skill is not equally distributed over the sites, and why, especially for large values of N_{obs} we observe a big difference between the site-averaged skills of the observed and unobserved variables.

Application: Model error and forecast bias

$\mathbf{P}_f$:

$\mathbf{P}_a$:

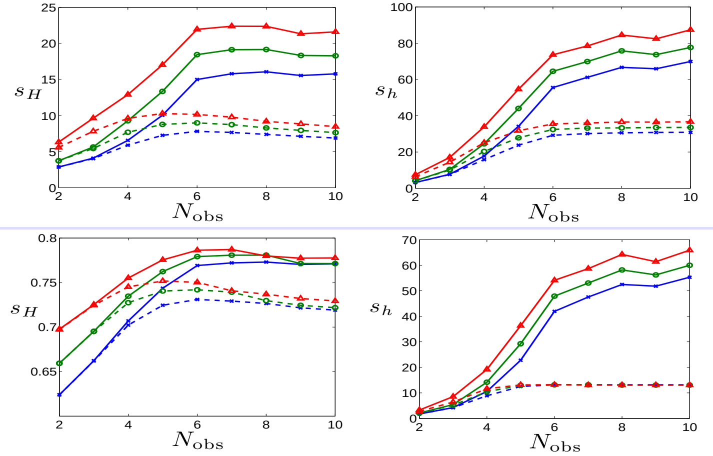


$\gamma = 1$ $\gamma = 0.9$ $\gamma = 0.8$ ($\Delta t_{\text{obs}} = 48$hrs) continuous: ETKF dashed: VLKF

Application: Model error and forecast bias

$\mathbf{P}_f$:

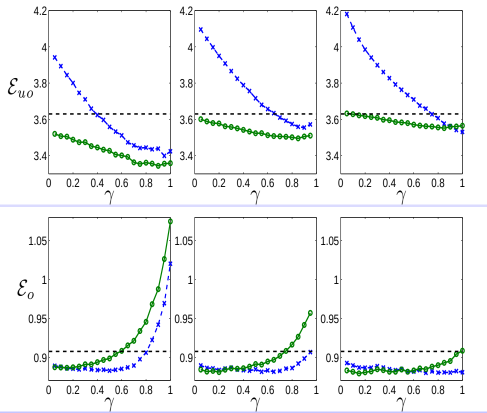
$\mathbf{P}_a$:



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

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

Distribution of RMS error over observables and pseudo-observables

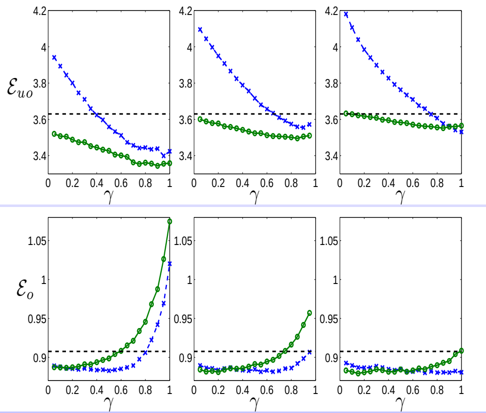


$N_{\text{obs}} = 3$ $\quad N_{\text{obs}} = 4$ $\quad N_{\text{obs}} = 8$

(ETKF/VLKF)

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

Distribution of RMS error over observables and pseudo-observables



$N_{\rm obs} = 3$ $\quad N_{\rm obs} = 4$ $\quad N_{\rm obs} = 8$

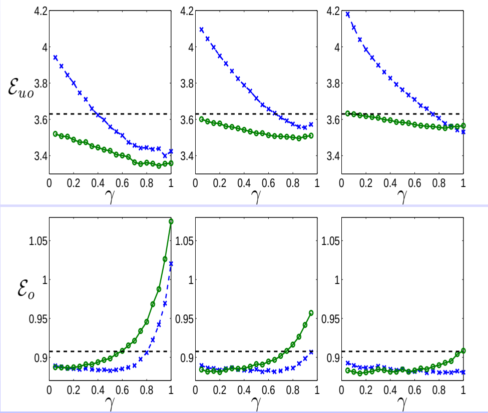
(ETKF/VLKF)

($\Delta t_{\rm obs} = 48$ hours)

pseudo-observables

- VLKF outperforms ETKF with increasing model error γ and increasing sparsity $N_{\rm obs}$
- VLKF becomes less effective compared to poor man's analysis with increasing model error γ and increasing sparsity $N_{\rm obs}$
- issue of overestimation

Distribution of RMS error over observables and pseudo-observables



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

(ETKF/VLKF)

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

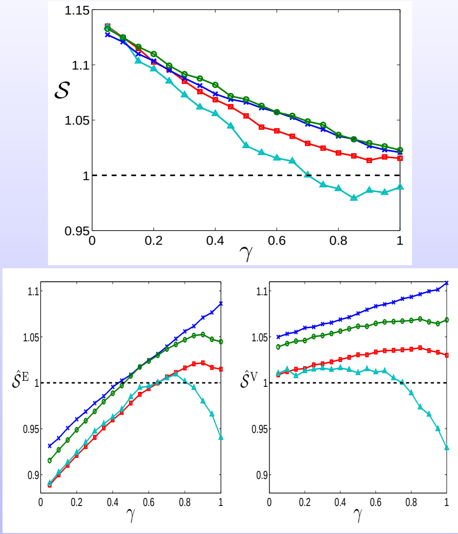
pseudo-observables

- VLKF outperforms ETKF with increasing model error γ and increasing sparsity N_{obs}
- VLKF becomes less effective compared to poor man's analysis with increasing model error γ and increasing sparsity N_{obs}
- issue of overestimation

observables

- ETKF outperforms VLKF for decreasing model error and decreasing sparsity
- ETKF performs worse than poor man's analysis for small model error
- convergence of ETKF andVLKF to $\mathcal{E}_o^\star < \sqrt{\mathbf{R}_{\mathrm{obs}}}$ for sufficiently large model error
- issue of underestimation

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



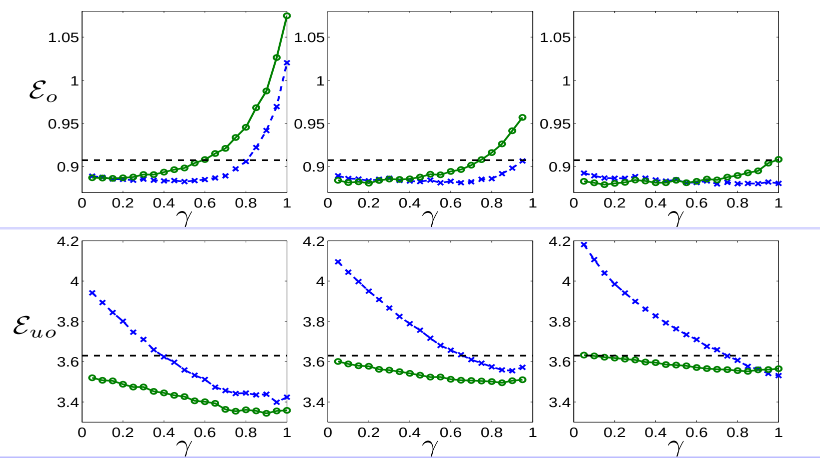
$\eta = 0.15$ $\eta = 0.15$ $\eta = 0.5$ $\eta = 1$

($\Delta t_{\mathrm{obs}} = 48$ hours, $N_{\mathrm{obs}} = 4$)

- VLKF outperforms ETKF and the poor man's analysis for sufficiently large observational noise since large noise gives more preference to the pseudo-observables which are controlled by the VLKF
- trade-off between skill over ETKF and efficacy over poor man's analysis with increasing model error

Application: Model error and forecast bias

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

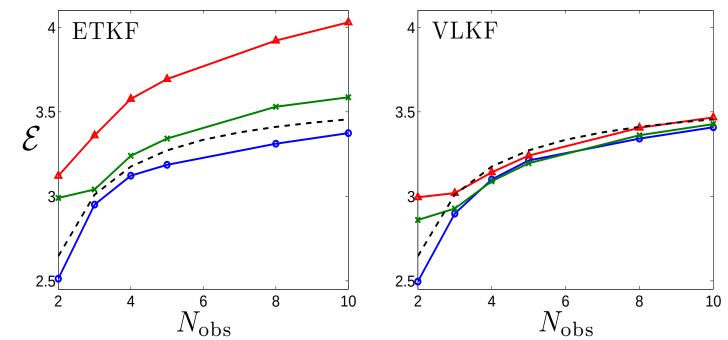


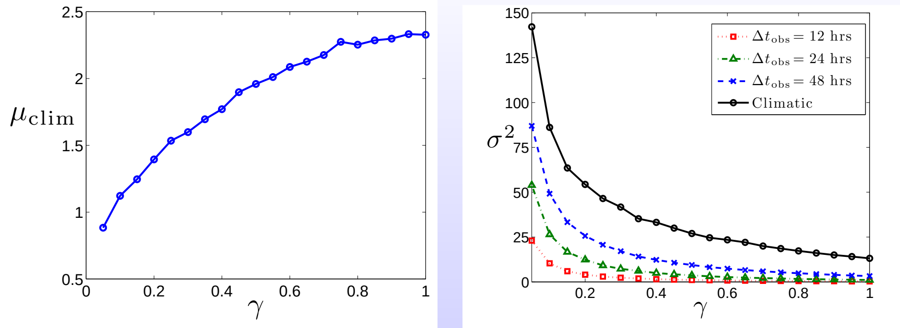
$N_{\text{obs}} = 3$ $N_{\text{obs}} = 4$ $N_{\text{obs}} = 8$

(ETKF/VLKF)

Model Error

Dependency of skill on sparsity N_{obs}





Dependence of climatic mean (left) and variance (right) on damping parameter γ. As well as the climatic variance (black, circles, solid) we show the average variance calculated over forecast intervals $\Delta t_{\rm obs} = 12$ hours (red, squares, dotted), $\Delta t_{\rm obs} = 24$ hours (green, triangles, dash-dotted) and $\Delta t_{\rm obs} = 48$ hours (blue, crosses, dashed).

Overestimation of forecast error covariance

Finite size ensemble sizes can lead to

- underestimation of diagonal elements of the forecast error covariance $\mathbf{P}_f$
- overestimation of off-diagonal elements of the forecast error covariance $\mathbf{P}_f$

To control off-diagonal terms one uses *localization* (*Houtekamer and Mitchell (1998)*), for example

$$\mathbf{P}_f \to C_{\mathrm{loc}} \circ \mathbf{P}_f$$

Can VLKF act as a form of localization?

Overestimation of forecast error covariance

Finite size ensemble sizes can lead to

- underestimation of diagonal elements of the forecast error covariance $\mathbf{P}_f$
- overestimation of off-diagonal elements of the forecast error covariance $\mathbf{P}_f$

To control off-diagonal terms one uses *localization* (*Houtekamer and Mitchell (1998)*), for example

$$\mathbf{P}_f \to C_{\mathrm{loc}} \circ \mathbf{P}_f$$

Can VLKF act as a form of localization?
2-d example $z = (x, y)$ where only x is observed:

$$\boldsymbol{\mathcal{K}}_{loc} = \frac{1}{P_{f\,11} + R} \begin{pmatrix} P_{f\,11} \\ C_{\mathrm{loc}\,21} P_{f\,21} \end{pmatrix}$$

$\boldsymbol{\mathcal{K}}_{VLKF} = \boldsymbol{\mathcal{K}}_{VLKF}(R_{\mathrm{w}})$:

- $R_{\mathrm{w}} \to \infty$: $\boldsymbol{\mathcal{K}}_{VLKF} \to \frac{1}{P_{f\,11}+R} \begin{pmatrix} P_{f\,11} & 0 \\ P_{f\,21} & 0 \end{pmatrix}$
- $R_{\mathrm{w}} \to \varepsilon R_{\mathrm{w}}$, $R \to \frac{1}{\varepsilon} R$: $\boldsymbol{\mathcal{K}}_{VLKF} \to \begin{pmatrix} \varepsilon \frac{\vartheta}{P_{f\,22} R} & \frac{P_{f\,12}}{P_{f\,22}} \\ \varepsilon^2 \frac{P_{f\,12}}{P_{f\,22}} \frac{R_{\mathrm{w}}}{R} & 1 \end{pmatrix}$ with $\vartheta = \det(\mathbf{P}_f)$