# 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 $J$:

$$\begin{aligned} J(\mathbf{x}(t_0)) &= \frac{1}{2}[\mathbf{x}(t_0) - \mathbf{x}^{\mathrm{b}}(t_0)]^T \mathbf{B}^{-1}[\mathbf{x}(t_0) - \mathbf{x}^{\mathrm{b}}(t_0)] \\ &+ \frac{1}{2}\sum_{i=0}^{N_{\Delta t}}[\mathbf{y}_i - H(\mathbf{x}(t_i))]^T \mathbf{R}_i^{-1}[\mathbf{y}_i - H(\mathbf{x}(t_i))] \end{aligned} \quad (1)$$

where:

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

1. 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.
2. 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):

$$\underbrace{\mathbf{x} - \mathbf{x}^{\mathrm{b}}}_{\delta\mathbf{x}} = \mathbf{L}\chi \tag{2}$$

where

$$\mathbf{B} = \mathbf{L}\mathbf{L}^{T} \tag{3}$$

The objective function:

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

$$J^{\mathrm{o}}(\mathbf{x}(t_0)) = \frac{1}{2}\sum_{i=0}^{N_{\Delta t}}[\mathbf{y}_i - H(\mathbf{x}(t_i))]^T \mathbf{R}_i^{-1}[\mathbf{y}_i - H(\mathbf{x}(t_i))] \quad (5)$$

becomes:

$$J^{\mathrm{b}}(\chi(t_0)) = \frac{1}{2}\chi(t_0)^T \chi(t_0) \quad (6)$$

$$J^{\mathrm{o}}(\chi(t_0)) = \frac{1}{2}\sum_{i=0}^{N_{\Delta t}}[\mathbf{d}_i - H(\mathbf{L}\chi(t_i))]^T \mathbf{R}_i^{-1}[\mathbf{d}_i - H(\mathbf{L}\chi(t_i))] \quad (7)$$

where $\mathbf{d}_i = \mathbf{y}_i - H[\mathbf{x}^{\mathrm{b}}(t_0)]$

Efficient minimization of $J$ requires $\nabla J$. In the classical formulation, we had:

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

$$\nabla J^{\mathrm{o}} = \sum_{i=0}^{N_{\Delta t}} \mathbf{M}^T_{1\to 0}\mathbf{M}^T_{2\to 1}\ldots\mathbf{M}^T_{i\to i-1}\left(\frac{\partial H(\mathbf{x}(t_i))}{\partial \mathbf{x}(t_i)}\right)^T \mathbf{R}_i^{-1}[\mathbf{y}_i - H(\mathbf{x}(t_i))] \tag{9}$$

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

$$\nabla J^{\mathrm{b}} = \chi_0 \tag{10}$$

$$\nabla J^{\mathrm{o}} = \mathbf{L}^T \sum_{i=0}^{N_{\Delta t}} \mathbf{M}^T_{1\to 0}\mathbf{M}^T_{2\to 1}\ldots\mathbf{M}^T_{i\to i-1}\left(\frac{\partial H(\mathbf{x}(t_i))}{\partial \mathbf{x}(t_i)}\right)^T \mathbf{R}_i^{-1}[\mathbf{y}_i - H(\mathbf{x}(t_i))]. \tag{11}$$

Again, as done by meteorological centers [Bannister, 2008a,b, QJ], we construct the BECM from a set of very sparce matrices. In our case, **L** is given by

$$\mathbf{L} = \mathbf{G}\mathbf{\Sigma}\mathbf{S}\mathbf{\Lambda}^{1/2} \tag{12}$$

where

- $\mathbf{\Lambda}$ is the correlation matrix defined in the spectral space; $\chi' = \mathbf{\Lambda}^{1/2}\chi$
- **S** operates the transformation from the spectral space to the gaussian grid. The control variable $\chi$ is thus a set of spherical harmonic coefficients; $\delta\mathbf{x}'' = \mathbf{S}\chi'$
- $\mathbf{\Sigma}$ is the background error variance matrix; $\delta\mathbf{x}' = \mathbf{\Sigma}\delta\mathbf{x}''$
- **G** operates the transformation from the gaussian grid to the lat/lon grid of BASCOE; $\delta\mathbf{x} \equiv \mathbf{x} - \mathbf{x}^{\mathrm{b}} = \mathbf{G}\delta\mathbf{x}'$

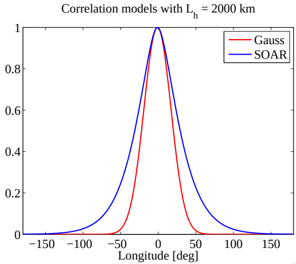
By considering homogeneous and isotropic correlations, $\mathbf{\Lambda}$ is a block diagonal matrix in the spectral space. For example, if $N_{lat} = 2$, $N_{lon} = 4$ and $N_{lev} = 3$ and if we organize $\chi$ as follow, $\mathbf{\Lambda}$ takes the form:

$$\chi = \begin{pmatrix} \chi^1_{00} \\ \chi^2_{00} \\ \chi^3_{00} \\ \chi^1_{1-1} \\ \chi^2_{1-1} \\ \chi^3_{1-1} \\ \chi^1_{10} \\ \chi^2_{10} \\ \chi^3_{10} \\ \chi^1_{11} \\ \chi^2_{11} \\ \chi^3_{11} \end{pmatrix}, \mathbf{\Lambda} = \begin{pmatrix} q^1_0 & c^{12}_0 & c^{13}_0 & & & & & & & & & \\ c^{12}_0 & q^2_0 & c^{23}_0 & & & & & & & & & \\ c^{13}_0 & c^{23}_0 & q^3_0 & & & & & & & & & \\ & & & q^1_1 & c^{12}_1 & c^{13}_1 & & & & & & \\ & & & c^{12}_1 & q^2_1 & c^{23}_1 & & & & & & \\ & & & c^{13}_1 & c^{23}_1 & q^3_1 & & & & & & \\ & & & & & & q^1_1 & c^{12}_1 & c^{13}_1 & & & \\ & & & & & & c^{12}_1 & q^2_1 & c^{23}_1 & & & \\ & & & & & & c^{13}_1 & c^{23}_1 & q^3_1 & & & \\ & & & & & & & & & q^1_1 & c^{12}_1 & c^{13}_1 \\ & & & & & & & & & c^{12}_1 & q^2_1 & c^{23}_1 \\ & & & & & & & & & c^{13}_1 & c^{23}_1 & q^3_1 \end{pmatrix}$$

- $\chi^l_{nm}$ are the expansion of $\mathbf{x}$ in spherical harmonics; $l$ is the level index, $n$ is the total wave number and $m$ is the zonal wave number
- $q^l_n$ are the horizontal correlation coefficients, depending on the level $l$ and on the total wavenumber $n$ (but not on the zonal wavenumber $m$)
- $c^{l_i,l_j}_n$ are the vertical correlation coefficients, depending on the pair of levels $(l_i, l_j)$ and the total wavenumber $n$

Current formulation of **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 indendent 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

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: $\mathbf{x}^{\mathrm{b}} = 1; \boldsymbol{\Sigma} = 0.1\mathbf{I}$

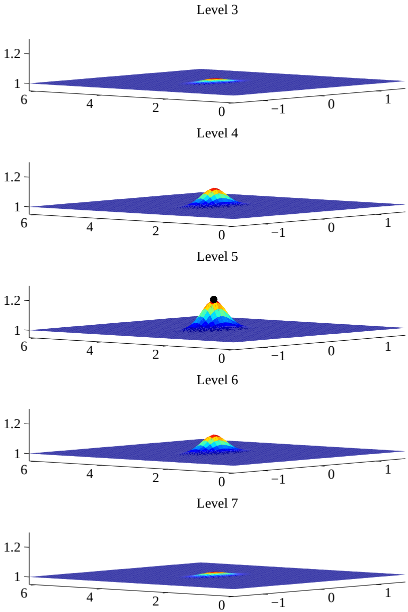
The new formulation of BASCOE is tested with the assimilation of a single pseudo observation:

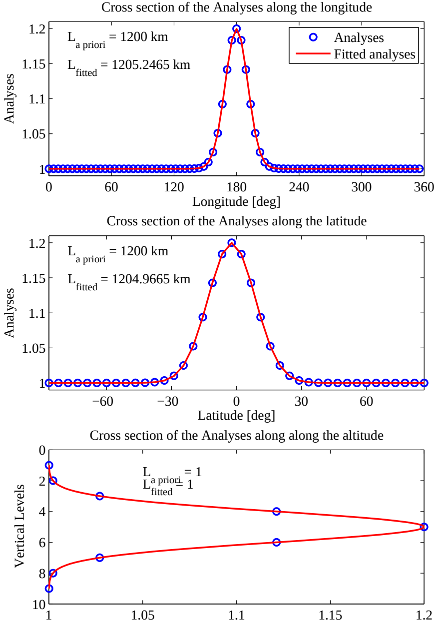
- Model grid size: 41 latitudes $\times$ 80 longitudes $\times$ 9 levels
- Uniform background field and variances: $\mathbf{x}^{\mathrm{b}} = 1; \mathbf{\Sigma} = 0.1\mathbf{I}$
- One observation is located on a grid point (20,40,5); $\mathbf{y}^{\mathrm{o}} = 1.2; \sigma^{\mathrm{o}} = 0.001$

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: $\mathbf{x}^{\mathrm{b}} = 1; \boldsymbol{\Sigma} = 0.1\mathbf{I}$
- One observation is located on a grid point (20,40,5); $\mathbf{y}^{\mathrm{o}} = 1.2; \sigma^{\mathrm{o}} = 0.001$
- Gaussian vertical and horizontal correlations. $L_v = 1; L_h = 1200\text{km}$

# Assimilation of a single pseudo observation (2/2)



BASCOE setup:

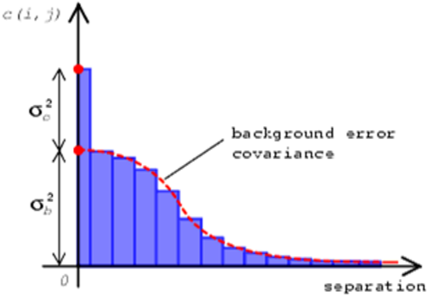
- Resolution: $2^\circ$ lat $\times$ $2^\circ$ lon $\times$ 37 levels
- Only $O_3$ is considered and the chemistry is turned off in order to reduce the CPU time. Ozone data are rejected above 0.5 hPa.
- wind and $T^\circ$ are taken from ERA-Interim

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):

$\mathbf{x}^{\mathrm{b}} = \mathbf{x}^{\mathrm{truth}} + \epsilon^{\mathrm{b}}$
$\mathbf{y}^{\mathrm{o}} = \mathbf{H}\mathbf{x}^{\mathrm{truth}} + \epsilon^{\mathrm{o}}$
then:
$\mathbf{y}^{\mathrm{o}} - \mathbf{H}\mathbf{x}^{\mathrm{b}} = \epsilon^{\mathrm{o}} - \mathbf{H}\epsilon^{\mathrm{b}}$

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: $\langle I(i,i), I(i,i)\rangle = (\sigma^{\mathrm{b}})^2 + (\sigma^{\mathrm{o}})^2$
$\langle I(i,j), I(i,j)\rangle = (\sigma^{\mathrm{b}})^2 \rho(r_{ij})$

Application of Hollingsworth and Lönnberg on BASCOE

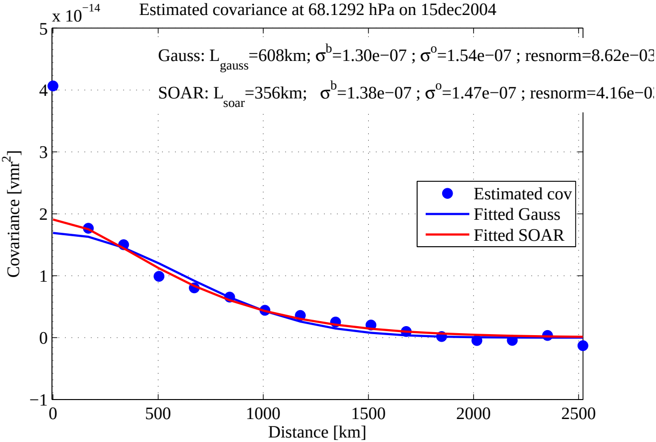
- Step 1: Assimilation of EOS MLS (Dec 2004) without spatial correlations, $\sigma^b = 30\%$; $\sigma^o = 15\%$
  $\Rightarrow$ EXP 1

Application of Hollingsworth and Lönnberg on BASCOE

- Step 2: Estimating the errors parameters from EXP 1

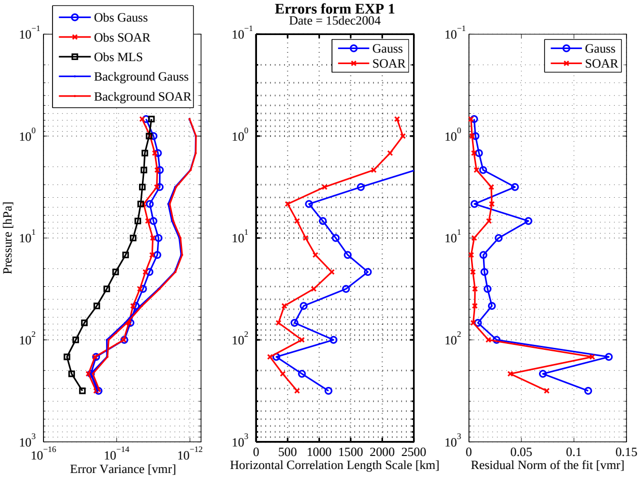
Application of Hollingsworth and Lönnberg on BASCOE

- Step 2: Estimating the errors parameters from EXP 1



Application of Hollingsworth and Lönnberg on BASCOE

- Step 2: Estimating the errors parameters from EXP 1



Application of Hollingsworth and Lönnberg on BASCOE

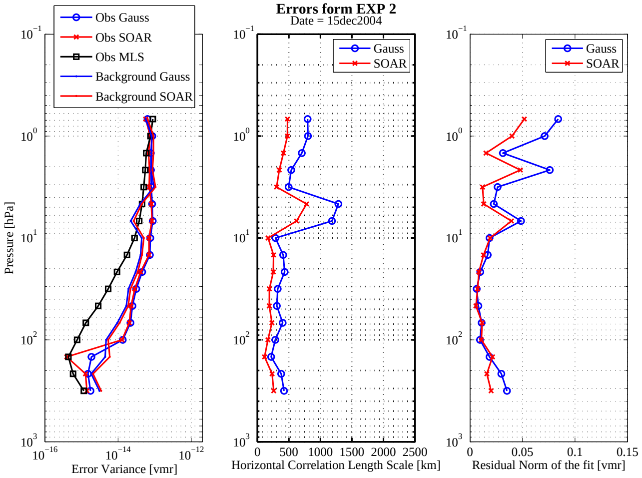
- Step 3: Re-assimilation of EOS MLS (Dec 2004).
  - $\sigma^b$ and $\sigma^o$ provided by tuning of EXP 1
  - Horizontal correlations: SOAR with $L_h = 600$ km
  - Vertical correlations: gaussian with $L_v = 1.5$ level
  - This is EXP 2

Application of Hollingsworth and Lönnberg on BASCOE

- Step 4: Estimating the errors parameters from EXP 2

Application of Hollingsworth and Lönnberg on BASCOE

- Step 4: Estimating the errors parameters from EXP 2



Application of Hollingsworth and Lönnberg on BASCOE

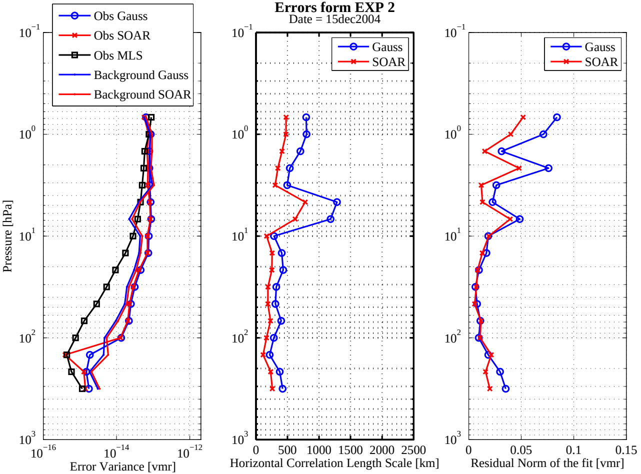
- Step 5: Re-assimilation of EOS MLS (Dec 2004).
  - $\sigma^b$ and $\sigma^o$ provided by tuning of EXP 2
  - Horizontal correlations: SOAR with $L_h$ tuned from EXP 2
  - Vertical correlations: gaussian with $L_v = 1.5$ level
  - This is EXP 3

Application of Hollingsworth and Lönnberg on BASCOE

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

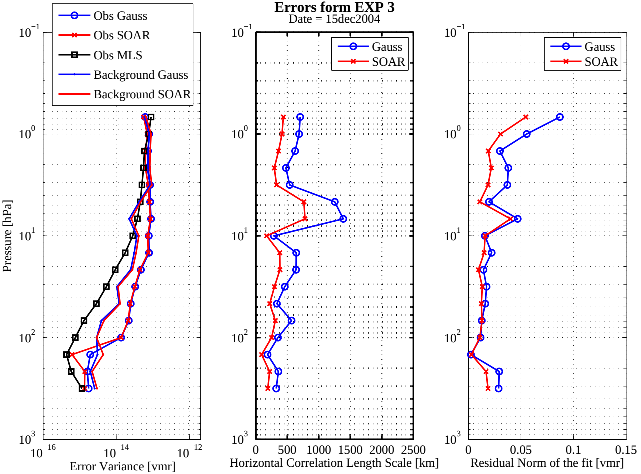
## Application of Hollingsworth and Lönnberg on BASCOE

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



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

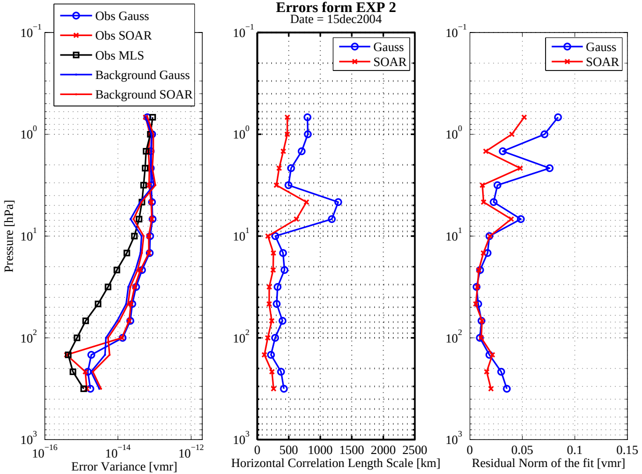
- Step 7: Re-assimilation of EOS MLS (Dec 2004).
  - $\sigma^b$ and $\sigma^o$ provided by tunning of EXP 3
  - Horizontal correlations: SOAR with $L_h$ tuned from EXP 3
  - Vertical correlations: gaussian with $L_v = 1.5$ level
  - This is EXP 4

Application of Hollingsworth and Lönnberg on BASCOE

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

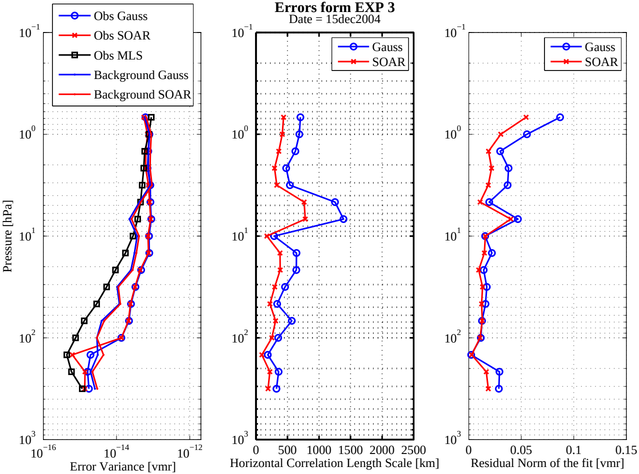
Application of Hollingsworth and Lönnberg on BASCOE

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



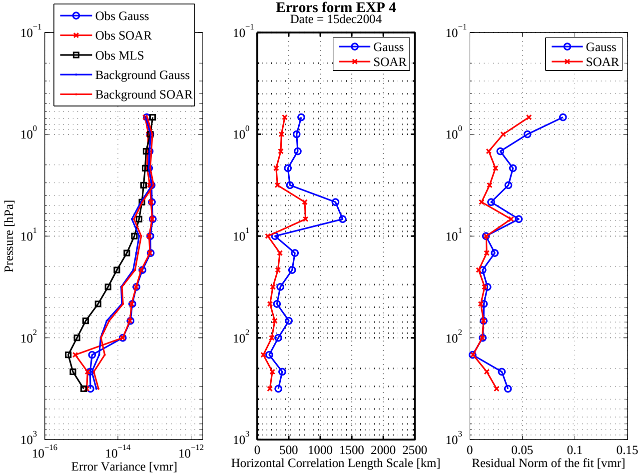
Application of Hollingsworth and Lönnberg on BASCOE

- 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

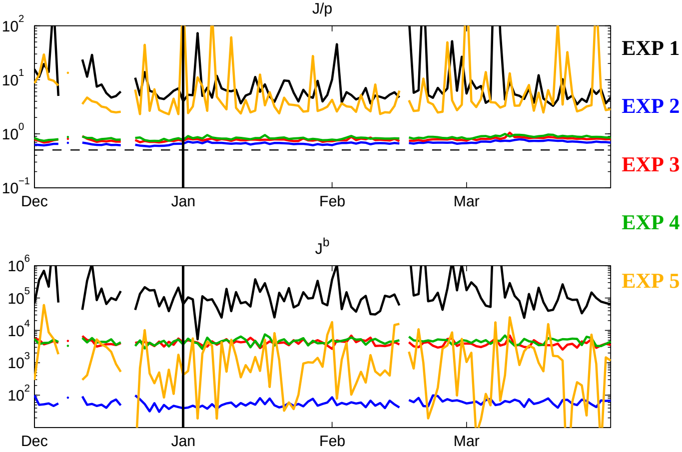
Application of Hollingsworth and Lönnberg on BASCOE

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

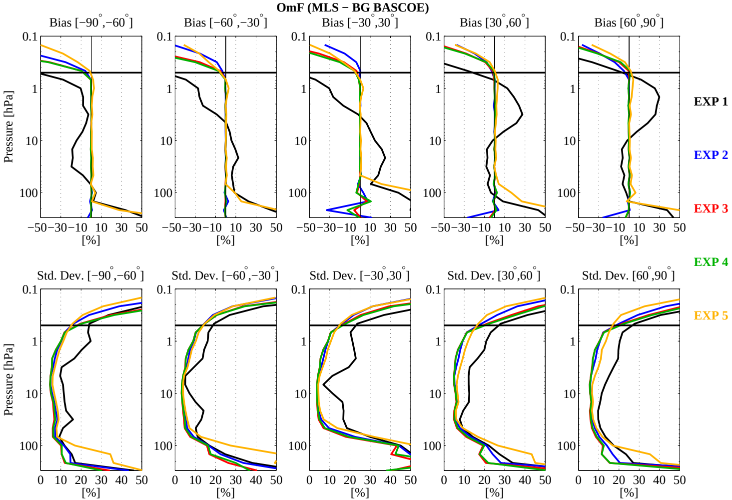
$$
\begin{aligned}
J(\mathbf{x}(t_0)) &= \frac{1}{2}[\mathbf{x}(t_0) - \mathbf{x}^{\mathrm{b}}(t_0)]^T \mathbf{B}^{-1}[\mathbf{x}(t_0) - \mathbf{x}^{\mathrm{b}}(t_0)] \\
&+ \frac{1}{2}\sum_{i=0}^{N_{\Delta t}}[\mathbf{y}_i - H(\mathbf{x}(t_i))]^T \mathbf{R}_i^{-1}[\mathbf{y}_i - H(\mathbf{x}(t_i))]
\end{aligned}
$$

Application of Hollingsworth and Lönnberg on BASCOE

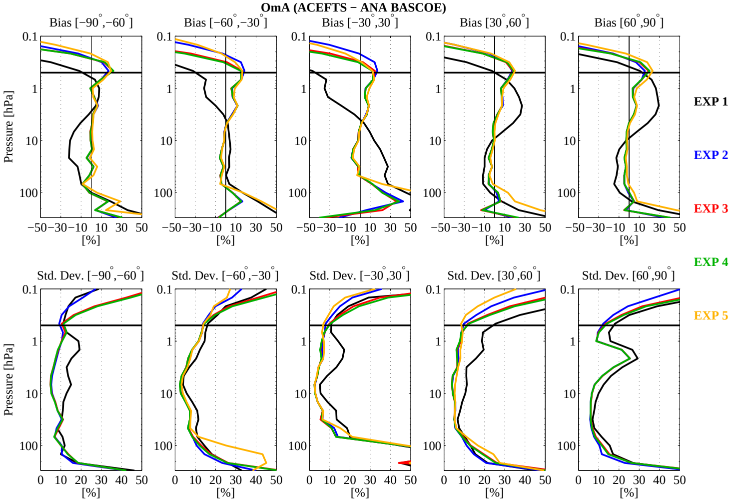
- Step 9: Checking whether $J/p$ has reached its theoretical value of 0.5



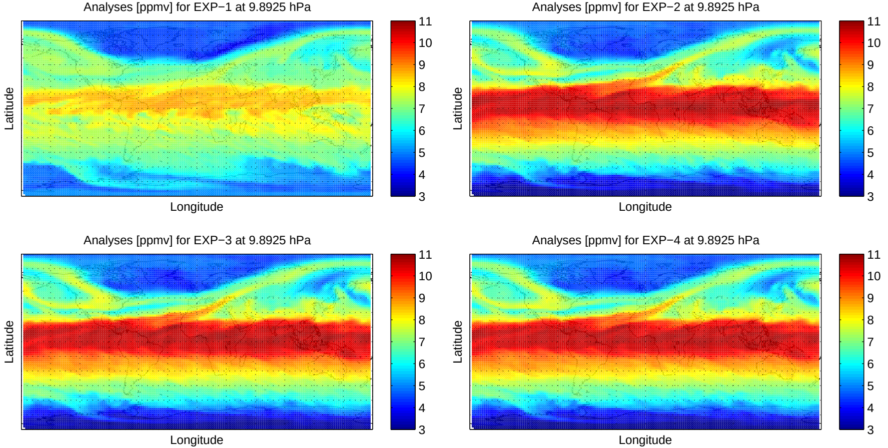
OmF: Deviation of the BASCOE background fields to EOS MLS observations Period considered: 01Jan2005 - 31Mar2005



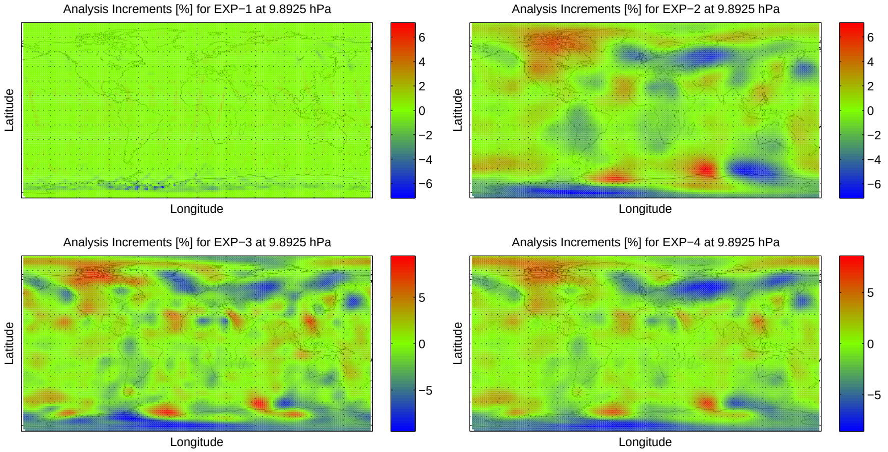
OmA: Deviation of the BASCOE analyses fields from independent ACE-FTS observations Period considered: 01Jan2005 - 31Mar2005



## Analyses on 22 Feb 2005



## Increments on 22 Feb 2005



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

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

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