Coping with Model Errors in Data Assimilation

Istvan Szunyogh

Texas A&M University
Department of Atmospheric Sciences

The 7th EnKF Data Assimilation Workshop, State College, PA, May 23-27, 2016

Models

“*The sciences do not try to explain, they hardly even try to interpret, they mainly make models. By a model is meant a mathematical construct which, with the addition of certain verbal interpretations, describes observed phenomena. The justification of such a mathematical construct is solely and precisely that it is expected to work."*– John von Neumann

- a spot-on description of our justification of the **mathematical model of data assimilation;**
- the “verbal interpretations” are crucial, because they guide our intuition, but they also tend to make us forget that we are working with models,

–IS

The Mathematical Model of Data Assimilation

For simplicity, assume that the **mathematical model of data assimilation** is the **model of sequential data assimilation:**

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}\delta\mathbf{y}, \qquad \delta\mathbf{y} = \mathbf{y}^o - \mathcal{H}\left(\mathbf{x}^b\right), \qquad \mathbf{K} = \mathbf{P}^b\mathbf{H}^T\left(\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R}\right)^{-1}$$

A particular scheme must be robust to errors in

- unexpected (gross) errors in the background and the observations
- the observation and background error statistics
- the model that defines the observation function

We can achieve this by replacing the statistics (e.g., background error covariance matrix and observation error covariance matrix) by robust statistics

Robust Statistics

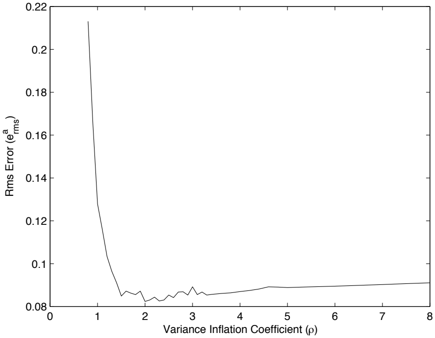
Robust statistics must satisfy the following criteria (Huber and Ronchetti 2009):

- **efficiency**–for **clean input data** (data that satisfy the assumptions of the original statistical model), the results are almost as good as for the original statistics (perfect model experiments)
- **stability–small errors** in the assumptions lead to small errors in the (state) estimates
- **breakdown–gross errors** in the input data do not lead to catastrophic breakdown

Example 1: Variance Inflation (from Szunyogh, 2014: Applicable Atmospheric Dynamics)

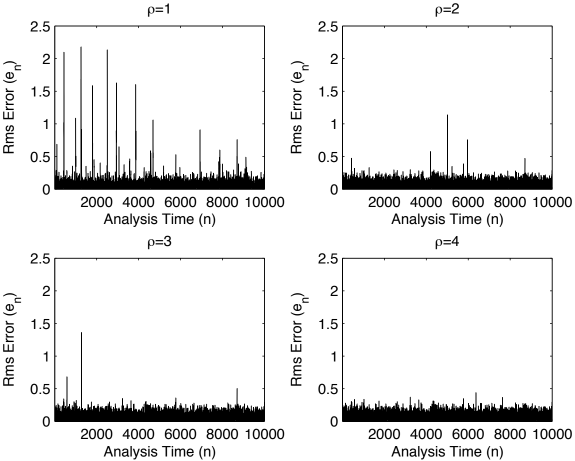
Assimilation of simulated **observations of the Henon Mapping** by an **Extended Kalman Filter**

- The **sources of the violation** of the assumptions of ETKF are the limitations of the TLM in describing the error dynamics



Example 1 (Continued)

Variance inflation reduces the magnitude and the frequency of **error bursts**



Example 2: Uncorrected Background Bia

From the Appendix of Holt et al., 2015, MWR, **143,** 3956–3980

- Assume that the model has a single state variable x and the scalar background x^b is biased by b, and we have a direct observation y^o of x ($\mathcal{H}\left(x^b\right) = x^b$, $H = H^T = 1$)
- The analysis still have minimum variance, but not minimum rms error
- The Kalman gain that minimizes the rms error is

$$\hat{K} = \left(P^b + b^2\right)\left(P^b + b^2 + R\right)^{-1}$$

  rather than $K = \left(P^b\right)\left(P^b + R\right)^{-1}$
- The same effect can be achieved by using K and replacing R by

$$\hat{R} = R(1 + b^2/P^b)^{-1}$$

Example 3: Continued

Assume that

- the data assimilation system uses $(P^b)^{1/2} = 4\ hPa$ for the SLP in a TC
- the data assimilation system uses $(R)^{1/2} = 5\ hPa$ for a TCVitals SLP observation
- x^b is biased with $b = 40\ hPa$

Using $\hat{R}$ rather than R

- increases the standard deviation of the analysis error from 3.12 hPa to 4.92 hPa, but reduces the rms error of the analysis from 24.59 hPa to 4.96 hPa
- A huge reduction of the analysis bias at the price of a small increase of the analysis error variance
- Can be used, if there is no reason to believe that the analysis with a smaller bias would upset the model

Example 4: Gross Observation Errors and/or Good Observations that May Shock the Model

- Roh et al., 2013: *Observation Quality Control with a Robust Ensemble Kalman Filter, MWR,* **141,** 4414–4428
- The analysis update equation can be Huberized as

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{KG}\left(\delta\mathbf{y}\right),$$

  where $\mathbf{G}\left(\delta\mathbf{y}\right)$ is the **Huber function,**
- For instance, a potential choice for the Huber function is

$$\mathbf{G}\left(\delta\mathbf{y}\right) = \begin{cases} \delta\mathbf{y} & \text{if} |\delta\mathbf{y}| < c \\ c & \text{if} \delta\mathbf{y} \geq c \\ -c & \text{if} \delta\mathbf{y} \geq -c \end{cases}$$

  where c is a prescribed clipping innovation
- **Main appeal:** It can easily implemented in an EnKF for QC (no need for variational minimization)

A Historical Note

- The **Huber norm QC** went into operations at ECMWF with cycle 35r3 on September 8 2009. Long before it was written up for the 2015 paper.
- An earliest citation is Tavolato and Isaksen, 2009: Huber norm quality control in the IFS. ECMWF Newsletter, 122, 27–31.

Illustration of Examples 3 and 4 for TC Observations

- Based on Holt et al., 2015, MWR, **143,** 3956–3980
- **Models:** NCEP GFS at resolution T62L28, RSM at resolution 48 km and 28 levels (a glorified toy system)
- **Data assimilation:** LETKF
- **Regular observations:** all operationally assimilated non-radiance observations
- **TC observations:** TCVitals SLP ($R^{1/2} = 0.5$ hPa), drospsondes from DOTSTAR, QuikSCAT (both with Huberized innovation)

Illustration: Sinlaku Analyses

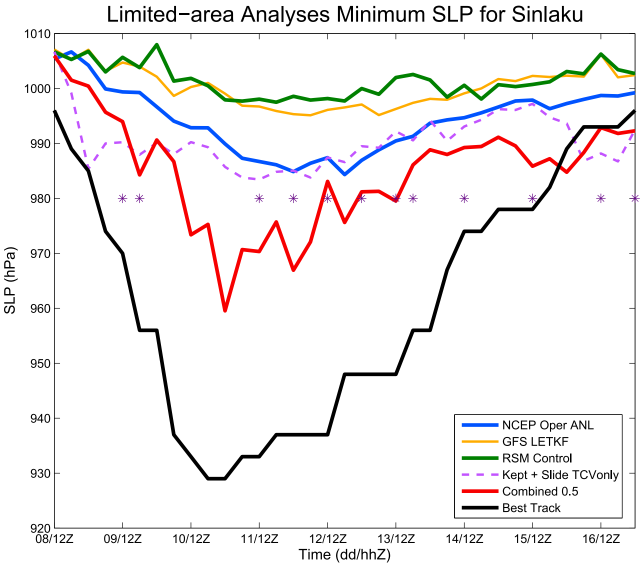
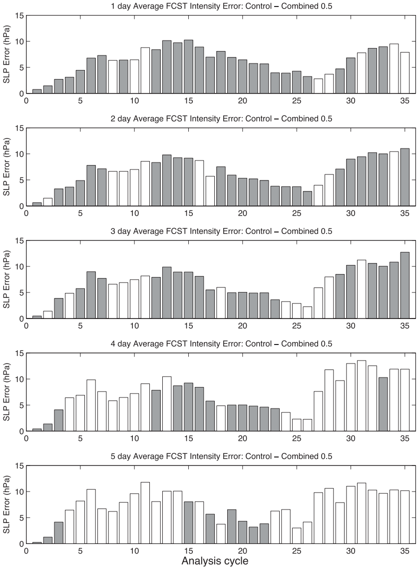


Illustration: Sinlaku Forecasts



Concluding Remarks

- People have always been working hard on making their data assimilation systems robust
- But, they do not like to talk about the adjustments they make to the error statistics, because they feel that these are hard to defend (reviewers make sure that they feel that way!)
- Keep in mind that the need for such adjustments is fully expected, as the mathematical model of data assimilation is not more than an extremely useful model