Performance verification of Complete Data Fusion of ozone profiles in the framework of the AURORA project


n.zoppetti@ifac.cnr.it

IFAC-CNR “Nello Carrara” Institute of Applied Physics, Florence, Italy
AURORA is a project financed by the European Commission in the framework of the Horizon 2020 Framework Program that concerns the application of fusion and assimilation algorithms to simulated VMR ozone profiles in different spectral bands, according to the specifications of atmospheric sentinels 4 (S4) and 5 (S5).

One of the main goals of the project is to evaluate the vantages/disadvantages of the assimilation of FUSED (FUS) products respect to the assimilation of standard simulated (L2) products (see block diagram below).
In this work we show that, if we want to reduce the number of assimilated data without loss of information, CDF must be used instead of arithmetic averages of L2 products.

AURORA is a project financed by the European Commission in the framework of the Horizon 2020 Framework Program that concerns the application of fusion and assimilation algorithms to simulated VMR ozone profiles in different spectral bands, according to the specifications of atmospheric sentinels 4 (S4) and 5 (S5).

One of the main goals of the project is to evaluate the vantages/disadvantages of the assimilation of FUSED (FUS) products respect to the assimilation of standard simulated (L2) products (see block diagram below).
Simulated Products: vertical profiles

- Satellite simulated (nadir) measurements of ozone volume mixing ratio (VMR) in the atmosphere.
- Each measurement is characterized by a **vertical profile**, a Covariance Matrix (CM) and an Averaging Kernels Matrix (AK).

Number of daylight measurements that fall in a $0.5^\circ \times 0.625^\circ$ latitude longitude grid cell in one day.
**Simulated Products: averaging kernels (AK)**

The Averaging Kernel Matrix $A$ is a matrix that describes the sensitivity of the retrieved/simulated profile $\hat{x}$ with respect to the (virtual) true profile $x$.

$$A = \frac{\partial \hat{x}}{\partial x_t}$$

The diagonal elements of AKs describe the information content as a function of the vertical coordinate. The degrees of freedom (DOFs) are a pure number that describes the information content of the measure.

Number of daylight measures that fall in a 0.5° x 0.625° latitude longitude grid cell in one day.

Data simulation formulas
CDF is an a posteriori method that accepts as inputs a generic number of standard retrieval products which can be considered co-located. Each of these products is represented by a profile characterized by its covariance matrix, its averaging kernel matrix and by the a priori information used in the retrieval.

The output of the fusion is a single product that has the same structure of the input ones and that collects all the information of the input products.
Here the CDF has been applied on 1000 coincident L2 pixels that share the same reference virtual true profile, the same AK and the same CM but with different errors $\sigma_x$ that are randomly generated according to a Gaussian distribution with zero mean and the given CM.

The FUSION of 1000 coincident L2 pixels is compared with the original simulated measurements and their arithmetic average.
Conclusions

The verification activity shows that:

- FUS $\Delta$(profiles-true) $\downarrow$ wrt $L_2$ and $<L_2>$
- FUS Total errors $\downarrow$ wrt $L_2$ only (*)
- FUS DOFs $\uparrow$ wrt $L_2$ and $<L_2>$

(*) The arithmetic average $<L_2>$ is biased by the a priori profiles used in the simulated measurements. This is the reason why the ($<L_2>$-true) differences are not consistent with $<L_2>$ total error.

The CDF is the only known algorithm able to correctly combine the information of several coincident measurements into a single product, taking into account the a priori information.
The error $\sigma_x$ is randomly generated according to a Gaussian distribution with zero mean and VCM given by (*)

\[
\tilde{x} = x_a + A_{\hat{x}} (x_t - x_a) + \sigma_x
\]

\[
A_{\hat{x}} = \left( K^T S_y^{-1} K + S_a^{-1} \right)^{-1} K^T S_y^{-1} K = GK
\]

\[
S_{\hat{x}} = \left( K^T S_y^{-1} K + S_a^{-1} \right)^{-1} K^T S_y^{-1} K \left( K^T S_y^{-1} K + S_a^{-1} \right)^{-1}
\]

Where the expression (*) is derived by the propagation through the gain matrix $G$ of an error $\sigma_y$ with zero mean and VCM $= S_y$. 

\[\sigma_y = \gamma \alpha \]
CDF formulas

- Fusion of N measures in perfect spatial coincidence (no coincidence error)
- Common fixed vertical grid (no interpolation error)

\[
\begin{align*}
  i & \quad 1..N \\
  \hat{x}_i & \quad L2 \text{ profiles} \\
  S_i & \quad L2 \text{ CM} \\
  A_i & \quad L2 \text{ AK} \\
  x_{ai} & \quad \text{a priori profile of } i_{th} \text{ product} \\
  x_a & \quad \text{a priori profile to be used for the output} \\
  S_a & \quad \text{CM a priori to be used for the output} \\

  x_f = \left( \sum_{i=1}^{N} A_i^T S_i^{-1} A_i + S_a^{-1} \right) \left( \sum_{i=1}^{N} A_i^T S_i^{-1} a_i + S_a^{-1} x_a \right)^{-1} \\
  \alpha_i = \hat{x}_i - (I - A_i)x_{ai} = A_i x_i + \sigma_i \\
  S_f = \left( \sum_{i=1}^{N} A_i^T S_i^{-1} A_i + S_a^{-1} \right) \left( \sum_{i=1}^{N} A_i^T S_i^{-1} A_i + S_a^{-1} \right)^{-1} \\
  A_f = \left( \sum_{i=1}^{N} A_i^T S_i^{-1} A_i + S_a^{-1} \right) \left( \sum_{i=1}^{N} A_i^T S_i^{-1} A_i + S_a^{-1} \right)^{-1} \\
\end{align*}
\]
Total error of $<\text{L2}>$

$$\hat{x}_i = A_i x_t + \left( I - A_i \right) x_a + \sigma_{i,\text{noise}}$$

$$\sigma_{i,\text{total}} = \hat{x}_i - x_t = \ldots = \left( I - A_i \right) \left( x_a - x_t \right) + \sigma_{i,\text{noise}}$$

$$\sigma_{<\text{L2>},\text{total}} = \left< \hat{x}_i \right> - x_t = \ldots = \left( I - A_i \right) \left( x_a - x_t \right) + \frac{1}{N} \sum_{i=1}^{N} \sigma_{i,\text{noise}}$$

Same true and a priori for all $N$ measurements

Bias

Gaussian term

Only the noise error is reduced by the arithmetic average

$<\text{L2}>$ total error in the figure is the estimate of that does not consider the bias

$$<\text{L2}>_{\text{total error}} = \frac{\sigma_{i,\text{total}}}{\sqrt{N}}$$
**What is the AURORA Call for Ideas?**
It is designed to engage European students in the debate about the use of satellite data to address today’s societal challenges.

**What should you do to win?**
Submit an idea for an application that uses satellite data and can have a positive impact on society. The focus is on Air Quality and UV radiation measurement.

**What could you win?**
10k euro for the winner in the “University and PhD level” category.