Simultaneous estimation of inflation and observation errors

Eugenia Kalnay
Hong Li
Takemasa Miyoshi

University of Maryland

Motivation

- Any data assimilation scheme requires accurate statistics for the observation and background errors. Unfortunately those statistics are not known and are usually tuned or adjusted by gut feeling.

- Ensemble Kalman filters need inflation (additive or multiplicative) of the background error covariance, but

1) Tuning the inflation parameter is expensive especially if it is regionally dependent, and it may depend on time

2) Miyoshi and Kalnay 2005 (MK) proposed a technique to objectively estimate the covariance inflation parameter.

3) This method works, but only if **the observation errors are known**.

- Here we introduce a method to simultaneously estimate observation errors and inflation.

MK method to estimate the inflation parameter (Miyoshi 2005, Miyoshi&Kalnay 2005, unpublished)

$$\mathbf{d}_{o-b} = \mathbf{y}_o - H(\mathbf{x}^b)$$

obs. increment in obs. space

$$< \mathbf{d}_{o-b}\mathbf{d}_{o-b}^T >= (\mathbf{1} + \Delta)\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R}$$

Should be satisfied if R, P^b and Δ are correct (they are not!)

So, at any given analysis time, and computing the inner product

$$\mathbf{d}_{o-b}^T\mathbf{d}_{o-b} = (1 + \Delta^o)Tr(\mathbf{H}\mathbf{P}^b\mathbf{H}^\mathbf{T}) + Tr(\mathbf{R})$$

$$\Delta^o = \frac{(\mathbf{d}_{o-b}^T\mathbf{d}_{o-b}) - Tr(\mathbf{R})}{Tr(\mathbf{H}\mathbf{P}^b\mathbf{H}^\mathbf{T})} - 1 \qquad \text{(1a)}$$

←

Assumption: R is known

This gives an "observation" of Δ

We use the "observation" of inflation to update the inflation online with a simple KF (adaptive inflation)

$$\Delta^o = \frac{(\mathbf{d}_{o-b}^T \mathbf{d}_{o-b}) - Tr(\mathbf{R})}{Tr(\mathbf{HP}^b\mathbf{H}^\mathrm{T})} - 1 \qquad (1a)$$

Assumption: **R** is known

This gives an "observation" of Δ

Online estimation: use scalar KF (adaptive regression), Kalnay 2003, App. C:

$$\Delta^a = \frac{v^o \Delta^f + v^f \Delta^o}{v^o + v^f} \qquad v^a = (1 - \frac{v^f}{v^f + v^o})v^f$$

where $\quad \Delta^f{}_{t+1} = \Delta^a{}_t \qquad v^f{}_{t+1} = (1+0.03)v^a$

This scalar KF is used for all the online estimation experiments discussed here

The MK method works very well to estimate the optimal inflation Δ if **R** is correct, but it fails if **R** is wrong: one equation (1a) with two unknowns…

Diagnosis of observation error statistics

(Desroziers et al, 2006, Navascues et al, 2006)

Desroziers et al, 2006, introduced two new statistical relationships:

$$< \mathbf{d}_{o-a}\mathbf{d}^{T}_{o-b} >= \mathbf{R}$$

$$< \mathbf{d}_{a-b}\mathbf{d}_{o-b}{}^{T} >= \mathbf{HBH}^{T}$$

if the **R** and **B** statistics are correct and errors are uncorrelated

Writing their inner products we obtain two more equations which we can use to "observe" **R** and Δ :

$$(\tilde{\sigma}_o)^2 = \mathbf{d}^{T}_{o-a}\mathbf{d}_{o-b} / p = \sum_{j=1}^{p} (y_j^o - y_j^a)(y_j^o - y_j^b) / p \qquad (2)$$

$$\Delta^o = \mathbf{d}^{T}_{a-b}\mathbf{d}_{o-b} / Tr(\mathbf{HP^fH}^T) - 1 = \sum_{j=1}^{p} (y_j^a - y_j^b)(y_j^o - y_j^b) / Tr(\mathbf{HP^fH}^T) - 1 \qquad (1b)$$

(an alternative to MK's online "observation" of the inflation factor)

Diagnosis of observation error statistics
(Desroziers et al, 2006, Navascues et al, 2006)

$$\Delta^o = \frac{(\mathbf{d}_{o-b}^T \mathbf{d}_{o-b}) - Tr(\mathbf{R})}{Tr(\mathbf{H}\mathbf{P}^b\mathbf{H}^\mathrm{T})} - 1 \qquad \text{(1a)}$$

$$\Delta^o = \sum_{j=1}^{p} (y_j^a - y_j^b)(y_j^o - y_j^b) / Tr(\mathbf{H}\mathbf{P}^\mathbf{f}\mathbf{H}^T) - 1 \qquad \text{(1b)}$$

$$(\tilde{\sigma}_o)^2 = \mathbf{d}_{o-a}^T \mathbf{d}_{o-b} / p = \sum_{j=1}^{p} (y_j^o - y_j^a)(y_j^o - y_j^b) / p \qquad \text{(2)}$$

Desroziers et al. (2006) and Navascues et al. (2006) have used these relations in a diagnostic mode, from past 3D-Var/4D-Var stats.

Here we use the simple KF to estimate both Δ and R online.

Tests within LETKF with Lorenz-40 model

In all these experiments we assume true Rt=1

Observations every 3 steps. Optimally tuned rms=0.264

Perfect R, estimate inflation using (1a) or (1b): both work

Δ method	R^s	Δ	rms
(1b)	1	0.096	0.265
(1a)	1	0.101	0.270

Wrong R, estimate inflation using (1a) or (1b): they both fail

Δ method	R^s	Δ	rms
(1b)	10.0	0.002	0.799
(1a)		0.008	1.088

Tests within LETKF with L40 model

Now we estimate observation error and optimal inflation simultaneously using (1a) or (1b) and (2): it works!

R method	Δ method	R_{init}	Estimated R	Estimated Δ	rms
(2)	(1b)	0.1	0.999	0.098	0.263
	(1a)		1.001	0.101	0.265
	(1a)	10.0	1.001	0.097	0.266
	(1b)		0.999	0.100	0.265

Tests within LETKF with SPEEDY

SPEEDY MODEL (Molteni 2003)

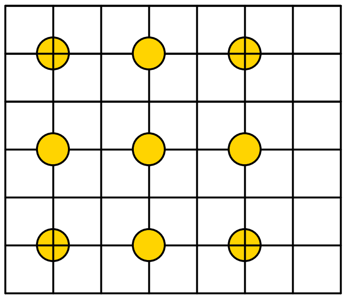
- Primitive equations, T30L7 global spectral model
- total 96x46 grid points on each level
- State variables u,v,T,Ps,q

Data Assimilation: Local ensemble transform Kalman filter (LETKF, Hunt et al. 2006)

Tests within LETKF with SPEEDY

OBSERVATIONS

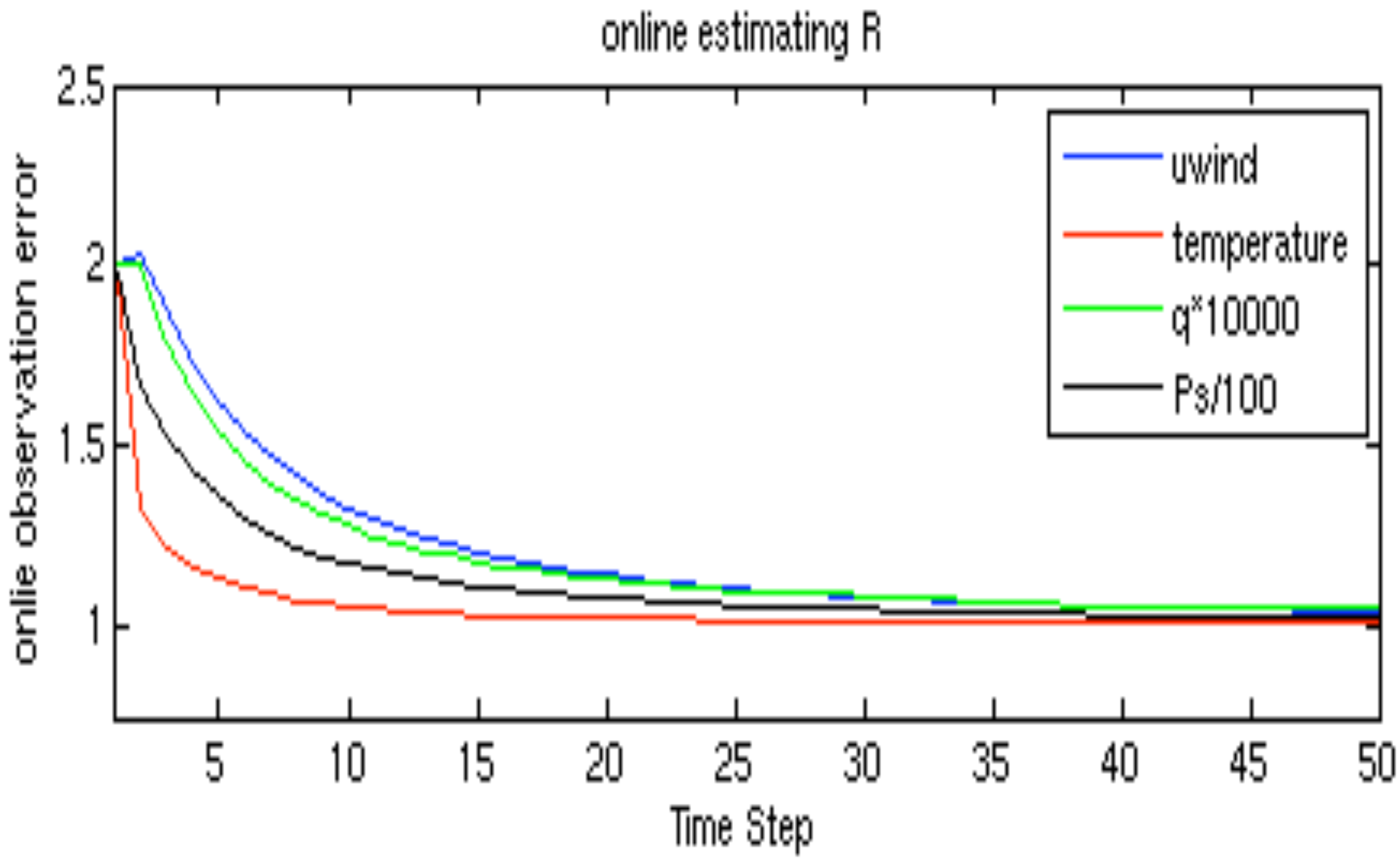
- Generated from the 'truth' plus "random errors" with error standard deviations of 1 m/s (u), 1 m/s(v), 1K(T), 10^{-4} kg/kg (q) and 100Pa(Ps).
- Dense observation network: 1 every 2 grid points in x and y direction



EXPERIMENTAL SETUP

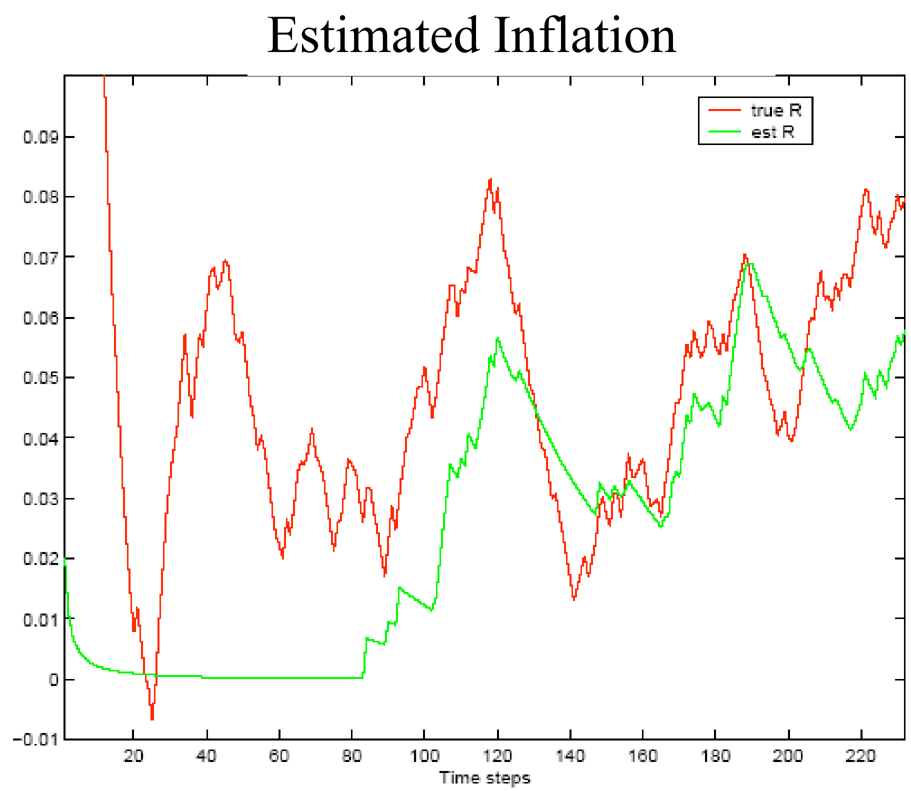
- Run SPEEDY with LETKF for two months (January and February 1982) , starting from wrong (doubled) observational errors of 2 m/s (u), 2 m/s(v), 2K(T), $2*10^{-4}$ kg/kg (q) and 200Pa(Ps).
- Estimate and correct the observational errors and inflation adaptively.

online estimated observational errors



The original wrongly specified R converges to the right R quickly (in about 5-10 days)

Estimation of the inflation



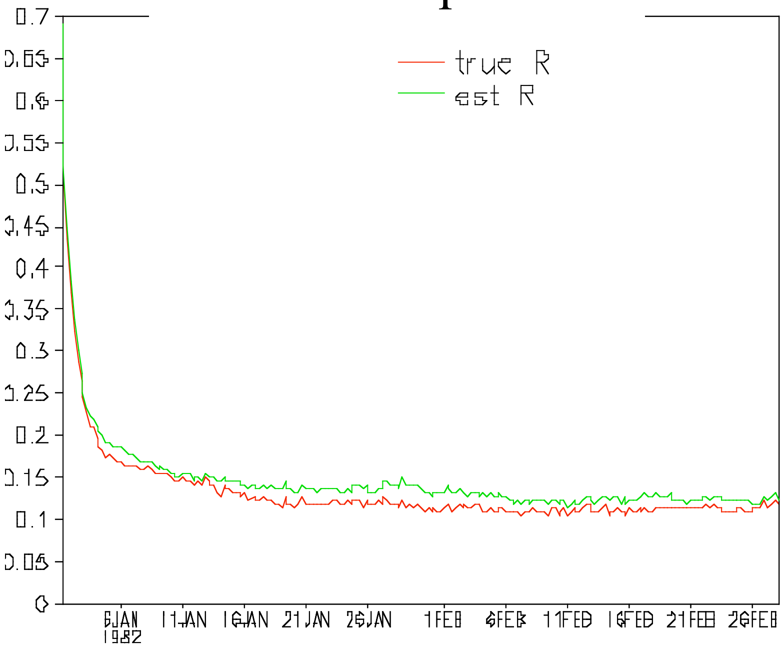
Using an initially wrong R and Δ but estimating them adaptively

Using a perfect R and estimating Δ adaptively

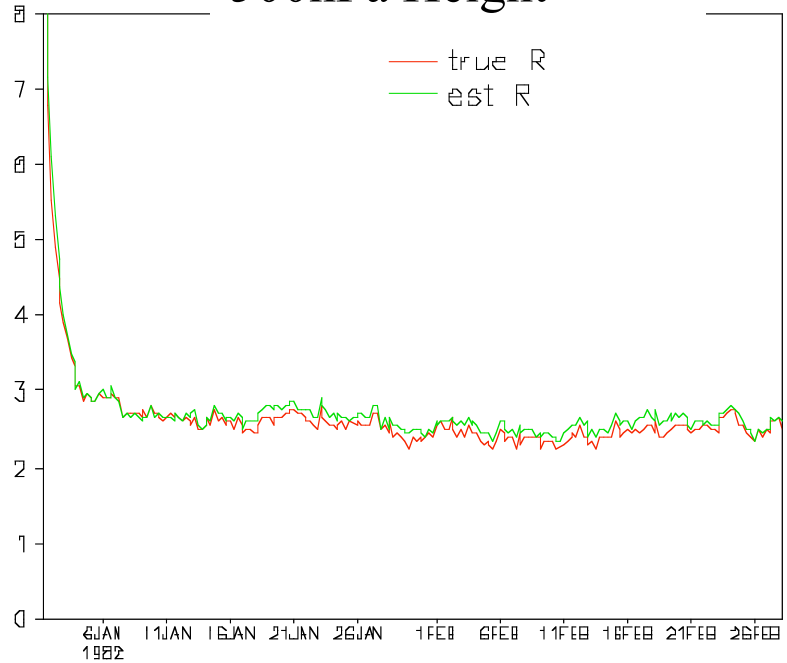
After **R** converges, they give similar inflation factors (time dependent)

Global averaged analysis RMS

500hPa Temperature



500hPa Height



Using an initially wrong R and Δ but estimating them adaptively

Using a perfect R and estimating Δ adaptively

Summary and future work

- The online (adaptive) estimation of inflation parameter alone does not work without estimating the observational errors.

- Estimating both of the observational errors and the inflation parameter simultaneously our approach works well on both the Lorenz-40 and the SPEEDY global model. It can also be applied to other ensemble based Kalman filters.

- SPEEDY experiments show our approach can simultaneously estimate observational errors for different instruments.

- We are extending our approach to estimate off-diagonal terms (correlation lengths) in the observation error covariance. We will use to estimate the correlated errors of AIRS retrievals.