The University of Warwick, Mathematics Institute
Multiscale Inverse Problems
17th – 19th June 2013

A Variational Smoothing Filter for Sequential Inverse Problems

Chris L. Farmer
OCCAM
Mathematical Institute, Oxford

farmer@maths.ox.ac.uk

UNIVERSITY OF OXFORD

OCCAM

Acknowledgements:

- This presentation is based on work supported by Award No. KUK-C1-013-04, made by King Abdullah University of Science and Technology (KAUST).
- Wonjung Lee for many useful discussions.

Topics for discussion

- The problem of sequential assimilation
- A variational smoothing filter (vsf)
- Numerical example - Lorenz '96
- Discussion
- References

Evolution of a probability density function

$\varphi = (\varphi_1, \varphi_2, \ldots, \varphi_N)$

Evolution equation from t^n to t^{n+1}: $\varphi^{n+1} = F^n(\varphi^n)$

$F^n =$ identity for parameters

e.g. as a solution of $\frac{d\varphi_i}{dt} = f_i(\varphi, t)$ with $f_i = 0$ for parameters

φ given at $t = 0$ as a pdf $\pi(\varphi^0)$

Simulation of the observing apparatus:

$s_k^n = h_k(\varphi^n) + \sigma_k w_k^n$ observe at discrete times

Let $S^n = \{s^1, s^2, \ldots, s^n)$ and where $\varphi^n = \varphi(t^n)$ and $t^n = n\Delta$

Problem: Compute the density function $\pi(\varphi, t \mid S^n) \quad t \geq t^n$

Evolution of a probability density function

$$\pi(\varphi^{n+1},\varphi^{n},s^{n+1} \mid S^{n}) = z\,\exp[-\sum_{k\in K}\frac{p_k}{2}(h_k(\varphi^{n+1})-s_k^{n+1})^2]\ \delta(\varphi^{n+1}-F^{n}(\varphi^{n}))\ \pi(\varphi^{n}\mid S^{n})$$

pdf of obs. given the new state

pdf of the old state

pdf of new state given the old state

$z =$ a generic normalisation constant

$$p_k = \frac{1}{\sigma_k^2}$$

F^n is the function implied by the dynamics taking states at t to states at $t+\Delta$

$$\pi(\varphi^{n+1}\mid S^{n+1}) = z\int \pi(\varphi^{n+1},\varphi^{n},s^{n+1}\mid S^{n})\,d\varphi^{n}$$

Gaussian sum representation of pdf's

mixture approximation with an ensemble of R 'centres' φ_r^n

$$\pi(\varphi^n) = \sum_r a_r^n z_r^n \exp[-H(\varphi^n - \varphi_r^n)]$$

where z_r are normalisation constants

weights

$$a_r^n \geq 0. \quad \sum_r a_r^n = 1$$

'energy of the r-th component'

$$H(\varphi^n - \varphi_r^n) = \frac{1}{2} \sum_{i,j} (\varphi_i^n - \varphi_{i,r}^n) L_{ij,r}^n (\varphi_j^n - \varphi_{j,r}^n)$$

Evolution of a probability density function

$$\int \delta(\varphi^{n+1} - F(\varphi^n))\,\mathrm{e}^{-\frac{1}{2}\sum_{i,j}(\varphi_i^n - \varphi_{i,r}^n)L_{ij,r}^n(\varphi_j^n - \varphi_{j,r}^n)}\,d\varphi^n$$

$$= |A^{n+1}|\,\mathrm{e}^{-\frac{1}{2}\sum_{i,j}(F_i^{-1}(\varphi^{n+1}) - \varphi_{i,r}^n)L_{ij,r}^n(F_j^{-1}(\varphi^{n+1}) - \varphi_{j,r}^n)}$$

$$\text{where} \quad |A^{n+1}| = \det\left(\frac{\partial F^{-1}(\varphi^{n+1})}{\partial \varphi^{n+1}}\right)$$

Use the change of variables theorem for multiple integrals and the properties of the delta function.

Assumes that the inverse is unique if the time step is sufficiently short.

Evolution of a probability density function

$$\pi(\varphi^{n+1}, s^{n+1} \mid S^n) =$$

$$\sum_r a_r^n z_r^n \mid A^{n+1} \mid \mathrm{e}^{-\sum_{k\in K} \frac{p_k}{2}(h_k(\varphi^{n+1}) - s_k^{n+1})^2 - \frac{1}{2}\sum_{i,j}(F_i^{-1}(\varphi^{n+1}) - \varphi_{i,r}^n) L_{ij,r}^n (F_j^{-1}(\varphi^{n+1}) - \varphi_{j,r}^n)}$$

An exact expression for the posterior mixture density given a Gaussian mixture density at the previous time step

The Variational Smoothing Filter builds a quadratic approximation to the argument of the exponential

Application of implicit Euler

Using an implicit Euler time discretisation with time step τ

$$\varphi^{\mathrm{n+1}} = \varphi^{\mathrm{n}} + \tau f(\varphi^{\mathrm{n+1}})$$

One finds: $$F^{-1}(\varphi^{\mathrm{n+1}}) = \varphi^{\mathrm{n+1}} - \tau f(\varphi^{\mathrm{n+1}})$$

Also: $$A_{ij}^{\mathrm{n+1}} = \delta_{ij} - \tau \frac{\partial f_i(\varphi^{\mathrm{n+1}})}{\partial \varphi_j^{\mathrm{n+1}}}$$

Notation: $A_r^{\mathrm{n+1}} = A^{\mathrm{n+1}}\Big|_{\varphi_r^{\mathrm{n+1}}}$

Variational Smoothing Filter

- * No adjoints needed
- * Analytical gradient
- * Analytical sparse Hessian
- * Sparse algebra
  - – but needs a drop tolerance in general

A: Update the centres

$$J_r^{n+1}(\varphi^{n+1}) = \sum_{k\in K} \frac{p_k}{2}[h_k(\varphi^{n+1}) - s_k^{n+1}]^2 +$$

$$\frac{1}{2}\sum_{i,j}(\varphi_i^{n+1} - \tau f_i(\varphi^{n+1}) - \varphi_{i,r}^n) L_{ij,r}^n (\varphi_i^{n+1} - \tau f_i(\varphi^{n+1}) - \varphi_{i,r}^n)$$

$$\varphi_r^{n+1} = \arg\min_{\varphi^{n+1}} \; J_r^{n+1}(\varphi^{n+1}) \qquad \left(\left.\frac{\partial J_r^{n+1}(\varphi^{n+1})}{\partial \varphi_i^{n+1}}\right|_{\breve{\varphi}_r^n} = 0 \right)$$

B: Update the precision matrices

$$L_{ij,r}^{n+1} = \left.\frac{\partial^2 J_r^{n+1}(\varphi^{n+1})}{\partial \varphi_i^n \partial \varphi_j^n}\right|_{\varphi_r^{n+1}}$$

Use prior centre as a first iterate & no Newtons == 'ensemble extended Kalman filter'

C: Update the weights

$$\breve{a}_r^{n+1} = a_r^n \left|A_r^{n+1}\right| \frac{\left|L_r^n\right|^{1/2}}{\left|L_r^{n+1}\right|^{1/2}} \exp(-J_r^{n+1}(\varphi_r^{n+1})) \qquad a_r^{n+1} = \varepsilon_w \frac{1}{R} + (1-\varepsilon_w)\frac{\breve{a}_r^{n+1}}{\sum_r \breve{a}_r^{n+1}}$$

Variational Smoothing Filter

Notes :

1. $$L_{ij,r}^{n+1} = \left. \frac{\partial^2 J_r^{n+1}(\varphi^{n+1})}{\partial \varphi_i^n \partial \varphi_j^n} \right|_{\varphi_r^{n+1}}$$

Provides a first order correction – in the time step - to the modified Kalman filter formulae

All of the component matrices are sparse

$$= p_i \delta_{ij} + \sum_{mk} (A_{im,r}^{n+1} L_{mk,r}^n A_{kj,r}^{n+1} - \boldsymbol{\tau} \frac{\partial^2 f_m(\varphi_i^{n+1})}{\partial \varphi_{i,r}^{n+1} \partial \varphi_{j,r}^{n+1}} L_{mk,r}^n (\varphi_{k,r}^{n+1} - \boldsymbol{\tau} f_k(\varphi_r^{n+1}) - \varphi_{k,r}^n))$$

2. Approximate $\dfrac{\left| L_r^n \right|^{1/2}}{\left| L_r^{n+1} \right|^{1/2}} \approx 1$

Example where *h* observes a single component

3. Reset $L_{mk,r}^n = L_{mk,r}^0 + \dfrac{\delta_{mk}}{\varepsilon + \mathrm{var}_m}$ every so many time steps

where $\mathrm{var}_m =$ empirical, ensemble variance of φ_m^{n+1}

Choosing L^0 via local random fields

clf 2007

$$H(\psi) = \frac{1}{2}\int [a\psi^2 + b(\nabla\psi)^2 + c(\nabla^2\psi)^2]\,d\omega \quad = \quad \frac{1}{2}\int [\psi L \psi]\,d\omega$$

where $L = a - b\nabla^2 + c\nabla^2\nabla^2$

Helmholtz Green's functions

$$g(r) = \frac{e^{-r\sqrt{\frac{a}{b}}}}{4\pi r b} \qquad \text{3D}$$

$$g(r) = \frac{e^{-r\sqrt{\frac{a}{b}}}}{4\sqrt{ab}} \qquad \text{1D}$$

$$\pi(\varphi) = z\exp(-H(\varphi - \bar{\varphi}))$$

$$g(x-y) := \langle \varphi(x)\varphi(y) \rangle$$

$Theorem:$ $Lg(x-y) = \delta(x-y)$

After discretisation L_{ij} is sparse and $C = (L)^{-1}$ is the covariance matrix

$L = a - b\nabla^2$ - the 'Helmholtz precision matrix' - is particularly convenient and was used in the numerical experiments on Lorenz-96

Set $L_r^0 = wL$ for some 'sharpness control' w that increases with R

Numerical example: Lorenz 96

$$\frac{du_i}{dt} = 0 \quad \text{: 'reality' and forward model}$$

$$\frac{dv_i}{dt} = (v_{i+1} - v_{i-2})v_{i-1} - v_i(u_i + e^{\varepsilon v_i^2} - 1) + F_i \quad \text{: 'reality' and forward model}$$

$$\varphi = (u, v)$$

See Yang & DelSole 2009, 2010 for related work

Reality:

$u_i = 0.5 + 0.2 \sin(0.3\, \mathrm{i}) + 0.02\, (0.5 - \zeta'_i);\quad \varepsilon_L = 1.0,\quad F_i = 10.0$

$v_i(0) = 10 + 2\zeta''_i \qquad \zeta_i, \zeta'_i, \zeta''_i \sim N(0,1)$

No parameters are observed. 1000 equilibration steps before observing

Observe variables at: $i = 1, 6, \ldots$ then every other five with $\sigma_i^2 = 0.01$

Initialised with balanced sampling (mirror images of each realisation about the mean)

Cor. lengths. 20.0 (parms) & 0.1 (vars) Init. var 0.1 parm and 1.0 var

Implicit Euler, time step = 0.05 between obs, and

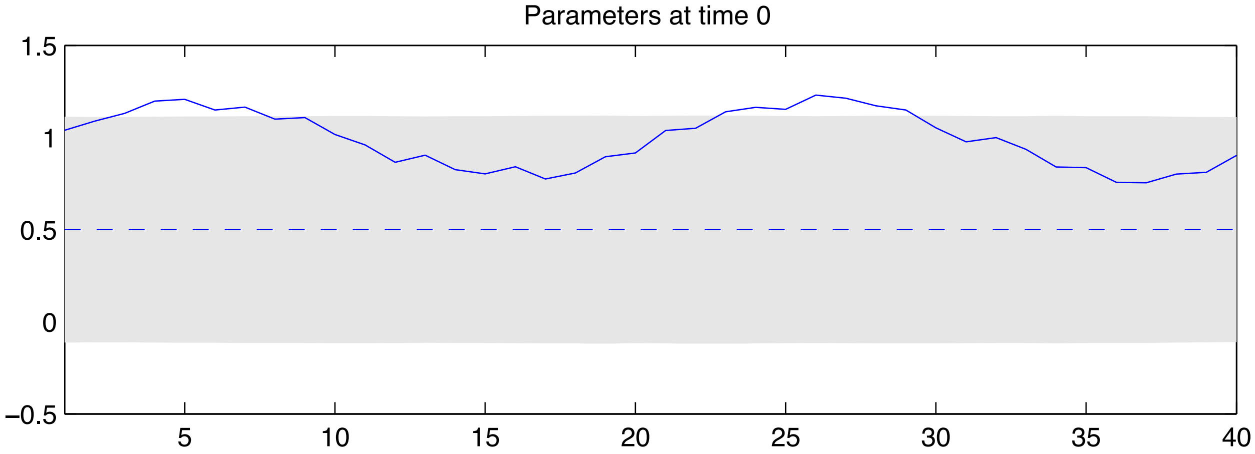
10 Newtons time stepping and optimisation

Lorenz 96 example – *R* = 50: initial condition

Balanced initial ensemble

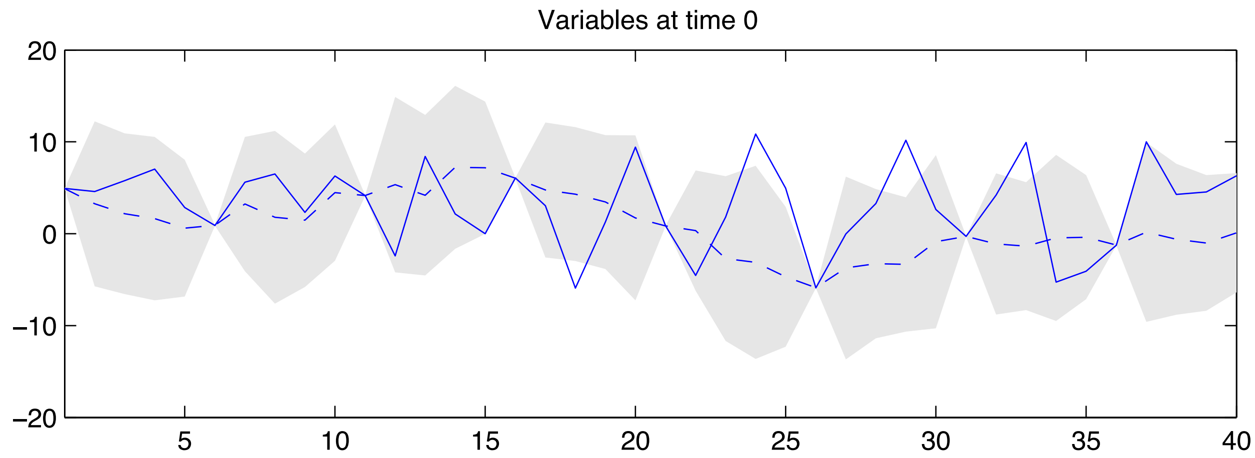
2 fields
40 parameters
40 variables

Parameters
prior correlation
length = 20.0



after 1000
Equilibration steps

vsf
compared
with reality



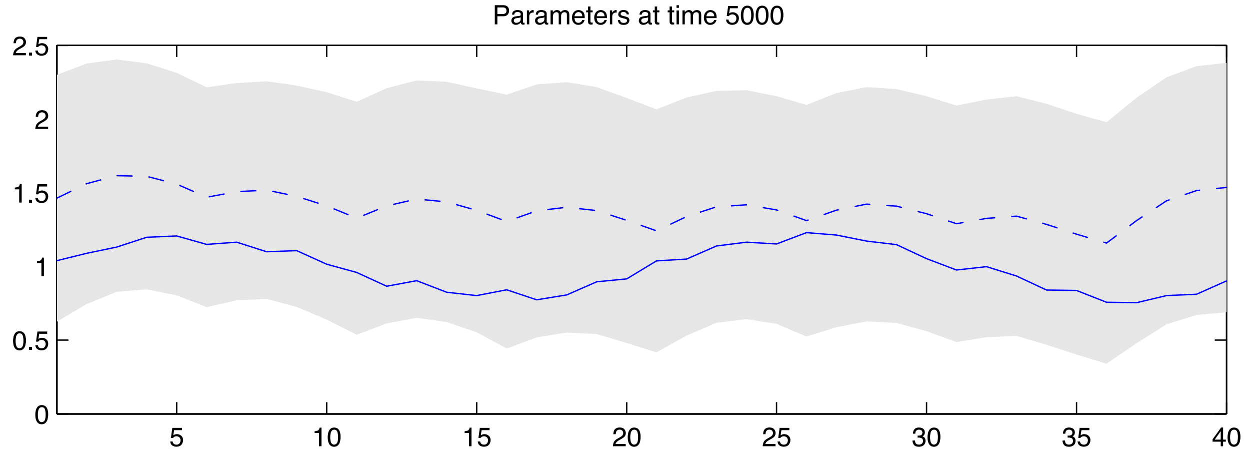
Variables
prior correlation
length = 0.1

Solid blue lines = ‘reality’. Dotted blue line = ‘ensemble mean’. Shading +/- 1 stand. dev.

Lorenz 96 example – R = 50: 5,000 time steps

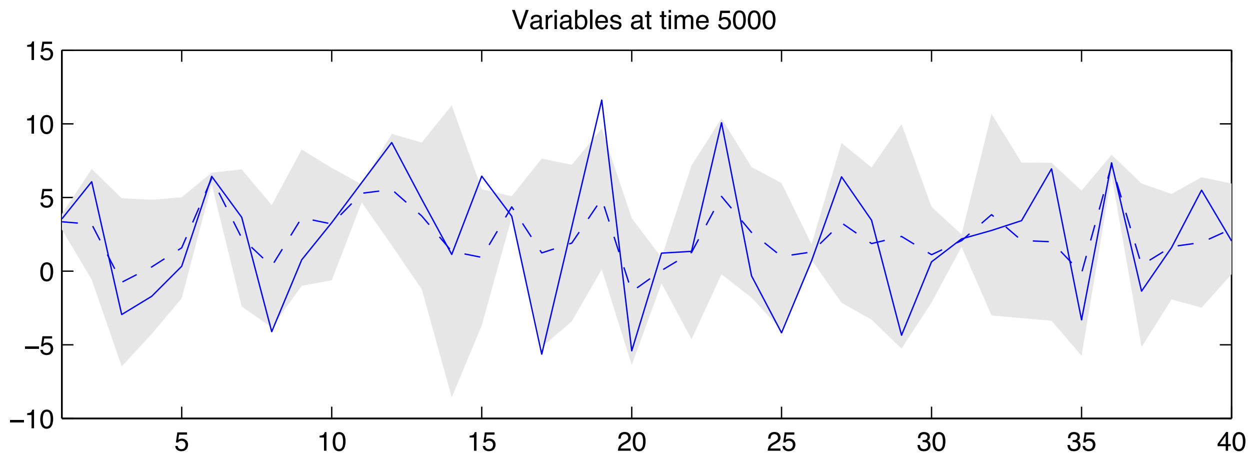
2 fields
40 parameters
40 variables

Parameters
prior correlation
length = 20.0



vsf
compared
with reality

Variables
prior correlation
length = 0.1

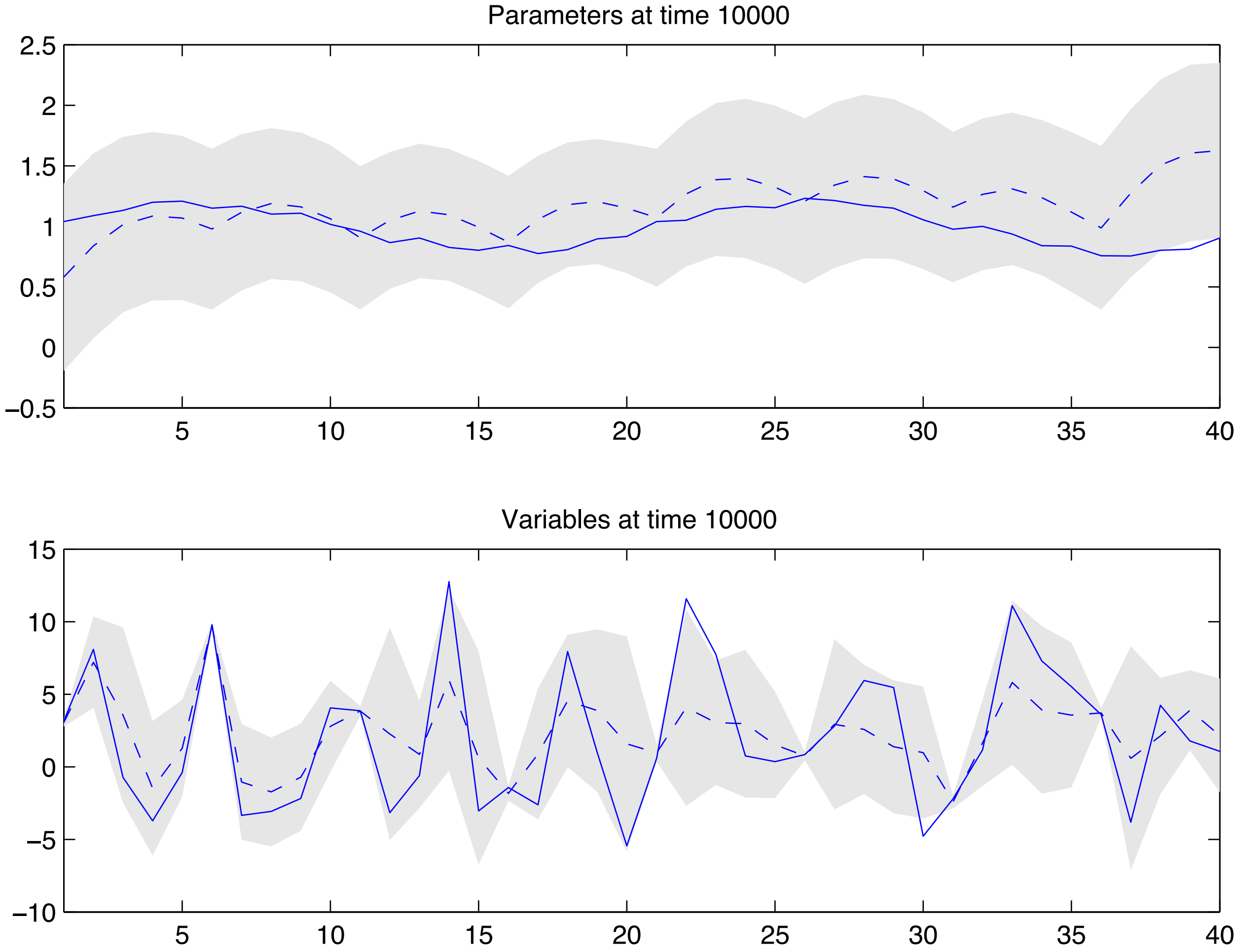


Solid blue lines = ‘reality’. Dotted blue line = ‘ensemble mean’. Shading +/- 1 stand. dev.

Lorenz 96 example – R = 50: 10,000 time steps

2 fields
40 parameters
40 variables

Parameters
prior correlation
length = 20.0



vsf
compared
with reality

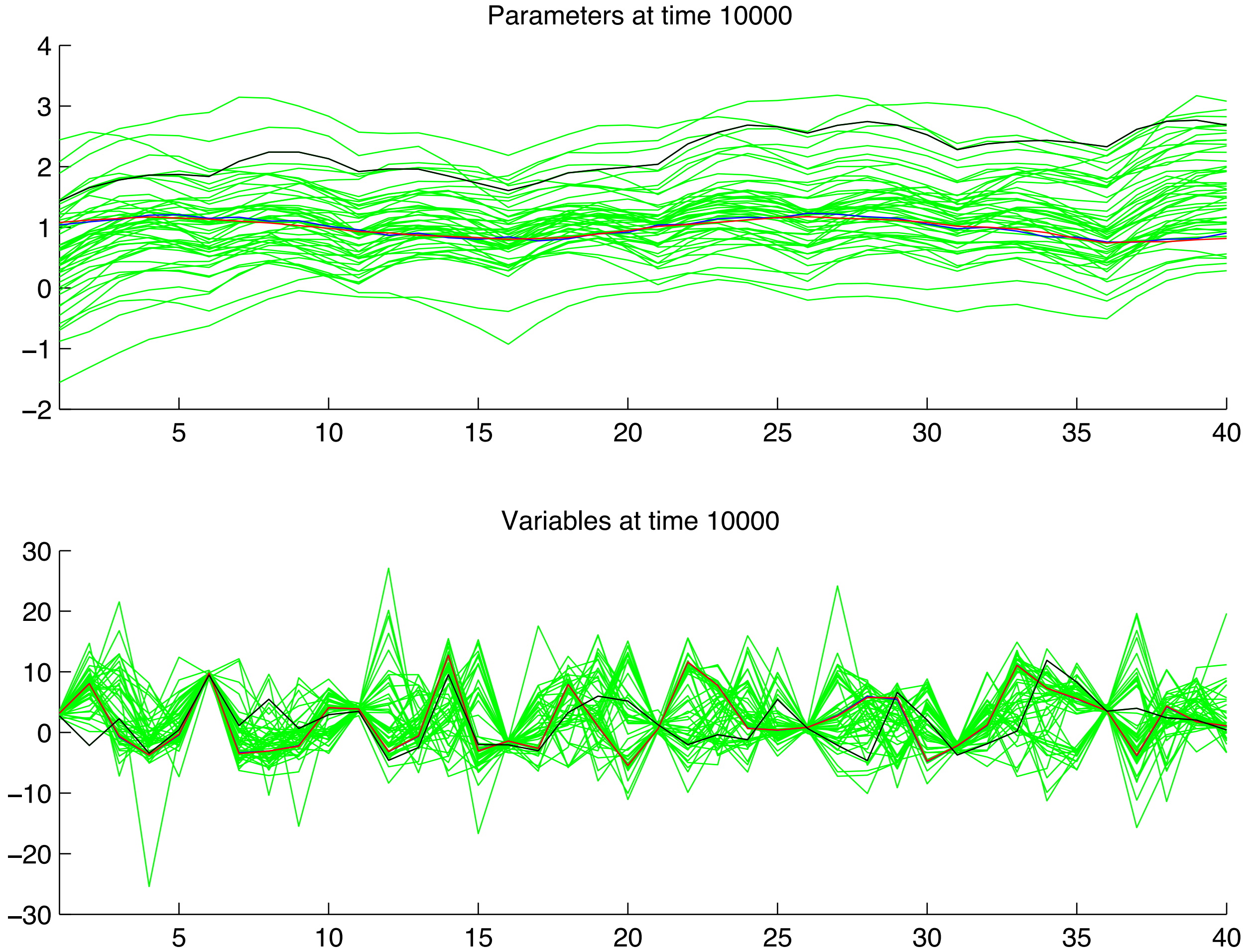
Variables
prior correlation
length = 0.1

Solid blue lines = 'reality'. Dotted blue line = 'ensemble mean'. Shading +/- 1 stand. dev.

Lorenz 96 example – R = 50: 10,000 time steps

2 fields
40 parameters
40 variables

Parameters
prior correlation
length = 20.0



vsf
compared
with reality

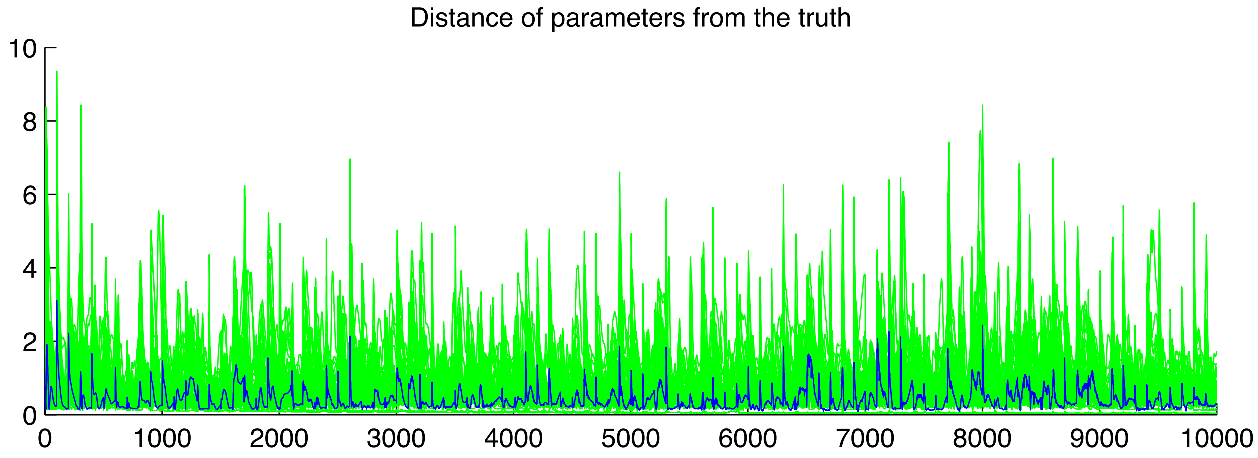
Variables
prior correlation
length = 0.1

Solid blue lines = 'reality'. Green lines = centres. Red/black = most/least probable centre

Lorenz 96 example – *R* = 50: 10,000 time steps

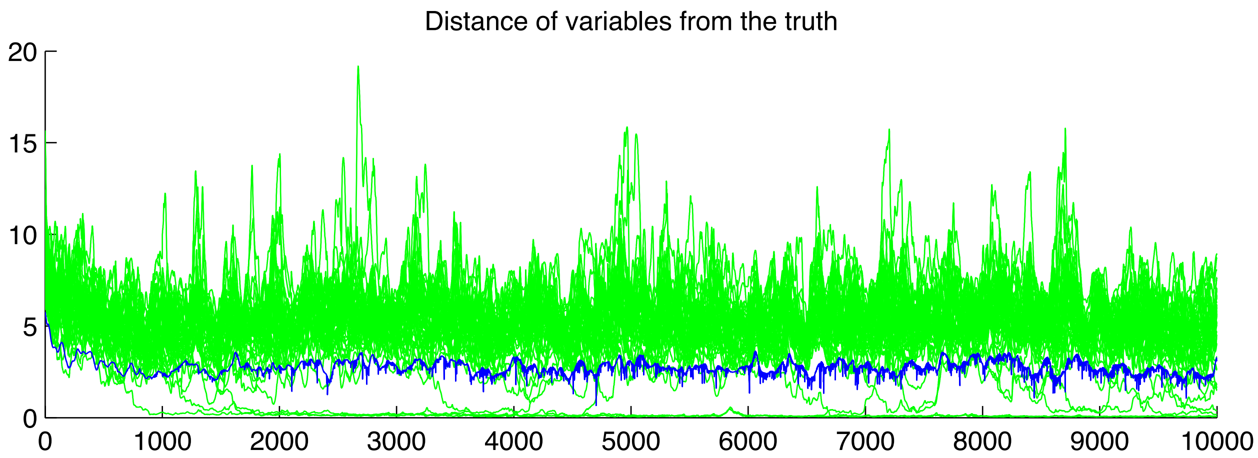
2 fields
40 parameters
40 variables

Parameters
prior correlation
length = 20.0



vsf
compared
with reality

Variables
prior correlation
length = 0.1



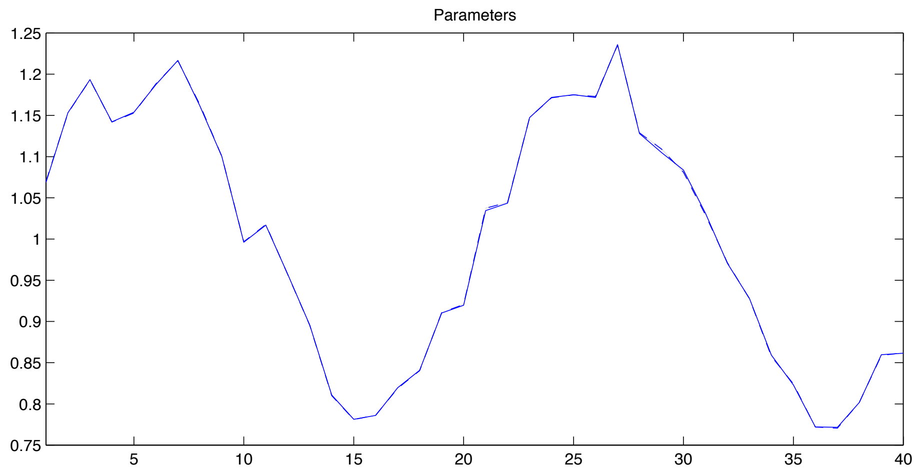
Root mean square distance from 'reality':
Solid blue lines = ensemble mean. Green lines – the centres.

Lorenz 96 – *R* = 4: Obs. every third variable

Balanced initial ensemble

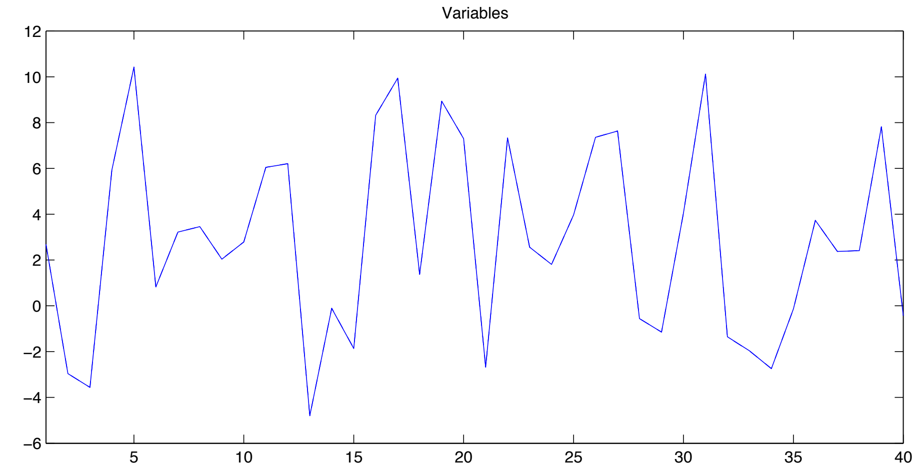
2 fields
40 parameters
40 variables

Parameters
prior correlation
length = 4.0



vsf
compared
with reality

Variables
prior correlation
length = 0.1



After 10000
steps the ensemble
converged to
the true values

Solid blue lines = 'reality'. Dotted blue line = 'ensemble mean'. Shading +/- 1 stand. dev.

Concluding remarks

- The aim of filtering theory is to *compute a best approximation to the posterior density*.
- The combination of variational and principled ensemble methods is promising. Even the first iterate – the 'ensemble modified Kalman Filter' works well in some cases.
- A 'principled' approach should allow us to control the numerical errors so that we can concentrate on the important questions such
  (i) how good is our forward model?
  (ii) how sensitive are we to the initial prior?
  (iii) what observations would reduce sensitivity to the initial prior?

References

- A. O'Hagan and J. Forster: Kendall's Advanced Theory of Statistics. Vol 2B, Bayesian Inference. 2nd Edition. Arnold Publishers, 2004.
- B.D.O. Anderson & J.B. Moore: Optimal Filtering, Prentice Hall 1979. (Dover 2005).
- A.H. Jazwinski: Stochastic Processes and Filtering Theory. Academic Press, 1970.
- C.L. Farmer: Bayesian Field Theory Applied to Scattered Data Interpolation and Inverse Problems. pp. 147-166. Algorithms for Approximation. Editors. A. Iske and J. Levesley, Springer-Verlag, Heidelberg, 2007.
- Inverse Theory for Petroleum Reservoir Characterization and History Matching. D. S. Oliver, A. C. Reynolds and N. Liu. Cambridge University Press, 2008.
- Silverman, B.W. 1986 Density Estimation for Statistics and Data Analysis, Chapman and Hall.
- Stordal, A.S., Karlsen, H.A., Nævdal, G., Skaug, H.J., &Vallès, B. 2011 Bridging the ensemble Kalman filter and particle filters: the adaptive Gaussian mixture filter. Computational Geosciences, 15(2), pp 293-305.
- Yang, X., and DelSole, T. 2009 Using the ensemble Kalman Filter to estimate multiplicative model parameters. Tellus, 61, pp 601-609.
- DelSole, T. and Yang, X. 2010 State and parameter estimation in stochastic dynamical models. Physica D, 239, pp 1781-1788.
- C.L. Farmer. 2013. A Variational Smoothing Filter for State Estimation. In preparation.