SANGOMA: Stochastic Assimilation for the Next Generation Ocean Model Applications SPA.2011.1.5-03 call, project 283580: Scale separation

J.-M. Beckers[a,c] & A. Barth[a,b] & I.Tomazic[a] & A.Alvera-Azcárate[a]

[a] GeoHydrodynamics and Environment Research, MARE, University of Liège
[b] Postdoctoral Researcher, F.R.S.-FNRS, BELGIUM
[c] Honorary Research Associate, F.R.S.-FNRS, BELGIUM

November 5-6, 2015, Liège

Additive processes

Different independent processes, each one with a different covariance matrix $\mathbf{B}_s$

$\Rightarrow$ the covariance of the total field is $\mathbf{B} = \sum_s \mathbf{B}_s$

Kalman gain:

$$\bar{\mathbf{K}} = \mathbf{B}\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right)^{-1} = \sum_s \mathbf{B}_s\mathbf{H}^{\mathrm{T}}\left(\sum_i \mathbf{H}\mathbf{B}_i\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right)^{-1} = \sum_s \bar{\mathbf{K}}_s \tag{1}$$

and

$$\bar{\mathbf{K}}_s = \mathbf{B}_s\mathbf{H}^{\mathrm{T}}\left(\sum_i \mathbf{H}\mathbf{B}_i\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right)^{-1}. \tag{2}$$

Additive processes

$$\bar{\mathsf{K}} = \sum_s \bar{\mathsf{K}}_s; \quad \bar{\mathsf{K}}_s = \mathsf{B}_s\mathsf{H}^{\mathrm{T}}\left(\sum_i \mathsf{H}\mathsf{B}_i\mathsf{H}^{\mathrm{T}} + \mathsf{R}\right)^{-1}. \qquad (3)$$

If $\left(\sum_i \mathsf{H}\mathsf{B}_i\mathsf{H}^{\mathrm{T}} + \mathsf{R}\right)^{-1}$ can be "inverted": No need to go further

Otherwise ...

Special individual covariance matrices

- Spline methods (covariances not explicitely calculated)
- 3DVar-NMC implementations
- Reduced rank approaches (EnKF, SEEK,...)
- OI with localization
- ...

using a covariance specification and Kalman gain

$$\mathbf{K}_s = \mathbf{B}_s\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\mathbf{B}_s\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right)^{-1}. \tag{4}$$

Special individual covariance matrices

Common property: Using a single covariance matrix specification does lead to efficient algorithms

$\Rightarrow \mathbf{K}_s$ can easily be applied to a (big) data array

So what ?

$$\mathsf{K}_s = \mathsf{B}_s\mathsf{H}^{\mathrm{T}}\left(\mathsf{H}\mathsf{B}_s\mathsf{H}^{\mathrm{T}} + \mathsf{R}\right)^{-1}. \tag{5}$$

can easily be applied to a (big) data array but

$$\bar{\mathsf{K}} = \sum_s \bar{\mathsf{K}}_s; \quad \bar{\mathsf{K}}_s = \mathsf{B}_s\mathsf{H}^{\mathrm{T}}\left(\sum_i \mathsf{H}\mathsf{B}_i\mathsf{H}^{\mathrm{T}} + \mathsf{R}\right)^{-1}. \tag{6}$$

can not.

Two processes (scales)

Matrix identity

$$\bar{\mathsf{K}}_1 = \mathsf{K}_1 - \mathsf{K}_1\mathsf{H}(\mathsf{I} - \mathsf{K}_2\mathsf{H}\mathsf{K}_1\mathsf{H})^{-1}\mathsf{K}_2\left(\mathsf{I} - \mathsf{H}\mathsf{K}_1\right) \tag{7}$$

$$= \mathsf{K}_1 - \mathsf{K}_1\mathsf{H}\mathsf{K}_2(\mathsf{I} - \mathsf{H}\mathsf{K}_1\mathsf{H}\mathsf{K}_2)^{-1}\left(\mathsf{I} - \mathsf{H}\mathsf{K}_1\right) \quad \text{Method P1a} \tag{8}$$

$$= (\mathsf{I} - \mathsf{K}_1\mathsf{H}\mathsf{K}_2\mathsf{H})^{-1}\mathsf{K}_1\left(\mathsf{I} - \mathsf{H}\mathsf{K}_2\right) \tag{9}$$

$$= \mathsf{K}_1(\mathsf{I} - \mathsf{H}\mathsf{K}_2\mathsf{H}\mathsf{K}_1)^{-1}\left(\mathsf{I} - \mathsf{H}\mathsf{K}_2\right) \quad \text{Method P1b} \tag{10}$$

but conservation of difficulty because of $(\mathsf{I} - \mathsf{H}\mathsf{K}_2\mathsf{H}\mathsf{K}_1)^{-1}$?

Lucky fact

$\mathbf{HK}_1$ and $\mathbf{HK}_2$ are so called "hat" matrices, "smaller than one". For $\mathbf{\Xi}$ "smaller than one"

$$(\mathbf{I}-\mathbf{\Xi})^{-1}=(\mathbf{I}+\mathbf{\Xi}\,(\mathbf{I}+\mathbf{\Xi}\,(\mathbf{I}+\mathbf{\Xi}(\ldots)\ldots))). \quad (11)$$

When applied to a vector $\mathbf{x}$, this immediately provides the algorithm to calculate $\mathbf{y}=(\mathbf{I}-\mathbf{\Xi})^{-1}\mathbf{x}$ as follows

$$\mathbf{y}\leftarrow\mathbf{x}$$

```
Loop
```

$$\mathbf{y}\leftarrow\mathbf{x}+\mathbf{\Xi}\mathbf{y} \quad (12)$$

```
End loop
```

Algorithm 1 : Approximate matrix inversion

Successive applications of existing tools

Even if iterations are stopped, some of the formulas yield identical results and there remain

$$\bar{\mathbf{K}}_1 = \mathbf{K}_1 - \mathbf{K}_1\mathbf{H}\mathbf{K}_2(\mathbf{I} - \mathbf{H}\mathbf{K}_1\mathbf{H}\mathbf{K}_2)^{-1}(\mathbf{I} - \mathbf{H}\mathbf{K}_1) \quad \text{Method P1a} \tag{13}$$

$$= \mathbf{K}_1(\mathbf{I} - \mathbf{H}\mathbf{K}_2\mathbf{H}\mathbf{K}_1)^{-1}(\mathbf{I} - \mathbf{H}\mathbf{K}_2) \quad \text{Method P1b} \tag{14}$$

and also

$$\bar{\mathbf{K}}_2 = \mathbf{K}_2 - \mathbf{K}_2\mathbf{H}\mathbf{K}_1(\mathbf{I} - \mathbf{H}\mathbf{K}_2\mathbf{H}\mathbf{K}_1)^{-1}(\mathbf{I} - \mathbf{H}\mathbf{K}_2) \quad \text{Method P2a} \tag{15}$$

$$= \mathbf{K}_2(\mathbf{I} - \mathbf{H}\mathbf{K}_1\mathbf{H}\mathbf{K}_2)^{-1}(\mathbf{I} - \mathbf{H}\mathbf{K}_1) \quad \text{Method P2b} \tag{16}$$

But which to chose ?

Synthetic test case

Controlled experiment leads to very clear guidelines:

- Label the process with the highest signal-to-noise ratio as 1 and the other process as 2 and apply P1a + P2b if we want to achieve the best analysis.
- If both processes have a similar signal-to-noise ratio, we should label the larger scale process as process 1 so that the same formulas are still the best.
- For individual processes, P1a should be used for the iterated versions, whereas $\mathbf{K}_1$ is the best choice for a simple analysis for process 1; for process 2 version P2b and $\mathbf{K}_2(\mathbf{I} - \mathbf{H}\mathbf{K}_1)$ are indicated respectively for the iterated approach or a simple approach.
- For iterative methods, if scales are well separated or at least one of the processes has a small signal-to-noise ratio, only very few iterations are needed.

Implemenation

$\phi \leftarrow \mathbf{K}_1\mathbf{d}$

$\mathbf{w}_1 \leftarrow \mathbf{H}\phi$ (Use only analysis at data points without clouds)

$\mathbf{w}_1 \leftarrow (\mathbf{d} - \mathbf{w}_1)$ (Residuals)

Calibrate or check OI parameters at this point using $\mathbf{w}_1$

$\mathbf{w}_2 \leftarrow \mathbf{w}_1$ (Start of iterative matrix inversion)

Loop n times

$\mathbf{w}_2 \leftarrow \mathbf{H}\mathbf{K}_2\mathbf{w}_2$ (apply tool 2 and retrieve solution at data points)

$\mathbf{w}_2 \leftarrow \mathbf{H}\mathbf{K}_1\mathbf{w}_2$ (apply tool 1 and retrieve solution at data points)

$\mathbf{w}_2 \leftarrow \mathbf{w}_1 + \mathbf{w}_2$

End loop

$\omega_1 \leftarrow \mathbf{K}_2\mathbf{w}_2$

ω_1 now contains scale 2 field $\bar{\mathbf{K}}_2\mathbf{d}$; can be saved for other uses

$\phi \leftarrow \phi + \omega_1$ (Add it to $\mathbf{K}_1\mathbf{d}$)

$\mathbf{w}_1 \leftarrow \mathbf{H}\omega_1$

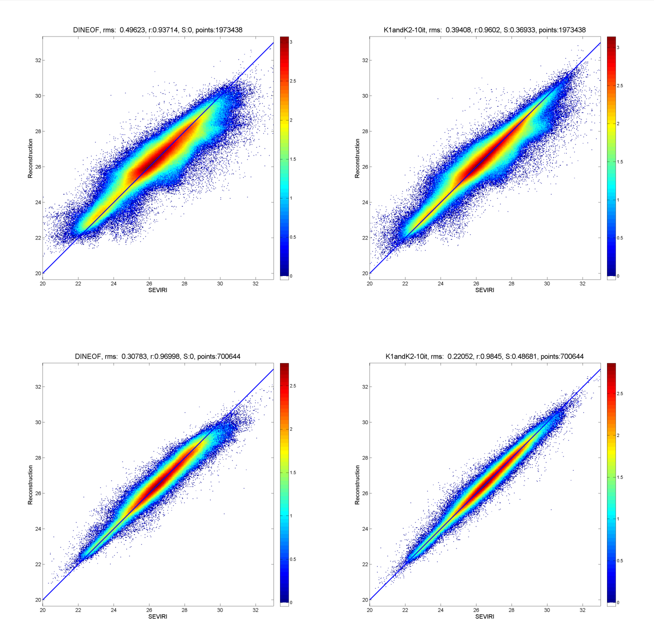
$\omega_1 \leftarrow \mathbf{K}_1\mathbf{w}_1$ (Calculates $\mathbf{K}_1\mathbf{H}\bar{\mathbf{K}}_2\mathbf{d}$)

$\phi \leftarrow \phi - \omega_1$ (Final analysis)

If scale 1 solution is requested, subtract saved scale 2 solution from ϕ and save it

DINEOF+OI, also outline for 3DVar-NMC

Example


DINEOF, rms: 0.49623, r:0.93714, S:0, points:1973438
K1andK2-10it, rms: 0.39408, r:0.9602, S:0.36933, points:1973438
DINEOF, rms: 0.30783, r:0.96998, S:0, points:700644
K1andK2-10it, rms: 0.22052, r:0.9845, S:0.48681, points:700644
Reconstruction
SEVIRI

Relevance for SANGOMA

- Possibly combine different covariance specifications (Reduced rank+classical OI)
- Possibility to include correlated observational errors into tools not allowing for them (treat **R** as a covariance matrix of a process (noise)) and isolate it
- Bias analysis

Bias analysis

- Process 1: model field $\mathbf{x}$
- Process 2: bias $\mathbf{b}$

and we want the unbiased field $\mathbf{u} = \mathbf{x} - \mathbf{b}$.
with property

$$\bar{\mathsf{K}}_1 = \mathsf{K}_1 \left(\mathsf{I} - \mathsf{H}\bar{\mathsf{K}}_2 \right) \quad (17)$$

and similarly

$$\bar{\mathsf{K}}_2 = \mathsf{K}_2 \left(\mathsf{I} - \mathsf{H}\bar{\mathsf{K}}_1 \right) \quad (18)$$

Analysis

$$\mathbf{x}^a = \mathbf{x}^f + \bar{\mathsf{K}}_1 \left(\mathbf{y} - \mathbf{H}\mathbf{u}^f \right) \tag{19}$$

$$\mathbf{b}^a = \mathbf{b}^f - \bar{\mathsf{K}}_2 \left(\mathbf{y} - \mathbf{H}\mathbf{u}^f \right) \tag{20}$$

and also

$$\mathbf{u}^a = \mathbf{u}^f + \left(\bar{\mathsf{K}}_1 + \bar{\mathsf{K}}_2 \right) \left(\mathbf{y} - \mathbf{H}\mathbf{u}^f \right) \tag{21}$$

A little formula shuffling

$$\mathbf{b}^a = \mathbf{b}^f - \bar{\mathsf{K}}_2 \left(\mathbf{y} - \mathbf{H}\mathbf{u}^f \right) \tag{22}$$

$$\mathbf{x}^a - \mathbf{b}^a = \mathbf{x}^f - \mathbf{b}^a + \mathsf{K}_1 \left(\mathbf{y} - \mathbf{H} \left(\mathbf{x}^f - \mathbf{b}^a \right) \right) \tag{23}$$

This is equivalent to the unsimplified Dee's approach:

Dee's method

Their method uses bias variables β in a reduced space from which bias on the model grid $\mathbf{b}$ is obtained by a linear relationship $\mathbf{b} = \mathbf{L}\beta$ (for bias parameters given in model space $\mathbf{L} = \mathbf{I}$). The method uses an augmented state vector $[\beta^\mathsf{T}\mathbf{x}^\mathsf{T}]^\mathsf{T}$ and the solution of the minimization as

$$\beta^a = \beta^f - \mathbf{K}_\beta \left[\mathbf{y} - \mathbf{H}\left(\mathbf{x}^f - \mathbf{b}^f\right)\right] \tag{24}$$

$$\mathbf{u}^a = \left(\mathbf{x}^f - \mathbf{b}^a\right) + \mathbf{K}_x \left[\mathbf{y} - \mathbf{H}\left(\mathbf{x}^f - \mathbf{b}^a\right)\right] \tag{25}$$

with Kalman gain matrices

$$\mathbf{K}_x = \mathbf{P}_x\mathbf{H}^\mathsf{T}\left(\mathbf{H}\mathbf{P}_x\mathbf{H}^\mathsf{T} + \mathbf{R}\right)^{-1} \tag{26}$$

$$\mathbf{K}_\beta = \mathbf{P}_\beta\mathbf{L}^\mathsf{T}\mathbf{H}^\mathsf{T}\left[\mathbf{H}\left(\mathbf{P}_x + \mathbf{L}\mathbf{P}_\beta\mathbf{L}^\mathsf{T}\right)\mathbf{H}^\mathsf{T} + \mathbf{R}\right]^{-1} \tag{27}$$

Is our multiscale version if $\mathbf{P}_x = \mathbf{B}_1$ and $\mathbf{P}_\beta = \mathbf{B}_2$.

Dee's method in practice

In Dee's approach then strong simplification are done because of the impossibility to afford inversion involved in the update of β.
Very strong assumption: $\mathbf{L} = \mathbf{I}$ and

$$\mathbf{P}_\beta = \gamma \mathbf{P}_x \tag{28}$$

with a tuning parameter γ.
Justification: seems to work AND is practicable see expression of

$$\mathbf{K}_\beta = \mathbf{P}_\beta \mathbf{L}^\mathsf{T} \mathbf{H}^\mathsf{T} \left[\mathbf{H} \left(\mathbf{P}_x + \mathbf{L} \mathbf{P}_\beta \mathbf{L}^\mathsf{T} \right) \mathbf{H}^\mathsf{T} + \mathbf{R} \right]^{-1} \tag{29}$$

Back to our idea

With iterations stopped early $\bar{\mathsf{K}}_2 \sim \mathsf{K}_2 (\mathsf{I} - \mathsf{HK}_1)$

$$\mathbf{r} = \mathbf{y} - \mathbf{H}\mathbf{u}^f \tag{30}$$

$$\tilde{\mathbf{r}} = \mathbf{r} - \mathbf{H}\mathbf{K}_1\mathbf{r} \tag{31}$$

$$\mathbf{b}^a = \mathbf{b}^f - \mathbf{K}_2\tilde{\mathbf{r}} \tag{32}$$

$$\mathbf{u}^a = \mathbf{x}^f - \mathbf{b}^a + \mathbf{K}_1\left(\mathbf{y} - \mathbf{H}\left(\mathbf{x}^f - \mathbf{b}^a\right)\right) \tag{33}$$

Yet another ordering of operations

$$\mathbf{u}^f = \mathbf{x}^f - \mathbf{b}^f \tag{34}$$

$$\tilde{\mathbf{y}} = \mathbf{y} - \mathsf{H}\mathsf{K}_2\mathbf{y} \tag{35}$$

$$\tilde{\mathbf{u}}^f = \mathbf{u}^f - \mathsf{K}_2\mathsf{H}\mathbf{u}^f \tag{36}$$

$$\mathbf{x}^a - \mathbf{b}^f = \mathbf{u}^f + \mathsf{K}_1\left(\tilde{\mathbf{y}} - \mathsf{H}\tilde{\mathbf{u}}^f\right) \tag{37}$$

$$\mathbf{b}^a = \mathbf{b}^f - \mathsf{K}_2\left(\mathbf{y} - \mathsf{H}(\mathbf{x}^a - \mathbf{b}^f)\right) \tag{38}$$

$$\mathbf{u}^a = \mathbf{x}^a - \mathbf{b}^f - (\mathbf{b}^a - \mathbf{b}^f) \tag{39}$$

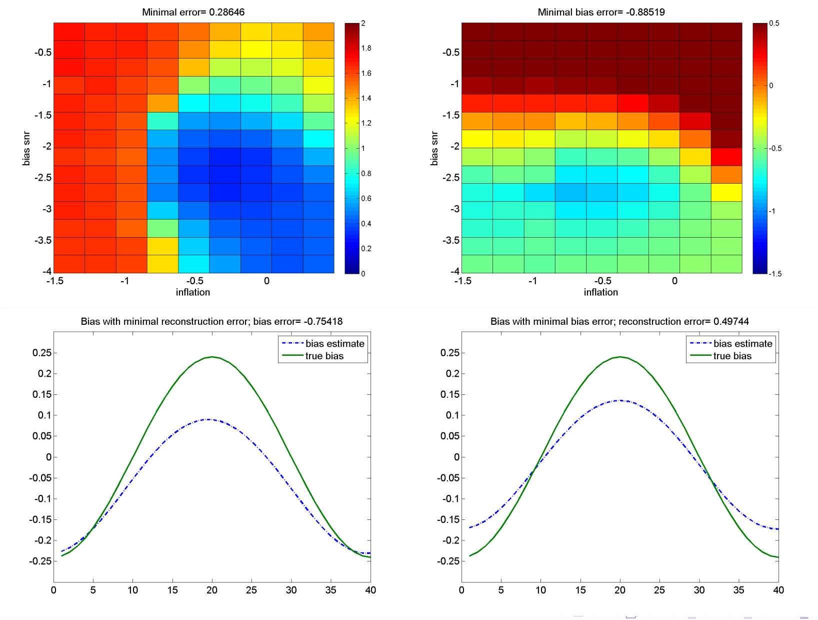
Test case

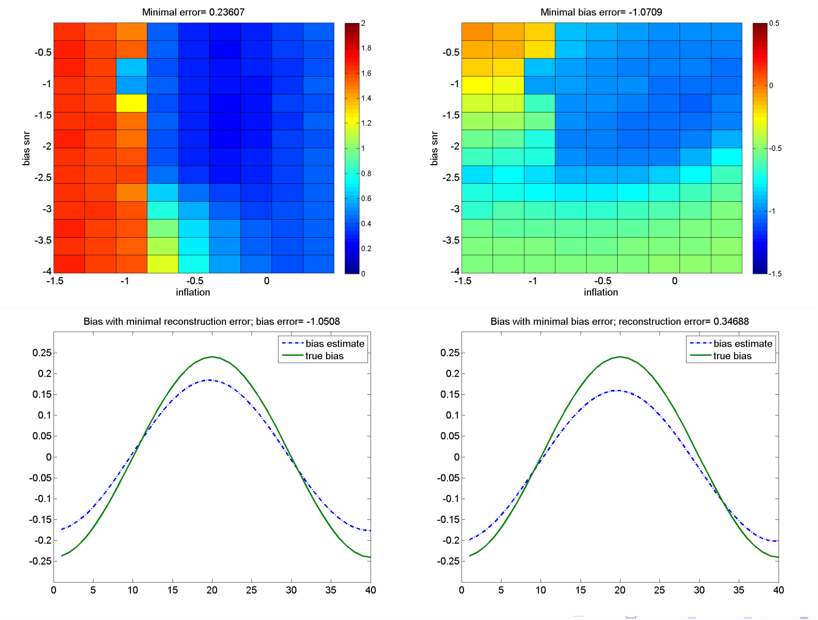
Small benchmark case with bias in equations.
Lorenz-96 with added bias. N =40 members, Model start from random initialization and then 10000 step. From there assimilation every step and we look after 20000 assimilation cycles. Initial members were constructed by adding random Gaussian noise of variance 1.3 in each location.
Additive bias $A \sin\left(\frac{2\pi(i-1)}{N}\right)$ all variables observed with $= \sigma^2 \mathbf{I}$, $\sigma^2 = 0.09$. Metric is rms of analysis averaged over time (over 30000 time steps).

Different strategies

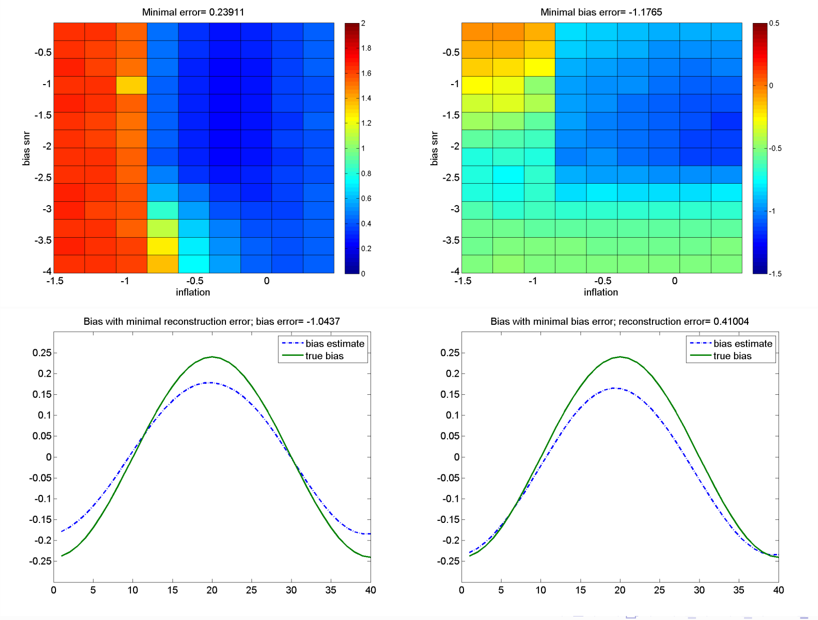
Bias forecast model is persistence in time. Spatial covariance specified by OI with empirical covariance function

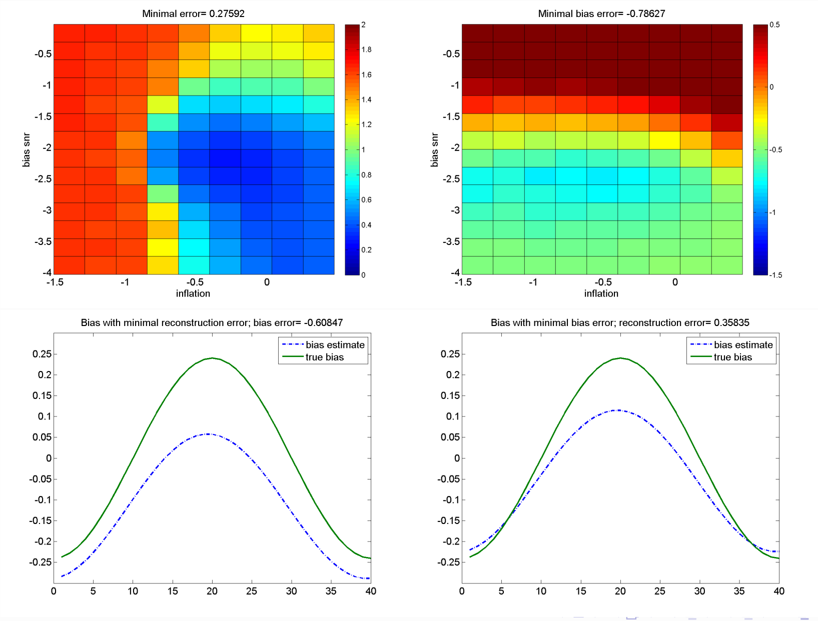
- EnKF for model.
- EnKS for model and constant bias over the smoother window (can be detangled as in EnKS approaches)
- First bias and then model estimate or inverse
- Iterations or not in scale separation
- Full observation or only partial observation
- Different spatial covariance model for bias

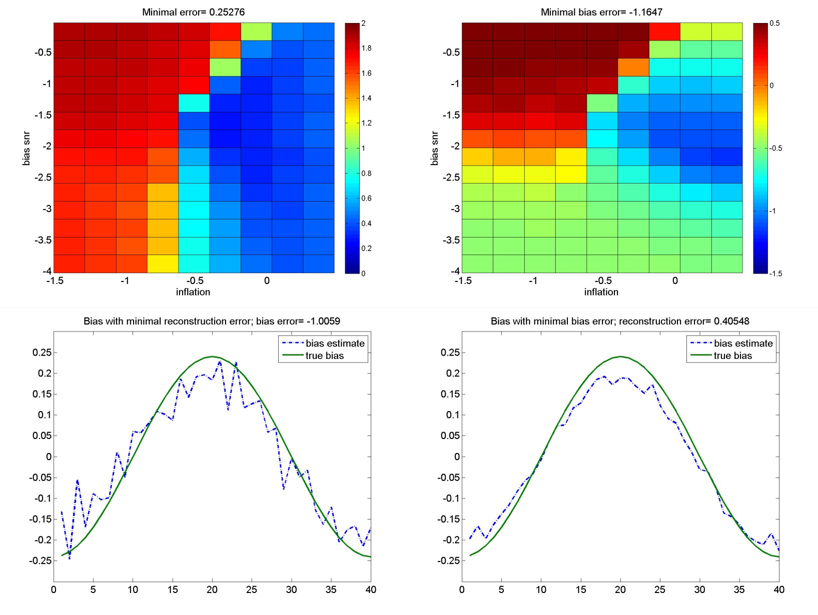

Minimal error= 0.28646
bias snr
inflation
Minimal bias error= -0.88519
bias snr
inflation
Bias with minimal reconstruction error; bias error= -0.75418
bias estimate
true bias
Bias with minimal bias error; reconstruction error= 0.49744
bias estimate
true bias

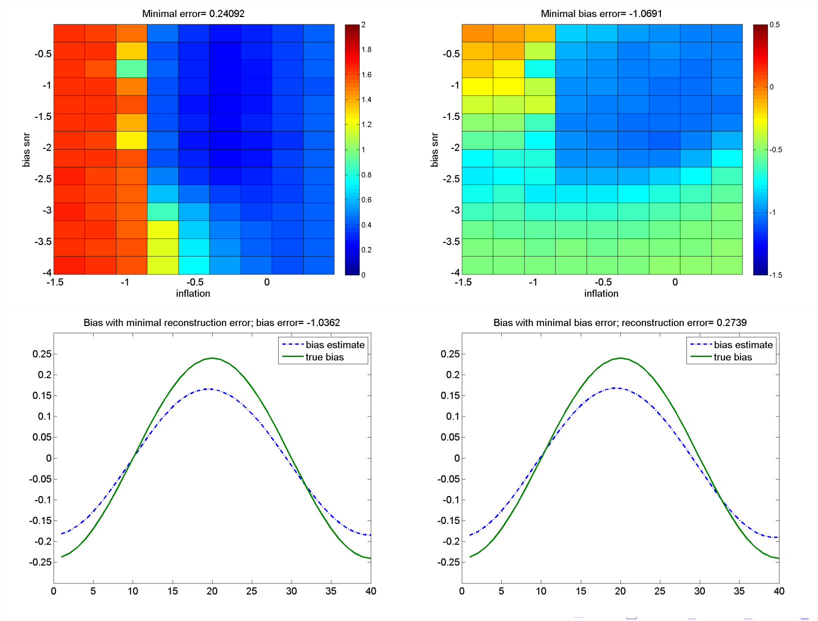


Minimal error= 0.23607
bias snr
inflation
Minimal bias error= -1.0709
bias snr
inflation
Bias with minimal reconstruction error; bias error= -1.0508
bias estimate
true bias
Bias with minimal bias error; reconstruction error= 0.34688
bias estimate
true bias


Minimal error= 0.23911
bias snr
inflation
Minimal bias error= -1.1765
Bias with minimal reconstruction error; bias error= -1.0437
bias estimate
true bias
Bias with minimal bias error; reconstruction error= 0.41004


Minimal error= 0.27592
bias snr
inflation
Minimal bias error= -0.78627
bias snr
inflation
Bias with minimal reconstruction error; bias error= -0.60847
bias estimate
true bias
Bias with minimal bias error; reconstruction error= 0.35835
bias estimate
true bias


Minimal error= 0.25276
Minimal bias error= -1.1647
bias snr
inflation
Bias with minimal reconstruction error; bias error= -1.0059
Bias with minimal bias error; reconstruction error= 0.40548
bias estimate
true bias


Minimal error= 0.24092
bias snr
inflation
Minimal bias error= -1.0691
bias snr
inflation
Bias with minimal reconstruction error; bias error= -1.0362
bias estimate
true bias
Bias with minimal bias error; reconstruction error= 0.2739
bias estimate
true bias

Discussion

1. Results with bias correction in the middle between results with unbiased model and result with biased model uncorrected for bias.
2. Only small differences between EnKF and EnKS? In theory bias at the end is anyway the best estimate ?
3. Iterations on scale separation does (slightly) improve results
4. Method seems more robust then Dee's (in particular with partial observations)
5. Worth continuing ?

References

Balmaseda, M. A., Dee, D., Vidard, A., and Anderson, D. L. T.: A multivariate treatment of bias for sequential data assimilation: Application to the tropical oceans, Quarterly Journal of the 10 Royal Meteorological Society, 133, 167179, doi:10.1002/qj.12, http://dx.doi.org/10.1002/qj. 12, 2007.

Beckers, J.-M., Barth, A., Tomazic, I., and Alvera-Azcárate, A.: A method to generate fully multi-scale optimal interpolation by combining efficient single process analyses, illustrated by a DINEOF analysis spiced with a local optimal interpolation, Ocean Sci., 10, 845-862, doi:10.5194/os-10-845-2014, 2014.

Dee, D. P.: Bias and data assimilation, Quarterly Journal of the Royal Meteorological Society, 131, 33233343, doi:10.1256/qj.05.137, http://dx.doi.org/10.1256/qj.05.137, 2005.