Importance of interpolation and coincidence errors in data fusion

Simone Ceccherini [1], Bruno Carli [1], Cecilia Tirelli [1], Nicola Zoppetti [1], Samuele Del Bianco [1], Ugo Cortesi [1], Jukka Kujanpää [2] and Rossana Dragani [3]

[1] *Istituto di Fisica Applicata "Nello Carrara" del Consiglio Nazionale delle Ricerche, Via Madonna del Piano 10, 50019 Sesto Fiorentino (Firenze), Italy*

[2] *Finnish Meteorological Institute, Earth Observation Unit, P.O. Box 503, FI-00101 Helsinki, Finland*

[3] *European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, RG2 9AX, United Kingdom*

AURORA – Advanced Ultraviolet Radiation and Ozone Retrieval for Applications

European Commission

INTRODUCTION

Many remote sensing observations of vertical profiles of atmospheric variables are obtained with instruments operating on space-borne and airborne platforms, as well as from ground-based stations. Recently, the Complete Data Fusion (CDF) method (Ceccherini et al., 2015) was proposed for use in the combination of independent measurements of the same profile in order to exploit all the available information and obtain a comprehensive and concise description of the atmospheric state. This is an a posteriori method that uses standard retrieval products. The CDF products are equivalent to those obtained with a simultaneous retrieval, considered to be the most comprehensive way of exploiting different observations of the same quantity, and offer a simpler implementation. However, so far, the CDF method was mainly applied to measurements performed by the same instrument while sounding the same air sample.

Here, we consider the general problem posed by the application of the CDF method to measurements performed by different instruments that are retrieved on different vertical grids and refer to different true profiles (which correspond to the case of fusing profiles measured in different geolocations). The analysis of this problem suggests a modification of the CDF method, taking into account interpolation and coincidence errors. We determine the expressions of these errors and show how they enter in the CDF formula. The study is performed using simulated measurements of ozone profiles obtained in the ultraviolet and in the thermal infrared in the framework of the Sentinel 4 (S4) mission of the Copernicus programme (http://www.copernicus.eu/main/sentinels).

The results presented in this poster arise from research activities conducted in the framework of the AURORA project (http://www.aurora-copernicus.eu) supported by the Horizon 2020 research and innovation programme of the European Union (call: H2020-EO-2015; topic: EO-2-2015) under grant agreement no. 687428.

THE CDF METHOD

Let us assume to have N independent and simultaneous measurements of the vertical profile of an atmospheric target referred to the same space-time location. Performing the retrieval of the N measurements with the optimal estimation method, we obtain N vectors $\hat{\mathbf{x}}_i$ (i=1, 2, …, N) here assumed to be estimates of the profiles made on a common vertical grid. The vectors $\hat{\mathbf{x}}_i$ are characterized by the covariance matrices (CMs) $\mathbf{S}_i$ and the averaging kernel matrices (AKMs) $\mathbf{A}_i$. The CDF solution for the considered profiles is given by

$$\mathbf{x}_f = \left(\sum_{i=1}^{N} \mathbf{A}_i^T \mathbf{S}_i^{-1} \mathbf{A}_i + \mathbf{S}_a^{-1} \right)^{-1} \left(\sum_{i=1}^{N} \mathbf{A}_i^T \mathbf{S}_i^{-1} \boldsymbol{\alpha}_i + \mathbf{S}_a^{-1} \mathbf{x}_a \right),$$

where

$$\boldsymbol{\alpha}_i \equiv \hat{\mathbf{x}}_i - (\mathbf{I} - \mathbf{A}_i) \mathbf{x}_{ai},$$

$\mathbf{x}_{ai}$ is the a priori profile used in the i-th retrieval, $\mathbf{I}$ is the identity matrix, $\mathbf{x}_a$ and $\mathbf{S}_a$ are the a priori profile and its CM used to constrain the data fusion.

When the fusing profiles $\hat{\mathbf{x}}_i$ are represented on different vertical grids it is necessary to perform a resampling of the AKMs, which makes their second index equal to that of the grid of the fusion product (in the following referred to as the *fusion grid*):

$$\mathbf{A}_i' = \mathbf{A}_i \mathbf{R}_i,$$

where $\mathbf{R}_i$ are the generalized inverse matrices of the linear interpolation matrices $\mathbf{H}_i$, which interpolate on the fusion grid the profiles obtained on different grids.

APPLICATION OF THE CDF METHOD TO PROFILES RETRIEVED ON DIFFERENT VERTICAL GRIDS AND RELATED TO DIFFERENT TRUE PROFILES

We test the CDF method on simulated data of S4. We simulate two S4 ozone vertical profile measurements as they could be obtained from the Infrared Sounder in the thermal infrared and from the Ultraviolet, Visible and Near-Infrared Sounding spectrometer in the ultraviolet on board the Meteosat Third Generation satellite. We refer to these two simulated measurements as TIR measurement and UV measurement, respectively.

In order to evaluate the effect of the variability of vertical grids and of true profiles, three cases are considered:

1) The simulated measurements refer to the same true profile and are retrieved on the same vertical grid.
2) The simulated measurements refer to the same true profile but are retrieved on different vertical grids.
3) The simulated measurements refer to different true profiles and are retrieved on the same vertical grid.

In all three cases, the true profile and the vertical grid of the UV measurement are kept fixed and, when pertinent, are changed for the TIR measurement. For simplicity, we define the fusion grid to coincide with the fixed grid of the UV measurement.

For a meaningful comparison of the quality of fusing and fused profiles, we used common a priori profiles and common a priori CMs.

The results obtained in the three test cases are reported in Figures 1-3. We observe that, while in case 1 the differences between the profile obtained from the fusion and the mean of the true profiles are smaller than, or comparable to, those of the profiles obtained from the TIR and UV measurements, in cases 2 and 3 these differences are significantly larger. Therefore, in cases 2 and 3 the fusion provides a product of poorer quality than that of the single products.

The problem encountered in case 2 is due to the fact that the data fusion is made using estimates of the AKMs on the fusion grid obtained by interpolation of the original AKMs, which are only an approximation of the real AKMs on the fusion grid. We refer to this effect as *interpolation error*. The problem encountered in case 3 is related to different true profiles and we refer to this effect as *coincidence error* because it occurs when fusing profiles that do not correspond to the same space-time location. These are additional errors, that are not accounted for in the CM of the measurements, and must be considered in the fusion procedure.

INTRODUCTION OF INTERPOLATION AND COINCIDENCE ERRORS

The CDF formula can be modified to account for the interpolation and coincidence errors by replacing $\boldsymbol{\alpha}_i$ with

$$\tilde{\boldsymbol{\alpha}}_i = \boldsymbol{\alpha}_i - \mathbf{A}_i \left(\mathbf{C}^{(i)} - \mathbf{R}_i \mathbf{C}^{(f)} \right) \mathbf{x}_a,$$

where $\mathbf{C}^{(i)}$ and $\mathbf{C}^{(f)}$ are the sampling matrices from a fine grid, that includes all the levels of the fusion grid (f) and of the N grids (i), to the grids (i) and to the grid (f), respectively, and $\mathbf{S}_i$ with

$$\tilde{\mathbf{S}}_i = \mathbf{S}_i + \mathbf{S}_{i,\mathrm{int}} + \mathbf{S}_{i,coin}.$$

The interpolation error is characterized by the CM:

$$\mathbf{S}_{i,\mathrm{int}} = \mathbf{A}_i \left(\mathbf{C}^{(i)} - \mathbf{R}_i \mathbf{C}^{(f)} \right) \mathbf{S}_a \left(\mathbf{C}^{(i)} - \mathbf{R}_i \mathbf{C}^{(f)} \right)^T \mathbf{A}_i^T.$$

The coincidence error is characterized by the CM:

$$\mathbf{S}_{i,coin} = \mathbf{A}_i \mathbf{C}^{(i)} \mathbf{S}_{coin} \mathbf{C}^{(i)T} \mathbf{A}_i^T,$$

where we introduce the CM $\mathbf{S}_{coin}$ that accounts for the dispersion of the true profiles, and, therefore, depends on the coincidence criteria.

TESTS WITH THE UPGRADED METHOD

The test cases of fusion 2 and 3 were repeated with the modified method. For case 3 we used a CM $\mathbf{S}_{coin}$, equal for both TIR and UV measurements, obtained considering an error of 5% of the a priori profile and a correlation length of 6 km.

In Figures 4 and 5, we report the noise errors, the interpolation errors and the coincidence errors related, respectively, to case 2 and case 3, for both TIR and UV measurements.

Figures 6 and 7 show the fused profiles and the residuals obtained with the modified algorithm compared with the same quantities reported in panels (b) and (c) of Figs. 2 and 3, respectively. In both tests, the modified method provides residuals that are significantly smaller than those obtained with the original CDF method.

These tests show that the upgrade of the CDF method solves the problems which occur when either the fusing profiles are retrieved on different vertical grids or they refer to different true profiles. The modified method is a generalization of the CDF that allows its application to a wide range of cases.

ORIGINAL METHOD

Case 1):
same true profile
same vertical grid

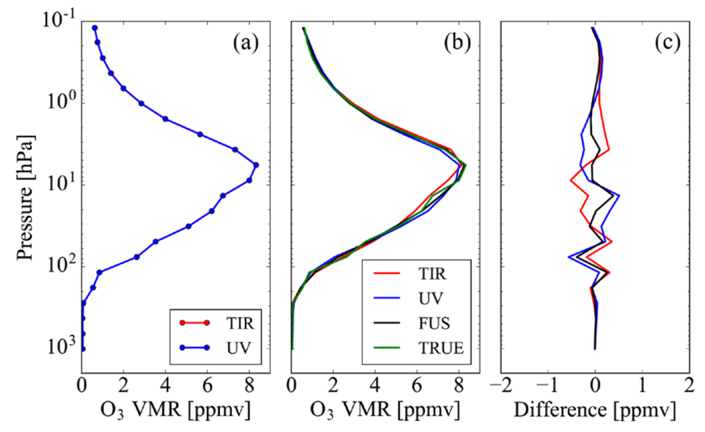

Figure 1. (a) True ozone profiles related to TIR (red line) and UV (blue line) measurements. (b) Ozone profiles obtained from TIR measurement (red line), from UV measurement (blue line), from the data fusion (black line) compared with the mean value of the true profiles (green line). (c) Residual errors obtained as differences of the ozone profiles obtained from TIR measurement (red line), from UV measurement (blue line) and from data fusion (black line) from the mean value of true profiles. All the reported quantities are related to case 1.

UPGRADED METHOD

Case 2):
same true profile
different vertical grids

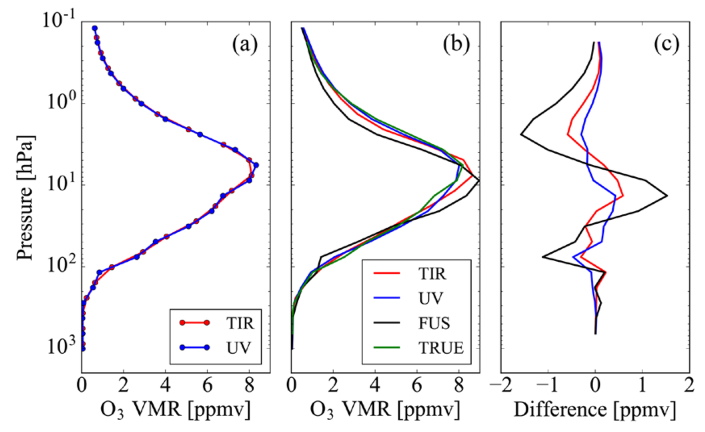

Figure 2. As Fig. 1 but for case 2.

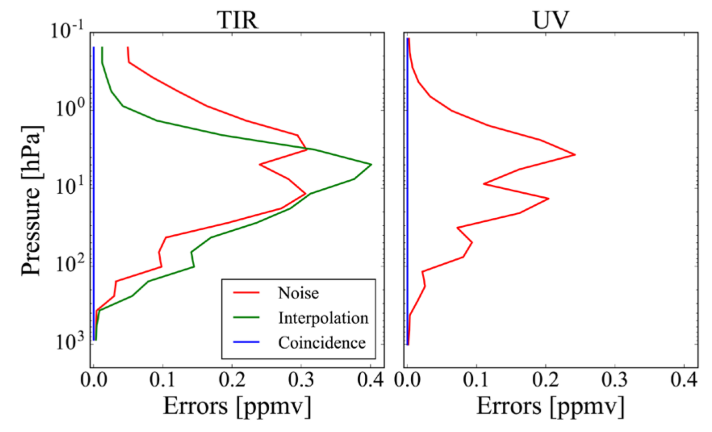

Figure 4. Noise errors (red lines), interpolation errors (green lines) and coincidence errors (blue lines) in case 2 for TIR and UV measurements.

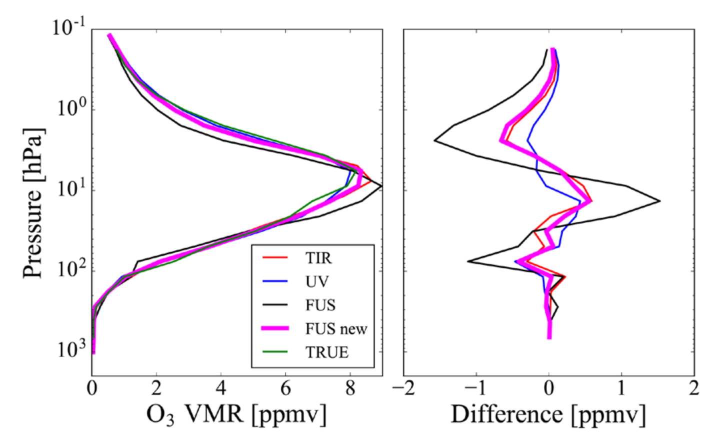

Figure 6. The fused profile and the residual error obtained with the modified algorithm (magenta lines) compared with the same quantities of Fig. 2b and c.

Case 3):
different true profiles
same vertical grid

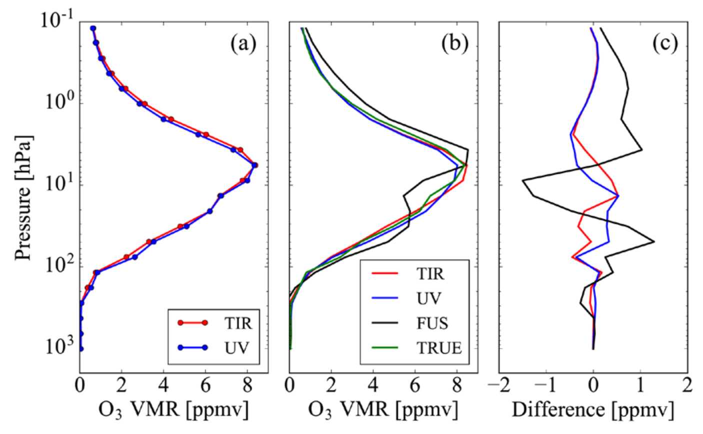

Figure 3. As Fig. 1 but for case 3.

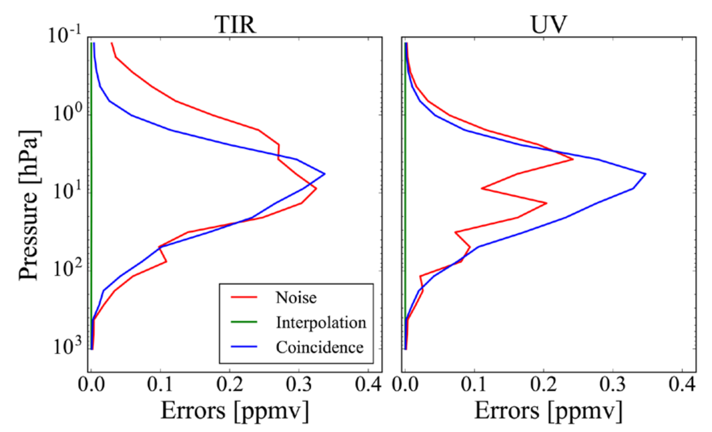

Figure 5. As Fig. 4 but for case 3.

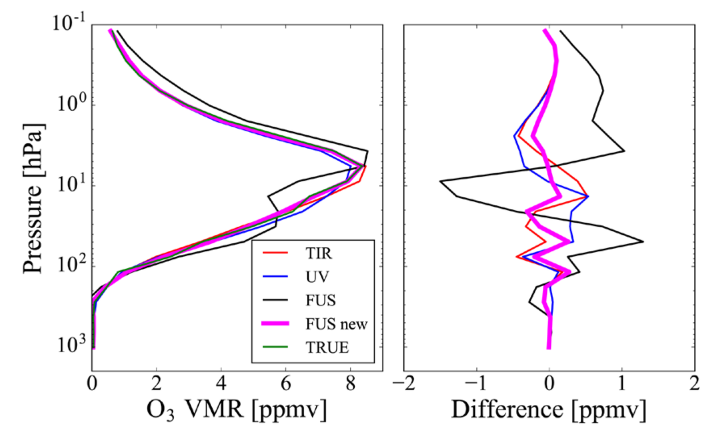

Figure 7. The fused profile and the residual error obtained with the modified algorithm (magenta lines) compared with the same quantities of Fig. 3b and c.

CONCLUSIONS

The study showed that the CDF algorithm works well when the fusing profiles are represented on the same vertical grid and refer to the same true profile, otherwise the algorithm provides products of degraded quality. To address this shortcoming, a generalization of the CDF method, which takes into account interpolation and coincidence errors, was presented. This upgrade overcomes the encountered problems and provides products of good quality when the fusing profiles are both retrieved on different vertical grids and referred to different true profiles.

REFERENCES

S. Ceccherini, B. Carli and P. Raspollini, *Equivalence of data fusion and simultaneous retrieval*, Optics Express, **23**, 8476-8488 (2015).

S. Ceccherini, B. Carli, C. Tirelli, N. Zoppetti, S. Del Bianco, U. Cortesi, J. Kujanpää and R. Dragani, *Importance of interpolation and coincidence errors in data fusion*, Atmospheric Measurement Techniques, **11**, 1009-1017 (2018).