Data Assimilation: Estimation and Prediction (group meeting)

Sujit Ghosh and Mariana Pensky

Time and place: March 30, 2005, 3:30-5:00, Rm.# 203 SAMSI

Summary report:
The meeting started with a review of the paper by E. Ott et al., "A local ensemble Kalman filter for atmospheric data assimilation," Tellus 56A pp. 415-428 (2004) (the paper is currently available on this website).

The procedure used in the paper was discussed in detail and the consensus was that the procedure is motivated by computational benefits rather than by (statistical) conceptual considerations. One of the main drawbacks of the method proposed in Ott et al. (2004) seems to the fact that local models do not generate a valid probability model for the entire space. The method of choosing the local neighborhood sounds rather *ad-hoc*. However, this paper does try to reduce the curse of dimensionality by focusing the predictions based on local neighborhoods. Hence, it appears that this method is related to our goal of developing a model that would reduce the curse of dimensionality. But, in principle, the method that we proposed and discussed during our meeting on March 16 is quite different to the method proposed in Ott et al. (2004).

One of main benefits of our proposed method is that it yields a valid probability distribution which is consistent with the local models that we develop based on neighborhood structure. We plan to use a very similar simulation study that is being conducted in Ott et al. (2004) and compare our results to theirs in terms of some predictive criteria (e.g., we could use the Lorenz–96 model for the dynamic system).

In the remaining part of the meeting we explored several possibilities to model the G matrices and the H function that appear in the model that we formulated on March 16, i.e., the model given by,

$$\vec{X}^a_I \sim N(\vec{X}^b_I + \sum_J G^a_{I,J}(\vec{X}^a_J - \vec{X}^b_J), \Sigma^a_I), \tag{1}$$

$$Y^0_l \sim N(H_l(\vec{X}^a) + \sum_k q_{l,k}(Y^0_k - H_l(\vec{X}^a)), \sigma^2_l), \tag{2}$$

where the weights matrices $G^a_{I,J}$ are of the form

$$G^a_{I,J} = \frac{\Lambda}{[\alpha|i-i'|^d + \beta|j-j'|^d + \gamma|z-z'|^d]^{1/d}}, \quad d>0,\ J \in N(I). \tag{3}$$

Here Λ is a symmetric matrix (compare this with the report for March 16 and notice that we replaced ρ by Λ since ρ is not a good notation for matrices!). Equations (1)–(3) completely specifies the model and in principle we should be able to work with this structural equations of the model, but some questions still remain:

1. What is a good choice for Σ_I^a?
2. What type of functions should we use for $H_l(\vec{X}^a)$?
3. How should we choose $q_{l,k}$?

We tried to come up with quite few solutions. Some choices are highlighted below. Other suggestions are very much welcome.

1. *The choice of* Σ_I^a*:* We decided to model Σ_I^a as

$$\Sigma_I^a = K_I \Sigma_0^a$$

   where K_I is a known positive number. This allows us to slightly modify modeling of the weights, namely

$$G_{I,J}^a = \frac{K_J}{[\alpha|i-i'|^d + \beta|j-j'|^d + \gamma|z-z'|^d]^{1/d}}\,\Lambda, \quad d>0.\ J \in N(I). \tag{4}$$

   Then definitions of Σ_I^a and $G_{I,J}^a$ are consistent and satisfy condition $G_{I,J}^a \Sigma_I^a = G_{J,I}^a \Sigma_J^a$ (see report for March 16). In the above, K_I is an adjustment factor which allows to work with neighborhoods of variable sizes, namely, one can take K_I to be inversely proportional to the number of neighbors.

2. *The choice of* $H_l(\vec{X}^a)$*:* We have decided to use *thin plate splines* to model the H functions. See Wahba (1990) "Spline Models for Observational Data", CBMS, 59, Siam. (Read Sections 2.4-2.5, Chapter 7 and the references therein). Another interesting paper of interest in this area would be Wahba (1985) "Multivariate thin plate spline smoothing...constraint", Statistical Image processing and Graphics, 275-290.

3. *The choice of* $q_{l,k}$*:* The weights $q_{l.k}$ can be taken as inversely proportional to the distance between the points l and k, i.e.

$$q_{l,k} = \frac{\rho}{\|s_k - s_l\|},$$

   where $\|\cdot\|$ is some kind of norm (not necessarily Euclidean) and s_k is the vector of coordinates of point k.

Participants: Sujit Ghosh, Marianna Pensky, Minjung Kyung and Dave Holland.