Smoothers: Types and Benchmarks

Patrick N. Raanes

Oxford University, NERSC

8th International EnKF Workshop

May 27, 2013

Chris Farmer, Irene Moroz

Laurent Bertino

Geir Evensen

Abstract

- Talk builds on “Smoothing problems in a Bayesian framework and their linear Gaussian solutions” (Cosme et al., 2012)
- Aim: Introduce new organisation of smoothers
- Aim: Comparisons with less known RTS smoother

Outline

Notation

- Kalman filter equations

$$x_{t|1:t-1} = M_{t-1} x_{t-1|1:t-1}$$
$$P_{t|1:t-1} = M_{t-1} P_{t-1|1:t-1} M_{t-1}^{\intercal} + Q_t$$

$$K_t = P_{t|t-1} H^{\intercal} C^{-1}$$
$$x_{t|1:t} = x_{t|1:t-1} + K_t d_t$$
$$P_{t|1:t} = (I - K_t H) P_{t|t-1}$$

where $C^{-1} = \left(H P_{t|t-1} H^{\intercal} + R_t\right)^{-1}$ and $d_t = (y_t - H x_{t|1:t-1})$.

- PDF of state x conditioned on data y : $p(x|\, y)$
- $x_{1:t} = (x_1;\, x_2;\, \ldots;\, x_{t-1};\, x_t)$
- Estimate for x_t given data $y_{1:t}$: $x_{t|1:t}$

Notation

- Kalman filter equations

$$x_{t|1:t-1} = M_{t-1} x_{t-1|1:t-1}$$
$$P_{t|1:t-1} = M_{t-1} P_{t-1|1:t-1} M_{t-1}^{\intercal} + Q_t$$

$$K_t = P_{t|t-1} H^{\intercal} C^{-1}$$
$$x_{t|1:t} = x_{t|1:t-1} + K_t d_t$$
$$P_{t|1:t} = (I - K_t H) P_{t|t-1}$$

where $C^{-1} = \left(H P_{t|t-1} H^{\intercal} + R_t\right)^{-1}$ and $d_t = (y_t - H x_{t|1:t-1})$.

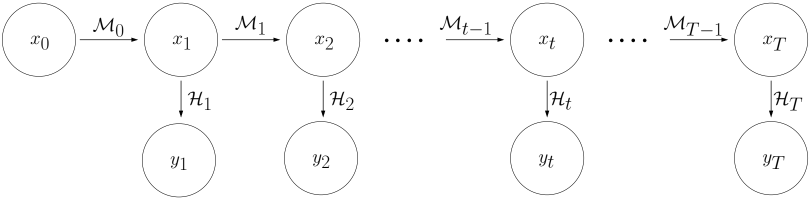
- PDF of state x conditioned on data y : $p(x|\, y)$
- $x_{1:t} = (x_1;\, x_2;\, \ldots;\, x_{t-1};\, x_t)$
- Estimate for x_t given data $y_{1:t}$: $x_{t|1:t}$
- The n^{th} ensemble member: $x^n_{t|1:t}$

  All ensemble members: $E_{t|1:t}$
- Non-linear models denoted by curly symbols: $\mathcal{H}, \mathcal{M}$

Bayesian DA

A useful generalisation

Hidden Markov Chain Model



Bayes' Rule

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

Marginalisation

$$p(x_1) = \int p(x_1, x_2)\, dx_2$$

Bayesian DA

A useful generalisation

3 Methods

- Augmented Filter (AF) smoothers
- Backward-Forward, Rauch-Tung-Striebel (RTS) smoothers
- Two-Filter (2F) smoothers

Bayesian DA

Augmented Filter (AF) smoothers

- Augment state vector: $\tilde{x}_t = \begin{pmatrix} x_t \\ x_{\Sigma_t} \end{pmatrix}$, for some $\Sigma_t \subseteq (0 : t-1)$

Bayesian DA

Augmented Filter (AF) smoothers

- Augment state vector: $\tilde{x}_t = \begin{pmatrix} x_t \\ x_{\Sigma_t} \end{pmatrix}$, for some $\Sigma_t \subseteq (0 : t-1)$
- Find "filter" solution

$$
\begin{aligned}
p(x_t, x_{\Sigma_t} | y_{1:t}) &= \cdots \\
&\propto \; p(y_t|x_t) \int p(x_t|x_{t-1})\, p(x_{t-1}, x_{\Sigma_{t-1}} | y_{1:t-1}) \, dx_{\Sigma_t^c \setminus \Sigma_{t-1}^c},
\end{aligned}
$$

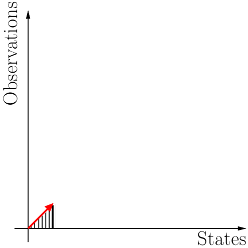
provided $\Sigma_{t-1}^c \subseteq \Sigma_t^c$, where $\Sigma_t^c = \{(0 : t-1) \setminus \Sigma_t\}$.

Bayesian DA

AF – special cases

The Filter

$\Sigma_t = \emptyset$, i.e. $(x_t; x_{\Sigma_t}) = x_t$



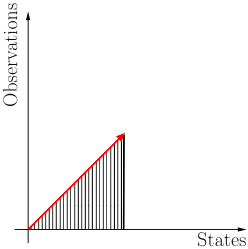
$$p(x_t|y_{1:t}) = p(y_t|x_t) \int p(x_t|x_{t-1})\, p(x_{t-1}|y_{1:t-1})\, dx_{t-1}$$

Bayesian DA

AF – special cases

The Filter

$\Sigma_t = \emptyset$, i.e. $(x_t; x_{\Sigma_t}) = x_t$



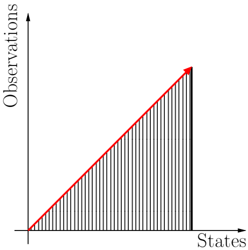
$$p(x_t|y_{1:t}) = p(y_t|x_t) \int p(x_t|x_{t-1})\, p(x_{t-1}|y_{1:t-1})\, dx_{t-1}$$

Bayesian DA

AF – special cases

The Filter

$\Sigma_t = \emptyset$, i.e. $(x_t; x_{\Sigma_t}) = x_t$



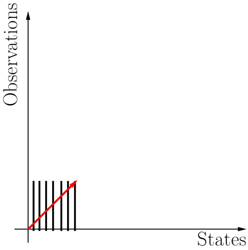
$$p(x_t|y_{1:t}) = p(y_t|x_t) \int p(x_t|x_{t-1})\, p(x_{t-1}|y_{1:t-1})\, dx_{t-1}$$

Bayesian DA

AF – special cases

The Global, Sequential Smoother

$\Sigma_t = (0 : t-1)$, i.e. $(x_t; x_{\Sigma_t}) = x_{0:t}$



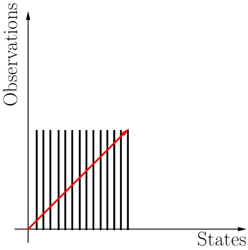
$$p(x_{0:t}|y_{1:t}) = p(y_t|x_t)p(x_t|x_{t-1})\; p(x_{0:t-1}|y_{1:t-1})$$

Bayesian DA

AF – special cases

The Global, Sequential Smoother

$\Sigma_t = (0 : t-1)$, i.e. $(x_t; x_{\Sigma_t}) = x_{0:t}$



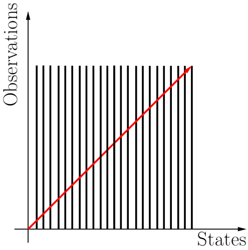
$$p(x_{0:t}|y_{1:t}) = p(y_t|x_t)p(x_t|x_{t-1})\; p(x_{0:t-1}|y_{1:t-1})$$

Bayesian DA

AF – special cases

The Global, Sequential Smoother

$\Sigma_t = (0 : t-1)$, i.e. $(x_t; x_{\Sigma_t}) = x_{0:t}$



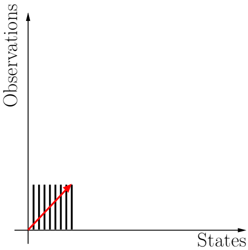
$$p(x_{0:t}|y_{1:t}) = p(y_t|x_t)p(x_t|x_{t-1})\, p(x_{0:t-1}|y_{1:t-1})$$

Bayesian DA

AF – special cases

The Fixed Lag Smoother

$\Sigma_t = (t - L : t{-}1)$, i.e. $(x_t; x_{\Sigma_t}) = x_{t-L:t}$



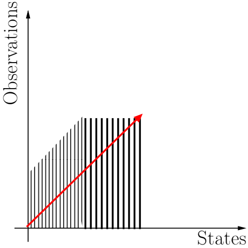
$$p(x_{t-L:t}|y_{1:t}) = p(y_t|x_t) \int p(x_t|x_{t-1})\, p(x_{t-1-L:t-1}|y_{1:t-1})\, dx_{t-1}$$

Bayesian DA

AF – special cases

The Fixed Lag Smoother

$\Sigma_t = (t - L : t{-}1)$, i.e. $(x_t; x_{\Sigma_t}) = x_{t-L:t}$



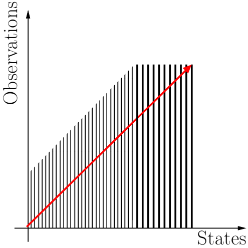
$$p(x_{t-L:t}|y_{1:t}) = p(y_t|x_t) \int p(x_t|x_{t-1})\, p(x_{t-1-L:t-1}|y_{1:t-1})\, dx_{t-1}$$

Bayesian DA

AF – special cases

The Fixed Lag Smoother

$\Sigma_t = (t - L : t{-}1)$, i.e. $(x_t; x_{\Sigma_t}) = x_{t-L:t}$



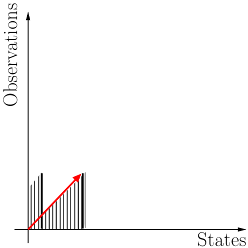
$$p(x_{t-L:t}|y_{1:t}) = p(y_t|x_t) \int p(x_t|x_{t-1})\, p(x_{t-1-L:t-1}|y_{1:t-1})\, dx_{t-1}$$

Bayesian DA

AF – special cases

The "Fixed, Lagging Point" Smoother

$\Sigma_t = t - L$, i.e. $(x_t; x_{\Sigma_t}) = (x_t; x_{t-L})$



Does not satisfy the condition $\Sigma_{t-1}^c \subseteq \Sigma_t^c$

$\implies$ no recursive relation available

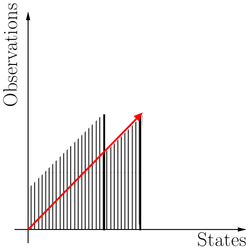
But the fixed-lag smoother should suffice

Bayesian DA

AF – special cases

The "Fixed, Lagging Point" Smoother

$\Sigma_t = t - L$, i.e. $(x_t; x_{\Sigma_t}) = (x_t; x_{t-L})$



Does not satisfy the condition $\Sigma_{t-1}^c \subseteq \Sigma_t^c$

$\implies$ no recursive relation available

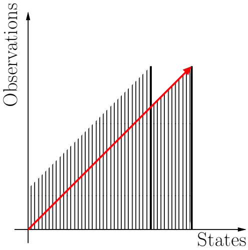
But the fixed-lag smoother should suffice

Bayesian DA

AF – special cases

The "Fixed, Lagging Point" Smoother

$\Sigma_t = t - L$, i.e. $(x_t; x_{\Sigma_t}) = (x_t; x_{t-L})$



Does not satisfy the condition $\Sigma_{t-1}^c \subseteq \Sigma_t^c$

$\implies$ no recursive relation available

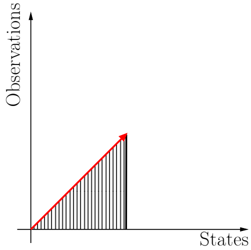
But the fixed-lag smoother should suffice

Bayesian DA

RTS smoother

Seek a recursive, marginal smoother for a fixed analysis window

$$p(x_t|y_{1:T}) = \cdots$$
$$\propto \int p(x_t|x_{t+1}, y_{1:t})\, p(x_{t+1}|y_{1:T})\, dx_{t+1}$$



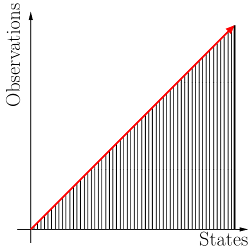
- First runs the filter; Then corrects backwards
- Recursive, but not with new data

Bayesian DA

RTS smoother

Seek a recursive, marginal smoother for a fixed analysis window

$$p(x_t|y_{1:T}) = \cdots$$
$$\propto \int p(x_t|x_{t+1}, y_{1:t})\, p(x_{t+1}|y_{1:T})\, dx_{t+1}$$



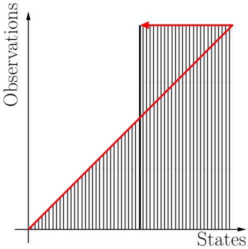
- First runs the filter; Then corrects backwards
- Recursive, but not with new data

Bayesian DA

RTS smoother

Seek a recursive, marginal smoother for a fixed analysis window

$$p(x_t|y_{1:T}) = \cdots$$
$$\propto \int p(x_t|x_{t+1}, y_{1:t})\, p(x_{t+1}|y_{1:T})\, dx_{t+1}$$



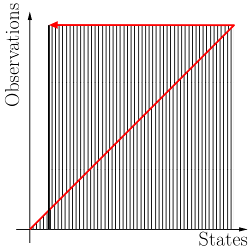
- First runs the filter; Then corrects backwards
- Recursive, but not with new data

Bayesian DA

RTS smoother

Seek a recursive, marginal smoother for a fixed analysis window

$$p(x_t|y_{1:T}) = \cdots$$
$$\propto \int p(x_t|x_{t+1}, y_{1:t})\, p(x_{t+1}|y_{1:T})\, dx_{t+1}$$



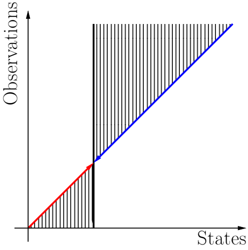
- First runs the filter; Then corrects backwards
- Recursive, but not with new data

Bayesian DA

Two-Filter (2F) smoother

Seek a smoother that directly utilises the filter solution

$$
\begin{aligned}
p(x_t|y_{1:T}) &= \cdots \\
&\propto p(x_t|y_{1:t})\, p(y_{t+1:T}|x_t)
\end{aligned}
$$



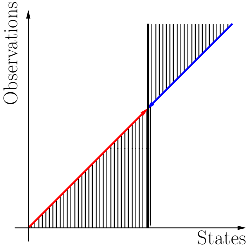
- Recursive, but not with new data
- Involves a backwards likelihood/information filter
  $\Longrightarrow$ Requires adjoint model
- Will not be discussed further

Bayesian DA

Two-Filter (2F) smoother

Seek a smoother that directly utilises the filter solution

$$p(x_t|y_{1:T}) = \cdots$$
$$\propto p(x_t|y_{1:t})\, p(y_{t+1:T}|x_t)$$



- Recursive, but not with new data
- Involves a backwards likelihood/information filter
  $\implies$ Requires adjoint model
- Will not be discussed further

- Bayesian DA
- **Classic Forms**
- Ensemble Formulations
- Benchmarking

Classic Form

- Derivable from the Bayesian formulations assuming Gaussian noise input and linear models
- Can simply insert augmented matrices into the Kalman filter equations

Classic Form

Augmented Filter Smoother

Augmented system matrices:

Observations
$$\tilde{H}_t = (H_t \mid 0 \cdots 0)$$

Fixed Point/Interval
$$\tilde{M}_t = \left(\begin{array}{c|c} M_t & 0 \cdots 0 \\ \hline 0 & \\ \vdots & I \\ 0 & \end{array}\right)$$

Fixed-Lag
$$\tilde{M}_t = \left(\begin{array}{c|c} M_t \; 0 \cdots & 0 \\ \hline & 0 \\ I & \vdots \\ & 0 \end{array}\right)$$

Classic Form

Augmented Filter Smoother

In either case, the other block matrices are

$$\tilde{Q}_t = \left(\begin{array}{c|c} Q_t & 0 \cdots 0 \\ \hline 0 & \\ \vdots & 0 \\ 0 & \end{array}\right), \qquad \tilde{P}_t = \left(\begin{array}{c|c} P_t & P_{t,\Sigma_t} \\ \hline P_{\Sigma_t,t} & P_{\Sigma_t} \end{array}\right),$$

which reduces the Kalman gain to

$$\tilde{K}_{t+1} = \begin{pmatrix} K_{t+1} \\ \\ K_{\Sigma_{t+1}} \end{pmatrix} = \begin{pmatrix} P_{t+1 \,|\, 1:t} \\ \\ P_{\Sigma_{t+1},t+1 \,|\, 1:t} \end{pmatrix} H_{t+1}^{\intercal} C^{-1}$$

Classic Form

Augmented Filter Smoother

Thus, in addition to the usual filter updates, we get

$$x_{\Sigma_t|1:t} = x_{\Sigma_t|1:t-1} + K_{\Sigma_t} d_t$$

$$P_{t,\Sigma_t|1:t} = (I - K_t H) P_{t,\Sigma_t|1:t-1}$$

$$P_{\Sigma_t|1:t} = P_{\Sigma_t|1:t-1} - K_{\Sigma_t} H P_{t,\Sigma_t|1:t-1}$$

Classic Form

RTS Smoother

- Derivable using the usual Linear, Gaussian assumptions
- The recursive, retrospective corrections become

$$
\begin{aligned}
J_t &= P_{t|1:t} M_t^\intercal P_{t+1|1:t}^{-1} \\
x_{t|1:T} &= x_{t|1:t} + J_t \left(x_{t+1|1:T} - x_{t+1|1:t} \right) \\
P_{t|1:T} &= P_{t|1:t} - J_t \left(P_{t+1|1:T} - P_{t+1|1:t} \right) J_t^\intercal
\end{aligned}
$$

- Bayesian DA
- Classic Forms
- **Ensemble Formulations**
- Benchmarking

Ensemble formulation

Augmented Filter Smoother

- Classic form

$$K_{\Sigma_t} = P_{\Sigma_t, t \mid 1:t-1} \, H^\intercal \left(H P_{t|1:t-1} H^\intercal + R \right)^{-1}$$

$$x_{\Sigma_t|1:t} = x_{\Sigma_t|1:t-1} + K_{\Sigma_t} d_t$$

Ensemble formulation

Augmented Filter Smoother

- Classic form

$$K_{\Sigma_t} = P_{\Sigma_t, t \mid 1:t\text{–}1} \; H^\intercal \left(H P_{t|1:t\text{–}1} H^\intercal + R \right)^{-1}$$

$$x_{\Sigma_t|1:t} = x_{\Sigma_t|1:t\text{–}1} + K_{\Sigma_t} d_t$$

- Ensemble formulation

$$\Longrightarrow \begin{cases} \text{EnKF (Evensen, 1994)} \\ \text{EnKS (Evensen and Van Leeuwen, 2000)} \end{cases}$$

Ensemble formulation

Augmented Filter Smoother

- Classic form

$$K_{\Sigma_t} = P_{\Sigma_t, t \mid 1:t-1} \, H^\intercal \left(H P_{t|1:t-1} H^\intercal + R \right)^{-1}$$

$$x_{\Sigma_t|1:t} = x_{\Sigma_t|1:t-1} + K_{\Sigma_t} d_t$$

- Ensemble formulation

$$\Longrightarrow \begin{cases} \text{EnKF (Evensen, 1994)} \\ \text{EnKS (Evensen and Van Leeuwen, 2000)} \end{cases}$$

$$K_{\Sigma_t} = A_{\Sigma_t \mid 1:t-1} \, S_t^\intercal \left(S_t S_t^\intercal + R \right)^{-1}$$

$$x^n_{\Sigma_t|1:t} = x^n_{\Sigma_t|1:t-1} + K_{\Sigma_t} d^n_t,$$

where $A_{t|1:t} = E_{t|1:t} \left[I - \frac{1}{N} \mathbb{1}\mathbb{1}^\intercal \right]$ and $S_t = \mathcal{H}\left(E_{t \mid 1:t-1} \right) \left[I - \frac{1}{N} \mathbb{1}\mathbb{1}^\intercal \right]$.

Ensemble formulation

Augmented Filter Smoother

- Possibility: "Save up" multiple updates before analysing.
  $\implies$ EnS (Van Leeuwen and Evensen, 1996)
  May save significant time because
  $$\max_n \sum_t T_t^n < \sum_t \max_n T_t^n,$$
  where T_t^n is the time required to propagate ensemble member n from one observation point to another.
- Possibility: Process observations sequentially.
  $\implies$ EnSS (Cosme et al., 2012)

Ensemble formulation

RTS Smoother

- Classic form

$$J_t = P_{t|1:t} M_t^\intercal P_{t+1|1:t}^{-1}$$

$$x_{t|1:T} = x_{t|1:t} + J_t \left(x_{t+1|1:T} - x_{t+1|1:t} \right)$$

- Ensemble formulation

$$A_{t+1|t} = U \Sigma V^\intercal$$

$$J_t = A_{t|1:t} V \Sigma^{-1} U^\intercal$$

$$x^n_{t|1:T} = x^n_{t|1:t} + J_t \left(x^n_{t+1|1:T} - x^n_{t+1|1:t} \right)$$

$\Longrightarrow$ EnRTS (Lermusiaux and Robinson, 1999)

- Bayesian DA
- Classic Forms
- Ensemble Formulations
- **Benchmarking**

Classic Form

Marginal and joint smoothers

- Joint and marginal estimates may differ
- But the marginal and joint *means* are always equal
- Besides, the marginal pdfs of a Gaussian RV are just the individual components of the joint pdf

$\Longrightarrow$ Classic smoothers, whether marginal or not, yield the same estimate if the system is linear and Gaussian.

Benchmarking

RMS

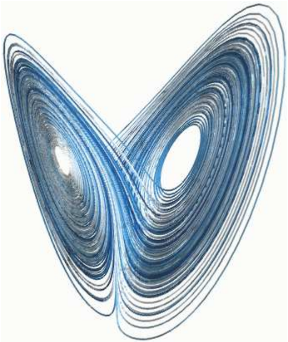
- RMS: $\sqrt{\frac{1}{T}\sum_{t=1}^{T}\|\bar{x}_t - x_t\|^2}$
- Normalised by the RMS of climatology.

Benchmarking

Lorenz'63

A convenient, chaotic, toy model

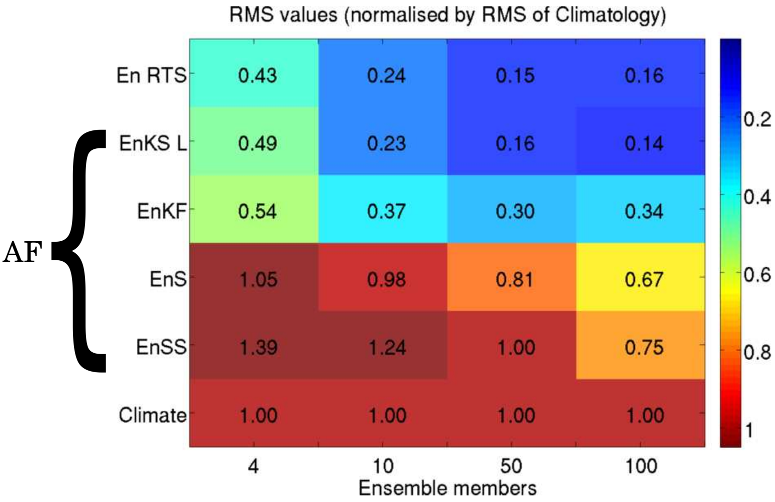
$$\frac{\mathrm{d}x}{\mathrm{d}t} = \sigma(y - x),$$
$$\frac{\mathrm{d}y}{\mathrm{d}t} = x(\rho - z) - y,$$
$$\frac{\mathrm{d}z}{\mathrm{d}t} = xy - \beta z.$$



with $\sigma = 10$, $\beta = 8/3$, $\rho = 28$, initial cond. $(x, y, z)_{t=0} = (1.508870, -1.531271, 25.460910)$, $Q_t = \text{diag}([2, 12.13, 12.31])$. Direct observations of all 3 components and $R_t = 2I$. $T = 10$. Δ_t between observations = 0.25.

Benchmarking

Lorenz'63

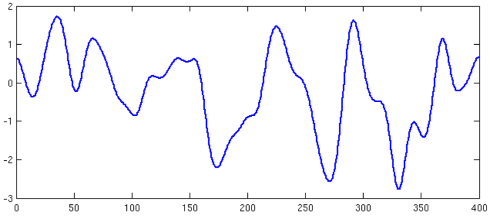


Benchmarking

1D Linear Advection

A linear, high-dimensional toy model

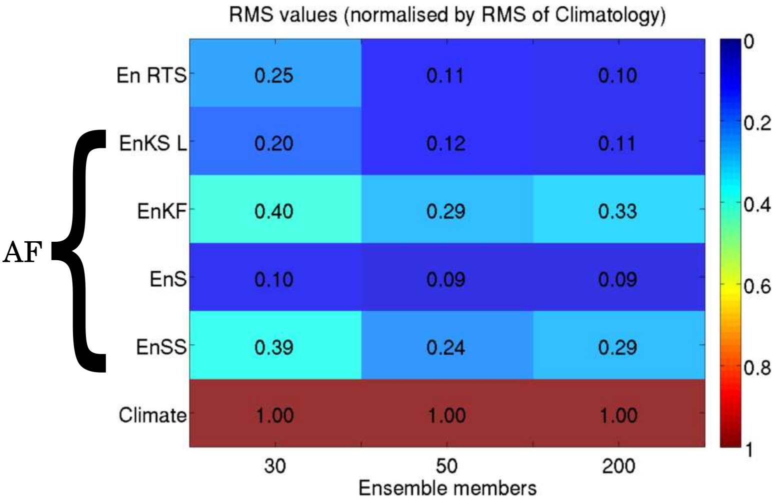
$$\frac{\partial u}{\partial t} + c\frac{\partial u}{\partial x} = 0$$



with $c = \Delta_x = \Delta_t = 1$, a periodic waveform with decorrelation length ≈ 20, a periodic domain with 300 nodes, and no forward model noise. Direct observations 60 nodes apart, assimilated every 10 time steps, with $R_t = 0.01I$. T = 305.

Benchmarking

1D Linear Advection



Benchmarking

Method performance

- EnS superior on LA model
- EnKS and RTS superior to EnS(S) on Lorenz'63
- EnKS and RTS perform similarly on both models
- EnS superior EnS-Sequential
- Difference SQRT/Pert comparatively insignificant

Benchmarking

Costs

- Asymptotic numerical flop costs.
- Not including application of $\mathcal{M}$ and $\mathcal{H}$ to ensemble
- Assumes $N \ll n_x, n_y$. (N: number of ensemble members)

Method	$\sim$ Flops		
EnKF	$TN^2(n_x + n_y)$		0
EnKS	$TN^2(n_x + n_y)$	$+$	$\frac{1}{2}T^2N^2n_x$
EnKS-Lag	$TN^2(n_x + n_y)$	$+$	LTN^2n_x
EnS	0	$+$	TN^2n_x
EnRTS	$TN^2(n_x + n_y)$	$+$	TN^2n_x

Future work

- More models
- More parameter test regimes
- More assessment scores than RMS
- More careful cost analysis
- More DA methods

References

Cosme, E., J. Verron, P. Brasseur, J. Blum, and D. Auroux, 2012: Smoothing problems in a bayesian framework and their linear gaussian solutions. *Monthly Weather Review*, **140**, 683–695.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research*, **99**, 10–10.

— 2009: The ensemble kalman filter for combined state and parameter estimation. *Control Systems, IEEE*, **29**, 83–104.

Evensen, G. and P. Van Leeuwen, 2000: An ensemble kalman smoother for nonlinear dynamics. *Monthly Weather Review*, **128**, 1852–1867.

Jazwinski, A., 1970: *Stochastic processes and filtering theory*, volume 63. Academic press.

Lermusiaux, P. F. and A. Robinson, 1999: Data assimilation via error subspace statistical estimation. part i: Theory and schemes. *Monthly Weather Review*, **127**, 1385–1407.

Skjervheim, J. and G. Evensen, 2011: An ensemble smoother for assisted history matching. *SPE Reservoir Simulation Symposium*.

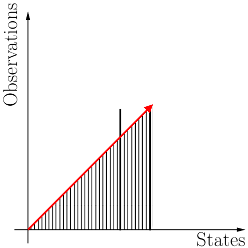
Van Leeuwen, P. J. and G. Evensen, 1996: Data assimilation and inverse methods in terms of a probabilistic formulation. *Monthly Weather Review-USA*, **124**, 2898–2913.

Bayesian DA

AF – special cases

The Fixed Point Smoother

$\Sigma_t = k$, i.e. $(x_t; x_{\Sigma_t}) = (x_t; x_k)$



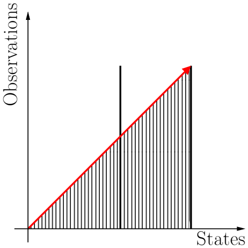
$$p(x_t, x_k | y_{1:t}) = p(y_t | x_t) \int p(x_t | x_{t-1})\, p(x_{t-1}, x_k | y_{1:t-1})\, dx_{t-1}$$

Bayesian DA

AF – special cases

The Fixed Point Smoother

$\Sigma_t = k$, i.e. $(x_t; x_{\Sigma_t}) = (x_t; x_k)$



$$p(x_t, x_k | y_{1:t}) = p(y_t | x_t) \int p(x_t | x_{t-1}) \, p(x_{t-1}, x_k | y_{1:t-1}) \, dx_{t-1}$$