Data Assimilation: Estimation and Prediction (group meeting)

Sujit Ghosh and Mariana Pensky

Time and place: March 16, 2005, 3:30-5:00, Rm.# 203 SAMSI

Summary report:
In our meeting we continued our discussions on how we can apply CAR and SAR models to the data assimilation problems. The main idea behind this approach is to avoid the required inversion of high dimensional matrices (of the order of $\approx 10^6$) but keeping the model assumptions as flexible as possible.

During our discussions it was pointed out that we need to consider multivariate generalizations of the CAR models, commonly known as the MCAR models. This is because the state vector X^a may contain several different physical measurements (temperatues, wind directions, humidity, etc.) measured at various spatial locations (grid points).

Hence, we can rewrite our state vector as a tensor

$$X^a = \{X_{I,l},\ I = (i_1, i_2, i_3), i_d = 1, \dots, N_d, d = 1, 2, 3.\},\ \vec{X}^a_I = \{X_{I,l},\quad l = 1, \dots, m\}.$$

where I is a three-dimensional spatial index for the I-th grid point and l refers to various physical measurements at location I.

We propose to model the state vector $\vec{X}^a_I$ and observation vector Y^0 using MCAR models extending the ideas that we discussed in our last meeting (02/23/2005).

We'd prefer to work the MCAR modelling approach in contrast to MSAR for the following reasons:

- The OLS estimators based on SAR models are known to be inconsistent (Ord, 1975, JASA).
- Alternative estimation methods (MLE, WLS etc.) based on SAR involves high dimensional matrix inversions, which in the furst place we want to avoid.
- Parameter estimation based on MCAR models, on the other hand are computationally less intensive and avoids high dimesional matrix operations.

The MCAR model for data assimilation:

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

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

where $I = (i_1, i_1, i_3)$, $J = (j_1, j_2, j_3)$ and l and k are scalar indices. In formulae (1) and (2), $G^a_{I,J}$ are $m \times m$ weight matrices, $g_{l,k}$ are scalar weights, Σ^a_I are conditional covariance matrices for the physical components in vector $\vec{X}^a_I$, H_l is the l-th component of the projection operator H and σ^2_l is the conditional variance of the l-th component of vector Y^0.

The above model formulation is general enough to capture all possible linear correlations among the variables. However, to make this model a tractable one, we'd make some structural assumptions to reduce the dimension of the model parameters, so that we can develop some efficient estimation methods.

Notice that the matrix $G^a_{I,J}$ captures both the inter-correlations within the components of the vactor $\vec{X}^a_I$ and also the intra-correlations between the vectors $\vec{X}^a_I$ and $\vec{X}^a_J$. On the other hand Σ^q_I captures the conditinal covariance between the $\vec{X}^a_I$ only.

A was decided to try some simple structures to start with, while modeling the matrices involved in the MCAR model. For instance, we may choose to use the following structure:

$$G^a_{I,J} = \frac{\rho}{[\sum_{d=1}^{3} \alpha_d |i_d - j_d|^p]^{1/p}} \quad p > 0,$$

where α_d's are unknown positive valued parameters to be estimated. In majority of papers so far, the coice of p has been $p = 1$ or $p = 2$. The necessity for three different α's expresses the fact that the scaling required in one dimension might be different than others (e.g., the correlations in the lattitute might be decreasing at a different rate than that in the longitude and/or altitute).

To check whether the above proposed model give rise to a valid joint distribution for X^a and Y^0, we need to use the Hammersely-Clifford type results. Our conjecture is that we might need some conditions like the following:

$$G^a_{I,J}\Sigma^a_I = G^a_{J,I}\Sigma^a_J. \tag{3}$$

During our next meetings we'd like to investigate the above type of conditions and to search alternative structures that might be suitable for the matrices like $G^a_{I,J}$ and Σ^a_I.

Participants: Mariana Pensky, Sujit Ghosh, Minjung Kyung, Prashant Pai, Kristen Foley and Dave Holland.