Adaptive and Non-Global MMSE of Covariance for Meteorologic Data Assimilation

Alexandre Guirland Nowosad

DMD-CPTEC-INPE
Av. dos Astronautas, 1758
12201-970 São José dos Campos-SP
alex@cptec.inpe.br
Summary:

- Introduction
- Development and Results
- Conclusions
Introduction

- Numerical Weather Prediction requires the initial state $w_i$ of the atmosphere to be known or estimated (Kalnay, 2003; Daley, 1991; Cohn, da Silva, Guo, Sienkiewicz e Lamich, 1998).

- Usually this state vector is estimated using a composition of observed data and the output of a meteorological model, as solution to the minimization problem:

$$
\min_{w_i} J(w_i) = \left( w_i - w^f \right)^T \left( P^f \right)^{-1} \left( w_i - w^f \right) + \\
\left[ w^o - H(w_i) \right]^T R^{-1} \left[ w^o - H(w_i) \right]
$$

$w^f$ is given by the model, $P^f$ is the covariance of the error in $w^f$, $w^o$ the observation, $H$ the observation function and $R$ the covariance of the observation error.

- This work concerns fast and optimal calculation of $P^f$. 
• Let

\[ e^f = w^f - w^r, \quad e^o = w^o - H(w^r) \]

\( w^r \) is the real atmosphere (inaccessible!)

• Definition of \( P^f \):

\[ P^f \equiv E \left\{ (e^f - E\{e^f\})(e^f - E\{e^f\})^T \right\} \]

• In practice one uses linear model for observation and:

\[ e \equiv H^+ w^o - w^f \]  
(Dee and daSilva, 1999)
• Another Problem: Computational complexity

\[ \tilde{e}_j = \begin{bmatrix} \tilde{e}_u & \tilde{e}_v & \cdots & \tilde{e}_\phi \end{bmatrix}_T^{1 \times L} \]

\[ P = \begin{bmatrix} \tilde{e}_1 \tilde{e}_1^T & \tilde{e}_1 \tilde{e}_2^T & \cdots & \tilde{e}_1 \tilde{e}_{N_x \cdot N_y \cdot N_z}^T \\ \tilde{e}_2 \tilde{e}_1^T & \tilde{e}_2 \tilde{e}_2^T & \cdots & \tilde{e}_2 \tilde{e}_{N_x \cdot N_y \cdot N_z}^T \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{e}_{N_x \cdot N_y \cdot N_z} \tilde{e}_1^T & \tilde{e}_{N_x \cdot N_y \cdot N_z} \tilde{e}_2^T & \cdots & \tilde{e}_{N_x \cdot N_y \cdot N_z} \tilde{e}_{N_x \cdot N_y \cdot N_z}^T \end{bmatrix} \]

\[ O \left( \left[ N_x \cdot N_y \cdot N_z \cdot L \right]^2 \right) = 10^{12} \]
- Aproximation 1: Vertical covariance is $s(z_1, z_2) = \sigma^2 \delta(z_1 - z_2)$ and independent from horizontal covariance.

- Aproximation 2: Homogeneous horizontal covariance

$$P^f(z, \Delta x, \Delta y) = s(z) \rho(\Delta x, \Delta y)$$

If uses FFT results $O([N_x \log N_x N_y \log N_y]. N_z L) = 10^9$
Development and Results

- Calculus of vertical variance in practice

\[ \bar{e}_{xy}(z) = \bar{e}_{xy}(z) = \frac{1}{T} \sum_{t} e_{xyzt} \]

\[ s_{xy}(z) = \langle s_{xy}(z) \rangle = \frac{1}{T} \sum_{t} \left\{ [e_{xyt}(z) - \bar{e}_{xy}(z)]^2 \right\} \]

\[ s(z) = \langle s_{xy}(z) \rangle = \frac{1}{N_x \cdot N_y - 1} \sum_{xy} s_{xy}(z) \]

- Calculus of horizontal covariance in practice

\[ \rho_z(\Delta x, \Delta y) = \left\langle \frac{\left[ e_{zt}(x + \Delta x, y + \Delta y) - \bar{e}_{zt}(x + \Delta x, y + \Delta y) \right] \times \left[ e_{zt}(x, y) - \bar{e}_{zt}(x, y) \right]}{\left[ e_{zt}(x, y) - \bar{e}_{zt}(x, y) \right]^2} \right\rangle \]
\[ \rho (\Delta x, \Delta y) = \left\langle \frac{\rho_z(\Delta x, \Delta y)}{\rho_z(0,0)} \right\rangle \]

- Space and time mean- and variance-ergodicity is assumed, therefore restrictions on autocovariance function of \( e^f \) apply.

(Papoulis, 1991)
Parameterizations (in order of calculus):

Parameterization of variance uses a polynomial of degree \( m \) conformal to the number of calculated variances valued at least 10% of the maximum of the set. Dynamic memory allocation is used in FORTRAN 90 to define the dimension \( m \) of the MMSE problem

\[
\hat{s}_{\text{norm}}(z) = \frac{1}{\hat{s}_{\text{max}}} \hat{\sigma}^2(z) = \sum_{i=0}^{m} a_i \left( \frac{z}{m} \right)^i
\]

\[
\tilde{a} = [a_0 \: \cdots \: a_m]^T
\]

To recover variances

\[
\hat{s}(z) = s_{\text{max}} \cdot \hat{s}_{\text{norm}}(z)
\]

Remaining parameterizations also use MMSE discarding calculated values smaller in amplitude than \( 10^{-6} \)

\[
\hat{\rho}_{\Delta y}(\Delta x) = \exp \left( - \alpha_{\Delta y}(\Delta x) \right)
\]

\[
\hat{\rho}_{\Delta x}(\Delta y) = \exp \left( - \alpha_{\Delta x}(\Delta y) \right)
\]

\[
\hat{\rho}(\Delta x, \Delta y) = \hat{\rho}_x(\Delta x)\hat{\rho}_y(\Delta y)
\]

(see Purser-Wu-Parrish-Roberts, 2003)
Figura 1: Vertical standard deviation.
Figure 2: Horizontal covariance.
Conclusions

- Dynamic memory allocation in FORTRAN 90 makes this software self-adaptive to available data providing smooth estimate (see Sztipanovits and Karsen, 2000).

- Results show estimated covariance is more realistic than calculated covariance.

- Computational complexity of problem is reduced.