Implementation of spatial correlations in the background error covariance matrix in the BASCOE system

Quentin Errera
Belgium Institute for Space Aeronomy (BIRA-IASB)
quentin@oma.be

and

Richard Ménard
Environment Canada

SPARC DA Workshop
June 20-22, 2011
The Belgium Assimilation System for Chemical ObsErvations (BASCOE)

- 4D-Var system dedicated to assimilation of stratospheric chemical observations
- 57 chemical species interact through 200 chemical reactions
- All species advected, usually by ECMWF dynamical fields
- The system includes a simple PSC parameterization
- Typical resolution: $3.75^\circ$ lon $\times$ $2.5^\circ$ lat $\times$ 37 levels (0.1 hPa to the surface)
- Up to now, $B$ has always been considered diagonal, with a fixed error usually between 20% to 50% of the background field
- This system has been used successfully to assimilate UARS MLS, MIPAS, GOMOS and EOS MLS
Why to implement spatial correlations in $B$?

- NRT assimilation of EOS MLS is done by BASCOE within MACC. Correlations in $B$ might allow to improve the analyses AND also allow to increase the resolution $\Rightarrow$ Double improvement

- We would also like to compare 4D-Var and EnKF. For this purpose, an optimal 4D-Var system is required.
1. New formulation of BASCOE
2. Numerical test: assimilation of one pseudo observation
3. Real test: assimilation of EOS MLS ozone
The classical formulation of 4D-Var (i.e. not incremental), as previously implemented in BASCOE, aims at minimizing the objective function $J$:

$$J(x(t_0)) = \frac{1}{2} [x(t_0) - x^b(t_0)]^T B^{-1} [x(t_0) - x^b(t_0)]$$

$$+ \frac{1}{2} \sum_{i=0}^{N_{\Delta t}} [y_i - H(x(t_i))]^T R_i^{-1} [y_i - H(x(t_i))]$$

where:

- $x^b(t_0)$ and $B$ are, respectively, the background model state (at time step 0) and its associated error covariance matrix.
- $y_i$ and $R_i$ are, respectively, the observation vector and its associated error covariance matrix at time step $i$, $N_{\Delta t}$ being the number of model time step.
- $x(t_i)$ is the model state vector at time step $i$ and is calculated from the previous time step by the model operator $M$: $x(t_i) = M_{i-1}[x(t_{i-1})]$.
- $H$ is the observation operator that maps the model state into the observation space.
- In the following, $J^b$ and $J^o$ will refer to the background and observation terms of $J$. 

$\text{Errera and Ménard}$
As the typical dimension of $\mathbf{x}$ is around $10^6$, a full $\mathbf{B}$ matrix is of size $10^{12}$. This is far too large for current computers.

The specification of the elements of such a matrix requires a huge amount of a priori information, more than available.

For those reasons, it is necessary to reduce the problem. Following the method used by meteorological centers [Bannister, 2008a,b, QJ, and references therein], we apply the following control variable transform (CVT):

\[
\underbrace{\mathbf{x} - \mathbf{x}^b}_{\delta\mathbf{x}} = \mathbf{L}\chi
\]  

(2)

where

\[
\mathbf{B} = \mathbf{L}\mathbf{L}^T
\]

(3)
The objective function:

\[ J^b(x(t_0)) = \frac{1}{2} [x(t_0) - x^b(t_0)]^T B^{-1} [x(t_0) - x^b(t_0)] \] (4)

\[ J^o(x(t_0)) = \frac{1}{2} \sum_{i=0}^{N\Delta t} [y_i - H(x(t_i))]^T R_i^{-1} [y_i - H(x(t_i))] \] (5)

becomes:

\[ J^b(\chi(t_0)) = \frac{1}{2} \chi(t_0)^T \chi(t_0) \] (6)

\[ J^o(\chi(t_0)) = \frac{1}{2} \sum_{i=0}^{N\Delta t} [d_i - H(L\chi(t_i))]^T R_i^{-1} [d_i - H(L\chi(t_i))] \] (7)

where \( d_i = y_i - H[x^b(t_0)] \)
Efficient minimization of \( J \) requires \( \nabla J \). In the classical formulation, we had:

\[
\nabla J^b = B^{-1}[x(t_0) - x^b(t_0)] \tag{8}
\]

\[
\nabla J^o = \sum_{i=0}^{N_{\Delta t}} M_{1 \rightarrow 0}^T M_{2 \rightarrow 1}^T \cdots M_{i \rightarrow i-1}^T \left( \frac{\partial H(x(t_i))}{\partial x(t_i)} \right)^T R_i^{-1} [y_i - H(x(t_i))] \tag{9}
\]

In contrast, with the new control variable \( \chi \), we have:

\[
\nabla J^b = \chi_0 \tag{10}
\]

\[
\nabla J^o = L^T \sum_{i=0}^{N_{\Delta t}} M_{1 \rightarrow 0}^T M_{2 \rightarrow 1}^T \cdots M_{i \rightarrow i-1}^T \left( \frac{\partial H(x(t_i))}{\partial x(t_i)} \right)^T R_i^{-1} [y_i - H(x(t_i))] \tag{11}
\]
Again, as done by meteorological centers [Bannister, 2008a,b, QJ], we construct the BECM from a set of very sparse matrices. In our case, $L$ is given by

$$L = G \Sigma S \Lambda^{1/2}$$  \hspace{1cm} (12)

where

- $\Lambda$ is the correlation matrix defined in the spectral space; $\chi' = \Lambda^{1/2} \chi$
- $S$ operates the transformation from the spectral space to the gaussian grid. The control variable $\chi$ is thus a set of spherical harmonic coefficients; $\delta x'' = S \chi'$
- $\Sigma$ is the background error variance matrix; $\delta x' = \Sigma \delta x''$
- $G$ operates the transformation from the gaussian grid to the lat/lon grid of BASCOE; $\delta x \equiv x - x^b = G \delta x'$
By considering homogeneous and isotropic correlations, $\Lambda$ is a block diagonal matrix in the spectral space. For example, if $N_{lat} = 2$, $N_{lon} = 4$ and $N_{lev} = 3$ and if we organize $\chi$ as follow, $\Lambda$ takes the form:

$$
\chi = \begin{pmatrix}
\chi_{00} & 0 & 0 \\
\chi_{00} & \chi_{11} & 0 \\
\chi_{11} & \chi_{12} & \chi_{13} \\
\chi_{12} & \chi_{13} & 0 \\
\chi_{10} & \chi_{10} & \chi_{10} \\
\chi_{10} & \chi_{10} & \chi_{13} \\
\chi_{11} & \chi_{11} & 0 \\
\chi_{11} & \chi_{11} & \chi_{11} \\
\chi_{11} & \chi_{11} & \chi_{11}
\end{pmatrix},
\Lambda = \begin{pmatrix}
q_{01} & c_{01}^{12} & c_{01}^{13} \\
q_{02} & c_{02}^{12} & c_{02}^{13} \\
q_{03} & c_{03}^{12} & c_{03}^{13} \\
q_{01} & c_{01}^{12} & c_{01}^{13} \\
q_{10} & c_{10}^{12} & c_{10}^{13} \\
q_{11} & c_{11}^{12} & c_{11}^{13} \\
q_{12} & c_{12}^{12} & c_{12}^{13} \\
q_{13} & c_{13}^{12} & c_{13}^{13} \\
q_{14} & c_{14}^{12} & c_{14}^{13}
\end{pmatrix}
$$

- $\chi_{lm}^l$ are the expansion of $x$ in spherical harmonics; $l$ is the level index, $n$ is the total wave number and $m$ is the zonal wave number.
- $q_{ln}^l$ are the horizontal correlation coefficients, depending on the level $l$ and on the total wavenumber $n$ (but not on the zonal wavenumber $m$).
- $c_{ln}^{l_i,l_j}$ are the vertical correlation coefficients, depending on the pair of levels $(l_i, l_j)$ and the total wavenumber $n$. 
Current formulation of $\mathbf{L}$ in BASCOE

- We assume that $c_{n_p l_i l_j} = c_{n_q l_i l_j}$, i.e. vertical correlations are independent of the total wave number $n$.
  $\Rightarrow$ vertical and horizontal correlations are separable

- Horizontal correlations: gaussian and second order autoregressive (SOAR) correlations can be modelled, given a correlation length scale $L_h$ (in km)

- Vertical correlations: gaussian correlations can be modelled, given $L_v$ (in level units)
The new formulation of BASCOE is tested with the assimilation of a single pseudo observation:

- Model grid size: 41 latitudes $\times$ 80 longitudes $\times$ 9 levels
Assimilation of a single pseudo observation (1/2)

The new formulation of BASCOE is tested with the assimilation of a single pseudo observation:

- Model grid size: 41 latitudes × 80 longitudes × 9 levels
- Uniform background field and variances: $x^b = 1; \Sigma = 0.1I$
The new formulation of BASCOE is tested with the assimilation of a single pseudo observation:

- Model grid size: 41 latitudes \( \times \) 80 longitudes \( \times \) 9 levels
- Uniform background field and variances: \( x^b = 1; \Sigma = 0.1I \)
- One observation is located on a grid point (20,40,5); \( y^o = 1.2; \sigma^o = 0.001 \)
The new formulation of BASCOE is tested with the assimilation of a single pseudo observation:

- Model grid size: 41 latitudes $\times$ 80 longitudes $\times$ 9 levels
- Uniform background field and variances: $x^b = 1; \Sigma = 0.1I$
- One observation is located on a grid point (20,40,5);
  $y^o = 1.2; \sigma^o = 0.001$
- Gaussian vertical and horizontal correlations. $L_v = 1; L_h = 1200km$
BASCOE setup:

- Resolution: $2^\circ \text{lat} \times 2^\circ \text{lon} \times 37$ levels
- Only O$_3$ is considered and the chemistry is turned off in order to reduce the CPU time. Ozone data are rejected above 0.5 hPa.
- wind and T$^\circ$ are taken from ERA-Interim

EOS MLS O$_3$ data are assimilated between December 2004 and March 2005
To make the system optimal, the error parameters must be tuned.

This is done using the innovations (Hollingsworth and Lönnberg, 1986, Tellus):

\[ x^b = x^{\text{truth}} + \epsilon^b \]
\[ y^o = Hx^{\text{truth}} + \epsilon^o \]

then:

\[ y^o - Hx^b = \epsilon^o - H\epsilon^b \]

Assuming that observations errors are horizontally uncorrelated and that background errors are spattially correlated, auto-covariance of \( I \) can be used to estimates the errors.

Thus:

\[ \langle I(i, i), I(i, i) \rangle = (\sigma^b)^2 + (\sigma^o)^2 \]
\[ \langle I(i, j), I(i, j) \rangle = (\sigma^b)^2 \rho(r_{ij}) \]
Application of Hollingsworth and Lönnberg on BASCOE

- Step 1: Assimilation of EOS MLS (Dec 2004) without spatial correlations, $\sigma^b = 30\%$; $\sigma^o = 15\%$

$\Rightarrow$ EXP 1
Application of Hollingsworth and Lönnberg on BASCOE

- Step 2: Estimating the errors parameters from EXP 1
Application of Hollingsworth and Lönnberg on BASCOE

- Step 2: Estimating the errors parameters from EXP 1

Estimated covariance at 68.1292 hPa on 15dec2004

Gauss: $L_{gauss} = 608 \text{km}$; $\sigma^b = 1.30 \times 10^{-7}$; $\sigma^o = 1.54 \times 10^{-7}$; resnorm = 8.62e-03

SOAR: $L_{soar} = 356 \text{km}$; $\sigma^b = 1.38 \times 10^{-7}$; $\sigma^o = 1.47 \times 10^{-7}$; resnorm = 4.16e-03
Assimilation of EOS MLS Ozone observations (5/9)

Application of Hollingsworth and Lönnberg on BASCOE

- Step 2: Estimating the errors parameters from EXP 1

![Graphs showing error variance, pressure, horizontal correlation length scale, and residual norm of the fit for Obs Gauss, Obs SOAR, Obs MLS, Background Gauss, Background SOAR, Gauss, and SOAR.]
Application of Hollingsworth and Lönnberg on BASCOE

- Step 3: Re-assimilation of EOS MLS (Dec 2004).
  - \( \sigma^b \) and \( \sigma^o \) provided by tuning of EXP 1
  - Horizontal correlations: SOAR with \( L_h = 600 \) km
  - Vertical correlations: gaussian with \( L_v = 1.5 \) level
  - This is EXP 2
Application of Hollingsworth and Lönnberg on BASCOE

- Step 4: Estimating the errors parameters from EXP 2
Assimilation of EOS MLS Ozone observations (5/9)

Application of Hollingsworth and Lönnberg on BASCOE

- Step 4: Estimating the errors parameters from EXP 2
Application of Hollingsworth and Lönnberg on BASCOE

- Step 5: Re-assimilation of EOS MLS (Dec 2004).
  - $\sigma^b$ and $\sigma^o$ provided by tuning of EXP 2
  - Horizontal correlations: SOAR with $L_h$ tuned from EXP 2
  - Vertical correlations: gaussian with $L_v = 1.5$ level
  - This is EXP 3
Application of Hollingsworth and Lönnberg on BASCOE

Step 6: Estimating the errors parameters from EXP 3 and checking for convergence
Assimilation of EOS MLS Ozone observations (5/9)

Application of Hollingsworth and Lönnerberg on BASCOE

- Step 6: Estimating the errors parameters from EXP 3 and checking for convergence

![Graphs showing error variance, pressure, horizontal correlation length scale, and residual norm of the fit for OBS Gauss, OBS SOAR, OBS MLS, Background Gauss, Background SOAR, Gauss, SOAR, and Errors form EXP 2.](image-url)
Application of Hollingsworth and Lönnberg on BASCOE

- Step 6: Estimating the errors parameters from EXP 3 and checking for convergence

⇒ Convergence not reached
Application of Hollingsworth and Lönnberg on BASCOE

- Step 7: Re-assimilation of EOS MLS (Dec 2004).
  - $\sigma^b$ and $\sigma^o$ provided by tuning of EXP 3
  - Horizontal correlations: SOAR with $L_h$ tuned from EXP 3
  - Vertical correlations: gaussian with $L_v = 1.5$ level
  - This is EXP 4
Application of Hollingsworth and Lönnberg on BASCOE

- Step 8: Estimating the errors parameters from EXP 4 and checking for convergence
Application of Hollingsworth and Lönnberg on BASCOE

- Step 8: Estimating the errors parameters from EXP 4 and checking for convergence
Application of Hollingsworth and Lönnberg on BASCOE

- Step 8: Estimating the errors parameters from EXP 4 and checking for convergence
Application of Hollingsworth and Lönnberg on BASCOE

- **Step 8**: Estimating the errors parameters from EXP 4 and checking for convergence

⇒ Convergence Ok
Application of Hollingsworth and Lönnberg on BASCOE

- New Exp 5
  - $\sigma^b = 30\%$; $\sigma^o = 15\%$
  - Horizontal correlations: SOAR with $L_h = 600$ km
  - Vertical correlations: gaussian with $L_v = 1.5$ level
  - This is EXP 5
Application of Hollingsworth and Lönnberg on BASCOE

- Step 9: Checking whether $J/p$ has reached its theoretical value of 0.5

\[
J(x(t_0)) = \frac{1}{2} [x(t_0) - x^b(t_0)]^T B^{-1} [x(t_0) - x^b(t_0)] \\
+ \frac{1}{2} \sum_{i=0}^{N_{\Delta t}} [y_i - H(x(t_i))]^T R_i^{-1} [y_i - H(x(t_i))] 
\]
Application of Hollingsworth and Lönberg on BASCOE

- Step 9: Checking whether $J/p$ has reached its theoretical value of 0.5
OmF: Deviation of the BASCOE background fields to EOS MLS observations
Period considered: 01Jan2005 - 31Mar2005

Bias [−90°,−60°]
Bias [−60°,−30°]
Bias [−30°,30°]
Bias [30°,60°]
Bias [60°,90°]

Exp 1
Exp 2
Exp 3
Exp 4
Exp 5
OmA: Deviation of the BASCOE analyses fields from independent ACE-FTS observations

Period considered: 01Jan2005 - 31Mar2005
Analyses on 22 Feb 2005

<table>
<thead>
<tr>
<th>Longitude</th>
<th>Latitude</th>
<th>Analyses [ppmv] for EXP−1 at 9.8925 hPa</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Longitude</th>
<th>Latitude</th>
<th>Analyses [ppmv] for EXP−2 at 9.8925 hPa</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Longitude</th>
<th>Latitude</th>
<th>Analyses [ppmv] for EXP−3 at 9.8925 hPa</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Longitude</th>
<th>Latitude</th>
<th>Analyses [ppmv] for EXP−4 at 9.8925 hPa</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Assimilation of EOS MLS Ozone observations (9/9)

Increments on 22 Feb 2005
Spatial correlation implemented in BASCOE with a new control variable defined in the spectral space
Conclusions

1. Spatial correlation implemented in BASCOE with a new control variable defined in the spectral space

2. Homogeneous and isotropic horizontal AND vertical correlations are modelled
Conclusions

1. Spatial correlation implemented in BASCOE with a new control variable defined in the spectral space
2. Homogeneous and isotropic horizontal AND vertical correlations are modelled
3. Case studies with EOS MLS $O_3$ data show:
   - Errors parameters are successfully tuned using Hollingsworth and Lönnber’s method
   - Tuned error parameters allow to improve the lower stratosphere