Localization Radius

Yicun Zhen

Sep 13, 2013

Outline

Steps

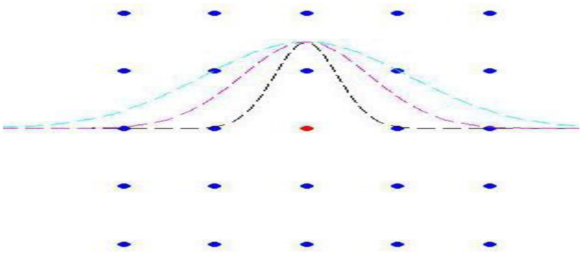
- Define a cost function F
- compute the value of cost function for several different localization parameter λ
- choose the λ that gives the least value of F
- use that λ to do sequential localized EnKF

Question

How to define the cost function λ?

Definition in last time

Fix the influence radius for each observation y_i^o, and compute the mean difference of the updates for the localized and un-localized sequential EnKF.



cost function in the last time

$$F(\lambda, i) = \int \sum_{j=1}^{n} (r_{ij}^{s} \rho_{\lambda}(d_{ij}) - r_{ij})^2 p(B|S) dB$$

where n is the number of gridpoints in the neighbor region in consideration.

notation

r_{ij}^s is the sample regression coefficient of using observation y_i^o to update mean state at grid point x_j

$$X_j^{mean} \leftarrow X_j^{mean} + r_{ij}^s \Delta y_i$$

r_{ij} is the true regression coefficient of using observation y_i^o to update mean state at grid point x_j

$$X_j^{mean} \leftarrow X_j^{mean} + r_{ij} \Delta y_i$$

Critical problem with last cost function

The resulting λ is always equal to 0 which means no localization is needed.

New cost function

$F(\lambda, n) = F_V(\lambda, n) - F_E(\lambda, n)$

where

- F_E is an updating effect function
- F_V is a pseudo-variance function

Definition of F_E and F_V

$$F_E(\lambda, n) = C(S)\left\{ \int \sum_{i=2}^{n} [r_i^2 - (r_i^s \rho_\lambda(d_i) - r_i)^2] p(B^s|B) p(B) dB \right\} \quad (1)$$

where $C(S)$ is a constant depending only on the sample and n such that:

$$C(S) \int p(B^s|B) p(B) dB = 1 \quad (2)$$

And we define the variance-like function:

$$F_V(\lambda, n) = C(S)\left\{ \int \sum_{i=2}^{n} \rho_\lambda(d_i)^2 (r_i^s - r_i)^2 p(B^s|B) p(B) dB \right\} \quad (3)$$

Theorem

$$F_E(\lambda, n) = \sum_{i=2}^{n} \left\{ 2\rho_\lambda(d_i)\theta_i e_i \frac{\int_0^\infty e^{-\frac{1}{2}t_1^2} \frac{t_1^{N-n-1}}{rt_1^2+|\epsilon_1|^2} dt_1}{\int_0^\infty e^{-\frac{1}{2}t_1^2} t_1^{N-n-1} dt_1} - \rho_\lambda^2(d_i)\theta_i^2 \right\} \quad (4)$$

$$F_V(\lambda, n) = \sum_{i=2}^{n} \rho_\lambda^2(d_i) \left\{ \theta_i^2 - 2\theta_i e_i \frac{\int_0^\infty e^{-\frac{1}{2}t_1^2} \frac{t_1^{N-n-1}}{rt_1^2+|\epsilon_1|^2} dt_1}{\int_0^\infty e^{-\frac{1}{2}t_1^2} t_1^{N-n-1} dt_1} + \left(\frac{\Delta_{ii}|\epsilon_1|^2}{N-n-1} + e_i^2 \right) \frac{\int_0^\infty e^{-\frac{1}{2}t_1^2} \frac{t_1^{N-n-1}}{(rt_1^2+|\epsilon_1|^2)^2} dt_1}{\int_0^\infty e^{-\frac{1}{2}t_1^2} t_1^{N-n-1} dt_1} \right\} \quad (5)$$

Algorithm

- Have some value of λ in mind.
- For each observation $y_i^o$and for those λ, find the influence radius (hence n_λ)for each λ compute the value of the new cost function F for each λ and find the λ that corresponds to the minimum F value.
- Hence for each observation y_i^o we have a unique λ_i that is optimal for y_i^o
- Use any method (for example, kernel density estimation) to find the maximum likelihood of λ
- use the maximum likelihood λ to be the λ we use in this assimilation cycle.

computational cost

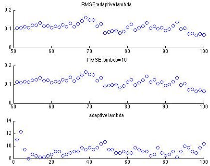
$O(N^3 m)$

Notations

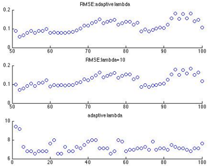
Model: Lorentz 96.
n: number of variables.
m:number of observations.
N: ensemble size.

Set 1

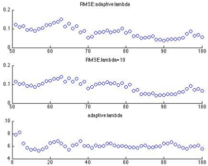
n=40,m=20,N=31,41,51,61



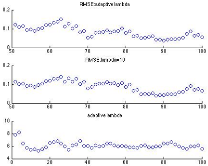
(a) N=31



(b) N=41



(c) N=51



(d) N=61

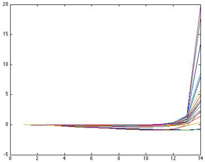
Conclusion:
larger ensemble size

$\Rightarrow$

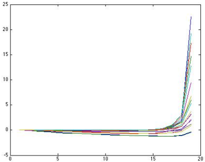
larger localization radius

Set 1

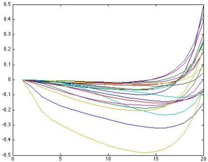
n=40,m=20,N=31,41,51,61



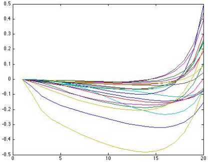
(d) N=31



(e) N=41



(f) N=51

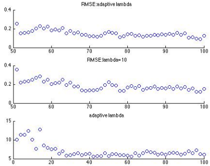


(g) N=61

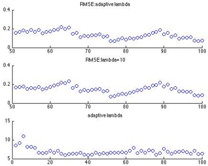
Plot of *F* values at different observation

Set 2

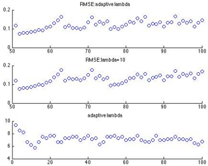
n=48,N=51,m=12,16,24



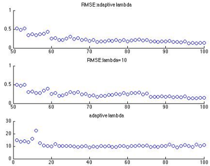
(g) m=12,N=51



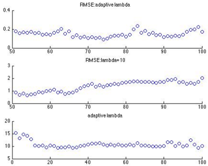
(h) m=16,N=51



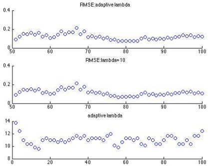
(i) m=24,N=51



(j) m=12,N=31



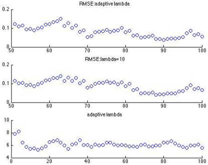
(k) m=16,N=31



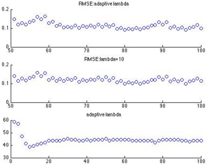
(l) m=24,N=31

Set 3

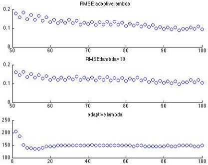
N=51,n=40,300,1000, m=20,150,500



(m) m=20,n=40



(n) m=150,n=300



(o) m=500,n=1000