https://doi.org/10.5194/amt-2022-182
Preprint. Discussion started: 29 August 2022

Atmospheric Measurement Techniques
Discussions

An improved formula for the Complete Data Fusion

Simone Ceccherini, Nicola Zoppetti, and Bruno Carli

Istituto di Fisica Applicata "Nello Carrara" del Consiglio Nazionale delle Ricerche, Via Madonna del Piano 10, 50019 Sesto Fiorentino, Italy

Correspondence to: Simone Ceccherini (S.Ceccherini@ifac.cnr.it) and Nicola Zoppetti (N.Zoppetti@ifac.cnr.it)

Abstract. The Complete Data Fusion is a method that combines independent measurements of an atmospheric vertical profile. Recently a new formula for the Complete Data Fusion, which does not contain matrices that can be singular and overcomes the generalized inverse approximation used when singular matrices have to be inverted, has been proposed. We show that the new formula is a generalization of the original one and analyze the analytical relationship between the two formulas when generalized inverse matrices are used for singular matrices. We extend the new formula to include interpolation and coincidence errors, which must be considered when the profiles to be fused are measured on different vertical grids and at either different times or locations. Finally, we use a real measurement of the IASI instrument to show the improved performances of the new formula with respect to the original one.

1 Introduction

The Complete Data Fusion (CDF) is a new data processing method that allows to combine several independent measurements of an atmospheric vertical profile (Ceccherini et al., 2015), and more generally of any vectorial quantity that is retrieved using the optimal estimation method (Rodgers, 2000). It is called ”complete” for its capability of considering all the features of the measurements that are being combined, that is not only their errors, but also their vertical resolution. The inputs of the method are the profiles retrieved from the individual measurements using the optimal estimation method together with their a priori profiles, averaging kernel matrices (AKMs) and noise covariance matrices (CMs), and an a priori profile with its CM is used to constrain the fused profile. The output of the method is a single profile (the fused profile) with its AKM and CM. The a priori information used to constrain the fused profile can be freely chosen independently of the a priori information used in the retrievals of the individual profiles. The method is equivalent to the simultaneous retrieval of all the measurements that are combined when the linear approximation of the forward models is appropriate in the variability range of the results of the individual retrievals.

The method has been extended to fuse profiles retrieved on different vertical grids for which an interpolation on a common grid is needed and to deal with measurements obtained either at different times or from different platforms and, therefore, referred to different true profiles. This extension required the introduction of interpolation and coincidence errors in the fusion process (Ceccherini et al., 2018).

The performances of the method have been studied on ozone profiles retrieved from simulated measurements in the ultraviolet, visible, and thermal infrared spectral ranges for the Sentinel-4 and Sentinel-5 missions of the Copernicus program (Tirelli et al., 2020, Zoppetti et al., 2021). The results of these studies show that the CDF is able to provide products of improved quality with respect to the input products in terms of reduced errors and increased number of degrees of freedom.

A problem connected with the application of the CDF formula is the presence of the inverse matrices of the noise CMs of the input profiles and this implies that the formula can be rigorously applied only when the noise CMs are nonsingular. When the profiles are retrieved solving ill-posed inverse problems (which is a very common case), this condition is not satisfied. In this case, we can still apply the CDF formula replacing the inverse matrices of the noise CMs with the generalized inverse matrices (Kalman, 1976), but the result is an approximation. Furthermore, a practical problem in the use of the generalized inverse matrices is the definition of the threshold for the eigenvalues for which eigenvalues smaller than this threshold have their inverses replaced with zeros. Too small values for this threshold determine significant numeric noise in the products; on the other hand, too large values of this threshold determine a loss of useful information.

Recently, following the approach of the Kalman filter (Rodgers, 2000), a different formula for the CDF has been derived (Ceccherini, 2021). This formula contains the inverse matrices of the retrieval error CMs, which include both the noise and the smoothing errors, instead of the inverse matrices of the noise CMs. Differently from the noise CMs, the retrieval error CMs are always nonsingular matrices and the new formula can be used without having to resort to the use of generalized inverse matrices.

In this paper we introduce the new formula showing that it is a generalization of the original CDF formula given in Ceccherini et al. (2015) and analytically analyze the differences between the new formula and the original one when the generalized inverse matrices are used for the inverse of the noise CMs. Since in the application of the CDF to real measurements it is common practice to interpolate between different grids and to consider not perfect coincidence of the fusing profiles, the new formula is also used to derive the operational expression that takes into account interpolation and coincidence errors.

Finally, we use a measurement of the IASI instrument (Clerbaux et al., 2009) to show the improved performances of the new formula with respect to the original one in the case of real data.

In Section 2, we show that the new formula is a generalization of the original one and extend it to handle the cases where coincidence and interpolation errors are present. In Section 3, we compare the performances of the two formulas using an IASI measurement and in Section 4 we draw the conclusions.

2 Theoretical analysis of the CDF formula

2.1 The new formula as a generalization of the original one

We assume to have *N* profiles $\hat{\mathbf{x}}_i$ retrieved on the same vertical grid with the optimal estimation method (Rodgers, 2000) from *N* independent measurements of a true atmospheric profile $\mathbf{x}_\mathrm{t}$. The profiles $\hat{\mathbf{x}}_i$ are characterized by the AKMs $\mathbf{A}_i = \frac{\partial \hat{\mathbf{x}}_i}{\partial \mathbf{x}_\mathrm{t}}$, which measure the sensitivities of the profiles $\hat{\mathbf{x}}_i$ to $\mathbf{x}_\mathrm{t}$ and by the CMs $\mathbf{S}_i$, which measure the retrieval errors.

Before introducing the new formula for the CDF, let us recall some useful relationships. The quantities $\mathbf{A}_i$ and $\mathbf{S}_i$ can be written as a function of the two quantities that characterize the retrievals, that is the Fisher information matrices (Fisher, 1935) $\mathbf{F}_i = \mathbf{K}_i^T \mathbf{S}_{\mathrm{n}\mathbf{y}_i}^{-1} \mathbf{K}_i$ ($\mathbf{K}_i$ being the Jacobian matrices of the forward models and $\mathbf{S}_{\mathrm{n}\mathbf{y}_i}$ the CMs of the noise errors of the measured radiances $\mathbf{y}_i$), which characterize the measurements, and the a priori CMs $\mathbf{S}_{\mathrm{a}i}$ used in the retrievals, which characterize the constraints. The expressions of $\mathbf{A}_i$ and $\mathbf{S}_i$ as a function of these two quantities are:

$$\mathbf{A}_i = \left(\mathbf{F}_i + \mathbf{S}_{\mathrm{a}i}^{-1}\right)^{-1} \mathbf{F}_i \qquad (1)$$

$$\mathbf{S}_i = \left(\mathbf{F}_i + \mathbf{S}_{\mathrm{a}i}^{-1}\right)^{-1}. \qquad (2)$$

We also recall that the $\mathbf{S}_i$ are the sum of two contributions: $\mathbf{S}_{\mathrm{n}i}$, the CMs of the noise errors, and $\mathbf{S}_{\mathrm{s}i}$, the CMs of the smoothing errors, that are respectively equal to:

$$\mathbf{S}_{\mathrm{n}i} = \left(\mathbf{F}_i + \mathbf{S}_{\mathrm{a}i}^{-1}\right)^{-1} \mathbf{F}_i \left(\mathbf{F}_i + \mathbf{S}_{\mathrm{a}i}^{-1}\right)^{-1} \qquad (3)$$

$$\mathbf{S}_{si}=\left(\mathbf{F}_i+\mathbf{S}_{ai}^{-1}\right)^{-1}\mathbf{S}_{ai}^{-1}\left(\mathbf{F}_i+\mathbf{S}_{ai}^{-1}\right)^{-1} \tag{4}$$

and as we can see from Eq. (2), the inverse matrices of $\mathbf{S}_i$ always exist.

The new formula for the CDF was obtained using the Kalman filter (Rodgers, 2000) in Ceccherini (2021) in the case of the fusion of two profiles. With an iterative procedure that adds one by one the extra profiles to the fused product, it can be generalized to the fusion of N retrieved profiles $\hat{\mathbf{x}}_i$ with the following formula:

$$\mathbf{x}_\mathrm{f}=\left(\sum_{i=1}^{N}\mathbf{S}_i^{-1}\mathbf{A}_i+\mathbf{S}_\mathrm{a}^{-1}\right)^{-1}\left(\sum_{i=1}^{N}\mathbf{S}_i^{-1}\boldsymbol{\alpha}_i+\mathbf{S}_\mathrm{a}^{-1}\mathbf{x}_\mathrm{a}\right), \tag{5}$$

where $\mathbf{x}_\mathrm{f}$ is the fused profile, $\mathbf{x}_\mathrm{a}$ and $\mathbf{S}_\mathrm{a}$ are the a priori profile and its CM used to constrain the fused profile and

$$\boldsymbol{\alpha}_i=\hat{\mathbf{x}}_i-\mathbf{x}_{ai}+\mathbf{A}_i\mathbf{x}_{ai}\,, \tag{6}$$

$\mathbf{x}_{ai}$ being the a priori profiles used in the retrievals of the individual $\hat{\mathbf{x}}_i$, in general different among them and from $\mathbf{x}_\mathrm{a}$. In the following we refer to the CDF formula given in Eq. (5) as CDF(2021).

From Eqs. (1-3) we see that we can express $\mathbf{S}_{ni}$ in terms of $\mathbf{A}_i$ and $\mathbf{S}_i$

$$\mathbf{S}_{ni}=\mathbf{S}_i\mathbf{A}_i^T=\mathbf{A}_i\mathbf{S}_i \tag{7}$$

and in the hypothesis that the CMs of the noise errors are nonsingular matrices we can obtain $\mathbf{S}_i^{-1}$

$$\mathbf{S}_i^{-1}=\mathbf{A}_i^T\mathbf{S}_{ni}^{-1}. \tag{8}$$

Substituting them in Eq. (5) we obtain the original formula for the CDF given in Ceccherini et al. (2015).

$$\mathbf{x}_\mathrm{f}=\left(\sum_{i=1}^{N}\mathbf{A}_i^T\mathbf{S}_{ni}^{-1}\mathbf{A}_i+\mathbf{S}_\mathrm{a}^{-1}\right)^{-1}\left(\sum_{i=1}^{N}\mathbf{A}_i^T\mathbf{S}_{ni}^{-1}\boldsymbol{\alpha}_i+\mathbf{S}_\mathrm{a}^{-1}\mathbf{x}_\mathrm{a}\right), \tag{9}$$

which, differently from Eq. (5), holds only in the case that the CMs of the noise errors $\mathbf{S}_{ni}$ are nonsingular matrices. Therefore, Eq. (5) is more general than Eq. (9). In the following we refer to the CDF formula given in Eq. (9) as CDF(2015).

As already stated, the output of the CDF is not only the fused profile, but also its AKM and CM. The AKM and the CM of the fused profile calculated using Eq. (9) also contained the inverse of $\mathbf{S}_{ni}$ in the formulas (Ceccherini et al., 2015). We can now calculate these quantities for the products of Eq. (5) aiming at obtaining expressions that do not contain the inverse of matrices that may be singular. From Eq. (5) the AKM of $\mathbf{x}_\mathrm{f}$ is given by:

$$\mathbf{A}_f=\frac{\partial\mathbf{x}_\mathrm{f}}{\partial\mathbf{x}_\mathrm{t}}=\left(\sum_{i=1}^{N}\mathbf{S}_i^{-1}\mathbf{A}_i+\mathbf{S}_\mathrm{a}^{-1}\right)^{-1}\sum_{i=1}^{N}\mathbf{S}_i^{-1}\frac{\partial\boldsymbol{\alpha}_i}{\partial\mathbf{x}_\mathrm{t}}=\left(\sum_{i=1}^{N}\mathbf{S}_i^{-1}\mathbf{A}_i+\mathbf{S}_\mathrm{a}^{-1}\right)^{-1}\sum_{i=1}^{N}\mathbf{S}_i^{-1}\mathbf{A}_i\,, \tag{10}$$

where we have used Eq. (6) for the calculation of the derivative.

The noise CM of $\mathbf{x}_\mathrm{f}$ is obtained exploiting the fact that the noise CMs of $\boldsymbol{\alpha}_i$ are $\mathbf{S}_{ni}$, therefore,

$$\mathbf{S}_\mathrm{nf}=\left(\sum_{i=1}^{N}\mathbf{S}_i^{-1}\mathbf{A}_i+\mathbf{S}_\mathrm{a}^{-1}\right)^{-1}\sum_{i=1}^{N}\mathbf{S}_i^{-1}\mathbf{S}_{ni}\mathbf{S}_i^{-1}\left(\sum_{i=1}^{N}\mathbf{S}_i^{-1}\mathbf{A}_i+\mathbf{S}_\mathrm{a}^{-1}\right)^{-1}. \tag{11}$$

Substituting $\mathbf{S}_{ni}$ given in Eq. (7) in Eq. (11), we obtain

$$\mathbf{S}_\mathrm{nf}=\left(\sum_{i=1}^{N}\mathbf{S}_i^{-1}\mathbf{A}_i+\mathbf{S}_\mathrm{a}^{-1}\right)^{-1}\sum_{i=1}^{N}\mathbf{S}_i^{-1}\mathbf{A}_i\left(\sum_{i=1}^{N}\mathbf{S}_i^{-1}\mathbf{A}_i+\mathbf{S}_\mathrm{a}^{-1}\right)^{-1}. \tag{12}$$

The CM of $\mathbf{x}_\mathrm{f}$ is obtained adding to Eq. (12) the CM of the smoothing errors

$$\mathbf{S}_{\mathrm{sf}} = \left(\sum_{i=1}^{N} \mathbf{S}_i^{-1}\mathbf{A}_i + \mathbf{S}_{\mathrm{a}}^{-1}\right)^{-1} \mathbf{S}_{\mathrm{a}}^{-1} \left(\sum_{i=1}^{N} \mathbf{S}_i^{-1}\mathbf{A}_i + \mathbf{S}_{\mathrm{a}}^{-1}\right)^{-1}, \tag{13}$$

obtaining

$$\mathbf{S}_{\mathrm{f}} = \mathbf{S}_{\mathrm{nf}} + \mathbf{S}_{\mathrm{sf}} = \left(\sum_{i=1}^{N} \mathbf{S}_i^{-1}\mathbf{A}_i + \mathbf{S}_{\mathrm{a}}^{-1}\right)^{-1}. \tag{14}$$

2.2 Relationship between CDF(2021) and CDF(2015) with generalized inverse matrices

In the introduction we mentioned that, using the approximation of the generalized inverse matrices (Kalman, 1976), the original formula CDF(2015) can also be used in the case of $\mathbf{S}_{ni}$ singular. Therefore, in this Section, we investigate the differences between CDF(2021) and CDF(2015) when in the latter the generalized inverse matrices of $\mathbf{S}_{ni}$ are used. In Eq. (9) we replace the matrices $\mathbf{S}_{\mathrm{n}i}^{-1}$ with the generalized inverse matrices $\mathbf{S}_{ni}^{\#}$

$$\mathbf{x}_{\mathrm{f}} = \left(\sum_{i=1}^{N} \mathbf{A}_i^T \mathbf{S}_{\mathrm{n}i}^{\#}\mathbf{A}_i + \mathbf{S}_{\mathrm{a}}^{-1}\right)^{-1} \left(\sum_{i=1}^{N} \mathbf{A}_i^T \mathbf{S}_{\mathrm{n}i}^{\#}\boldsymbol{\alpha}_i + \mathbf{S}_{\mathrm{a}}^{-1}\mathbf{x}_{\mathrm{a}}\right). \tag{15}$$

$\mathbf{S}_{\mathrm{n}i}^{\#}$ appear in two terms. For the first term it has already been demonstrated in the appendix of Ceccherini et al. (2012) that

$$\mathbf{A}_i^T \mathbf{S}_{\mathrm{n}i}^{\#}\mathbf{A}_i = \mathbf{F}_i = \mathbf{S}_i^{-1}\mathbf{A}_i\,, \tag{16}$$

where the second equality follows from Eqs. (1) and (2). Therefore, the first term is equal in the two CDF formulas.

We can elaborate the second term using Eqs. (1-3)

$$\mathbf{A}_i^T \mathbf{S}_{\mathrm{n}i}^{\#} = \mathbf{F}_i\left(\mathbf{F}_i + \mathbf{S}_{\mathrm{a}i}^{-1}\right)^{-1}\mathbf{S}_{\mathrm{n}i}^{\#} = \left(\mathbf{F}_i + \mathbf{S}_{\mathrm{a}i}^{-1}\right)\left(\mathbf{F}_i + \mathbf{S}_{\mathrm{a}i}^{-1}\right)^{-1}\mathbf{F}_i\left(\mathbf{F}_i + \mathbf{S}_{\mathrm{a}i}^{-1}\right)^{-1}\mathbf{S}_{\mathrm{n}i}^{\#} = \mathbf{S}_i^{-1}\mathbf{S}_{\mathrm{n}i}\mathbf{S}_{\mathrm{n}i}^{\#}\,, \tag{17}$$

which, in general, are different from $\mathbf{S}_i^{-1}$, because $\mathbf{S}_{ni}\mathbf{S}_{ni}^{\#}$ are different from the identity matrices when $\mathbf{S}_{ni}$ are singular matrices.

Therefore, in the case of singular $\mathbf{S}_{ni}$, the CDF(2015) used with the generalized inverse matrices of $\mathbf{S}_{ni}$, Eq. (15), is equivalent to

$$\mathbf{x}_{\mathrm{f}} = \left(\sum_{i=1}^{N} \mathbf{S}_i^{-1}\mathbf{A}_i + \mathbf{S}_{\mathrm{a}}^{-1}\right)^{-1} \left(\sum_{i=1}^{N} \mathbf{S}_i^{-1}\mathbf{S}_{\mathrm{n}i}\mathbf{S}_{ni}^{\#}\boldsymbol{\alpha}_i + \mathbf{S}_{\mathrm{a}}^{-1}\mathbf{x}_{\mathrm{a}}\right). \tag{18}$$

Therefore, the CDF(2015) used with the generalized inverse matrices is an approximation of the more rigorous CDF(2021) and the quality of the approximation depends on how much $\mathbf{S}_{\mathrm{n}i}\mathbf{S}_{\mathrm{n}i}^{\#}$ is close to the identity matrix.

2.3 The new formula in presence of coincidence and interpolation errors

We know that in the applications of the CDF to real measurements it is often necessary to fuse vertical profiles measured on different grids and at either different times or locations, so that, interpolation and coincidence errors must also be considered. The expression of the CDF with interpolation and coincident errors, that can be called the operational CDF, was calculated in Ceccherini et al. (2018) and was derived from the equation of the CDF(2015) that, as we have seen above, is not valid when there are singular matrices. In this Section, we show how the expression of the operational CDF, can be written in a more general form, using the CDF(2021) exploiting the equivalence of CDF(2015) and CDF(2021) in the case that the CMs of the noise errors are nonsingular.

We start from the formula that deals with interpolation and coincidence errors, given in Ceccherini et al., (2018) based on the CDF(2015) and equal to:

$$\mathbf{x}_\mathrm{f} = \left(\sum_{i=1}^{N} \mathbf{R}_i^T \mathbf{A}_i^T \tilde{\mathbf{S}}_{\mathrm{n}i}^{-1} \mathbf{A}_i \mathbf{R}_i + \mathbf{S}_\mathrm{a}^{-1} \right)^{-1} \left(\sum_{i=1}^{N} \mathbf{R}_i^T \mathbf{A}_i^T \tilde{\mathbf{S}}_{\mathrm{n}i}^{-1} \tilde{\boldsymbol{\alpha}}_i + \mathbf{S}_\mathrm{a}^{-1} \mathbf{x}_\mathrm{a} \right), \tag{19}$$

where $\mathbf{R}_i$ are the generalized inverse matrices of the linear interpolation matrices $\mathbf{H}_i$, which interpolate the profiles from the retrieval grids to the fusion grid. Furthermore,

$$\tilde{\boldsymbol{\alpha}}_i = \boldsymbol{\alpha}_i - \mathbf{A}_i \left(\mathbf{C}^{(i)} - \mathbf{R}_i \mathbf{C}^{(\mathrm{f})} \right) \mathbf{x}_\mathrm{a,fine} \tag{20}$$

$$\tilde{\mathbf{S}}_{\mathrm{n}i} = \mathbf{S}_{\mathrm{n}i} + \mathbf{A}_i \left(\mathbf{C}^{(i)} - \mathbf{R}_i \mathbf{C}^{(\mathrm{f})} \right) \mathbf{S}_\mathrm{a,fine} \left(\mathbf{C}^{(i)} - \mathbf{R}_i \mathbf{C}^{(\mathrm{f})} \right)^T \mathbf{A}_i^T + \mathbf{A}_i \mathbf{C}^{(i)} \mathbf{S}_\mathrm{coin} \mathbf{C}^{(i)T} \mathbf{A}_i^T , \tag{21}$$

where $\mathbf{x}_\mathrm{a,fine}$ is the a priori profile used to constrain the data fusion represented on a fine grid that includes all the levels of the fusion grid and of the N retrievals grids. $\mathbf{C}^{(i)}$ and $\mathbf{C}^{(\mathrm{f})}$ are the sampling matrices from this fine grid to the grid of the i-th retrieval and to the fusion grid, respectively. $\mathbf{S}_\mathrm{a,fine}$ and $\mathbf{S}_\mathrm{coin}$ are respectively the fusion a priori CM and the CM describing the variability of the true profiles related to the measurements that we fuse: both CMs are represented on the fine grid. The same limit of Eq. (9) applies also to Eq. (19) that, evidently, can be written only in the hypothesis that $\tilde{\mathbf{S}}_{\mathrm{n}i}$ are nonsingular matrices.

In order to write an equation similar to Eq. (7) for $\tilde{\mathbf{S}}_{\mathrm{n}i}$, we define the matrix $\tilde{\mathbf{S}}_i$

$$\tilde{\mathbf{S}}_i = \mathbf{S}_i + \mathbf{A}_i \left(\mathbf{C}^{(i)} - \mathbf{R}_i \mathbf{C}^{(\mathrm{f})} \right) \mathbf{S}_\mathrm{a,fine} \left(\mathbf{C}^{(i)} - \mathbf{R}_i \mathbf{C}^{(\mathrm{f})} \right)^T + \mathbf{A}_i \mathbf{C}^{(i)} \mathbf{S}_\mathrm{coin} \mathbf{C}^{(i)T} \tag{22}$$

and from Eqs. (7, 21 and 22) we see that the following equation holds

$$\tilde{\mathbf{S}}_{\mathrm{n}i} = \tilde{\mathbf{S}}_i \mathbf{A}_i^T . \tag{23}$$

We observe that the matrix $\tilde{\mathbf{S}}_i$ is not symmetric and, therefore, does not represent a CM. However, this only concerns the physical meaning of the quantities and does not interfere with the validity of the equations. On the other hand, we can see from Eq. (21) that $\tilde{\mathbf{S}}_{\mathrm{n}i}$ is symmetric and, therefore, equal to its transpose, so that also the following equation holds:

$$\tilde{\mathbf{S}}_{\mathrm{n}i} = \mathbf{A}_i \tilde{\mathbf{S}}_i^T . \tag{24}$$

We substitute Eq. (23) in Eq. (19) and obtain:

$$\mathbf{x}_\mathrm{f} = \left(\sum_{i=1}^{N} \mathbf{R}_i^T \mathbf{A}_i^T \left(\tilde{\mathbf{S}}_i \mathbf{A}_i^T \right)^{-1} \mathbf{A}_i \mathbf{R}_i + \mathbf{S}_\mathrm{a}^{-1} \right)^{-1} \left(\sum_{i=1}^{N} \mathbf{R}_i^T \mathbf{A}_i^T \left(\tilde{\mathbf{S}}_i \mathbf{A}_i^T \right)^{-1} \tilde{\boldsymbol{\alpha}}_i + \mathbf{S}_\mathrm{a}^{-1} \mathbf{x}_\mathrm{a} \right). \tag{25}$$

From Eq. (23) we see that the hypothesis of $\tilde{\mathbf{S}}_{\mathrm{n}i}$ nonsingular implies that also $\mathbf{A}_i$ and $\tilde{\mathbf{S}}_i$ are nonsingular, therefore, from Eq. (25) we obtain the new formula for operational CDF that does no longer contain inverse of matrices that can be singular

$$\mathbf{x}_\mathrm{f} = \left(\sum_{i=1}^{N} \mathbf{R}_i^T \tilde{\mathbf{S}}_i^{-1} \mathbf{A}_i \mathbf{R}_i + \mathbf{S}_\mathrm{a}^{-1} \right)^{-1} \left(\sum_{i=1}^{N} \mathbf{R}_i^T \tilde{\mathbf{S}}_i^{-1} \tilde{\boldsymbol{\alpha}}_i + \mathbf{S}_\mathrm{a}^{-1} \mathbf{x}_\mathrm{a} \right). \tag{26}$$

It is simple to see that in case of absence of interpolation and coincidence errors (that is all the vertical grids coincide and $\mathbf{S}_\mathrm{coin}$ is zero) Eq. (26) becomes Eq. (5). Therefore, Eq. (26), which coincides with the operational CDF of Eq. (19) when $\tilde{\mathbf{S}}_{\mathrm{n}i}$ are nonsingular and coincides with the CDF(2021) in absence of interpolation and coincidence errors, can be used as the new operational CDF rigorously valid also when the noise CMs of the retrieved products are singular matrices.

We can also calculate the AKM and the CMs of the fused profile obtained using Eq. (26). The AKM of $\mathbf{x}_\mathrm{f}$ is given by

$$\mathbf{A}_\mathrm{f} = \frac{\partial \mathbf{x}_\mathrm{f}}{\partial \mathbf{x}} = \left(\sum_{i=1}^{N} \mathbf{R}_i^T \tilde{\mathbf{S}}_i^{-1} \mathbf{A}_i \mathbf{R}_i + \mathbf{S}_\mathrm{a}^{-1} \right)^{-1} \sum_{i=1}^{N} \mathbf{R}_i^T \tilde{\mathbf{S}}_i^{-1} \frac{\partial \tilde{\boldsymbol{\alpha}}_i}{\partial \mathbf{x}} = \left(\sum_{i=1}^{N} \mathbf{R}_i^T \tilde{\mathbf{S}}_i^{-1} \mathbf{A}_i \mathbf{R}_i + \mathbf{S}_\mathrm{a}^{-1} \right)^{-1} \sum_{i=1}^{N} \mathbf{R}_i^T \tilde{\mathbf{S}}_i^{-1} \mathbf{A}_i \mathbf{R}_i , \tag{27}$$

where $\bar{\mathbf{x}}$ is the unknown profile estimated by the data fusion, which for example can be the mean value of the true profiles of the measurements that are fused. The value of the derivative $\frac{\partial \tilde{\boldsymbol{\alpha}}_i}{\partial \mathbf{x}} = \mathbf{A}_i \mathbf{R}_i$ is obtained from Eq. (17) of Ceccherini et al. (2018).

Exploiting the fact that the noise CMs of $\tilde{\boldsymbol{\alpha}}_i$ are $\tilde{\mathbf{S}}_{\mathrm{n}i}$ (Ceccherini et al., 2018), the noise CM of $\mathbf{x}_\mathrm{f}$ is equal to:

$$\mathbf{S}_{\mathrm{n}f} = \left(\sum_{i=1}^{N} \mathbf{R}_i{}^T \tilde{\mathbf{S}}_i{}^{-1} \mathbf{A}_i \mathbf{R}_i + \mathbf{S}_\mathrm{a}{}^{-1} \right)^{-1} \sum_{i=1}^{N} \mathbf{R}_i{}^T \tilde{\mathbf{S}}_i{}^{-1} \tilde{\mathbf{S}}_{\mathrm{n}i} \left(\tilde{\mathbf{S}}_i{}^{-1} \right)^T \mathbf{R}_i \left(\sum_{i=1}^{N} \mathbf{R}_i{}^T \mathbf{A}_i{}^T \left(\tilde{\mathbf{S}}_i{}^{-1} \right)^T \mathbf{R}_i + \mathbf{S}_\mathrm{a}{}^{-1} \right)^{-1} . \qquad (28)$$

In order to simplify this equation, we consider the symmetric matrix given by the product $\tilde{\mathbf{S}}_i{}^{-1} \tilde{\mathbf{S}}_{\mathrm{n}i} \left(\tilde{\mathbf{S}}_i{}^{-1} \right)^T$ and use Eq. (24)

$$\tilde{\mathbf{S}}_i{}^{-1} \tilde{\mathbf{S}}_{\mathrm{n}i} \left(\tilde{\mathbf{S}}_i{}^{-1} \right)^T = \tilde{\mathbf{S}}_i{}^{-1} \mathbf{A}_i \tilde{\mathbf{S}}_i{}^T \left(\tilde{\mathbf{S}}_i{}^{-1} \right)^T = \tilde{\mathbf{S}}_i{}^{-1} \mathbf{A}_i = \mathbf{A}_i{}^T \left(\tilde{\mathbf{S}}_i{}^{-1} \right)^T , \qquad (29)$$

where the last equality is obtained making the transpose and exploiting the fact that the matrix $\tilde{\mathbf{S}}_i{}^{-1} \tilde{\mathbf{S}}_{\mathrm{n}i} \left(\tilde{\mathbf{S}}_i{}^{-1} \right)^T$ is symmetric.

Using Eq. (29) in Eq. (28), the noise CM of $\mathbf{x}_\mathrm{f}$ becomes:

$$\mathbf{S}_{\mathrm{n}f} = \left(\sum_{i=1}^{N} \mathbf{R}_i{}^T \tilde{\mathbf{S}}_i{}^{-1} \mathbf{A}_i \mathbf{R}_i + \mathbf{S}_\mathrm{a}{}^{-1} \right)^{-1} \sum_{i=1}^{N} \mathbf{R}_i{}^T \tilde{\mathbf{S}}_i{}^{-1} \mathbf{A}_i \mathbf{R}_i \left(\sum_{i=1}^{N} \mathbf{R}_i{}^T \tilde{\mathbf{S}}_i{}^{-1} \mathbf{A}_i \mathbf{R}_i + \mathbf{S}_\mathrm{a}{}^{-1} \right)^{-1} . \qquad (30)$$

The smoothing error CM of $\mathbf{x}_\mathrm{f}$ is equal to

$$\mathbf{S}_\mathrm{sf} = \left(\sum_{i=1}^{N} \mathbf{R}_i{}^T \tilde{\mathbf{S}}_i{}^{-1} \mathbf{A}_i \mathbf{R}_i + \mathbf{S}_\mathrm{a}{}^{-1} \right)^{-1} \mathbf{S}_\mathrm{a}{}^{-1} \left(\sum_{i=1}^{N} \mathbf{R}_i{}^T \tilde{\mathbf{S}}_i{}^{-1} \mathbf{A}_i \mathbf{R}_i + \mathbf{S}_\mathrm{a}{}^{-1} \right)^{-1} \qquad (31)$$

and the CM of $\mathbf{x}_\mathrm{f}$, obtained adding to the $\mathbf{S}_\mathrm{nf}$ given in Eq. (30) the smoothing error CM given in Eq. (31), is equal to:

$$\mathbf{S}_\mathrm{f} = \mathbf{S}_\mathrm{nf} + \mathbf{S}_\mathrm{sf} = \left(\sum_{i=1}^{N} \mathbf{R}_i{}^T \tilde{\mathbf{S}}_i{}^{-1} \mathbf{A}_i \mathbf{R}_i + \mathbf{S}_\mathrm{a}{}^{-1} \right)^{-1} . \qquad (32)$$

3 Performance comparison of the original and the new formula using an IASI measurement

In this Section, we show an example of the error that we make using CDF(2015) instead of CDF(2021) on real data, using a METOP-B IASI ozone measurement acquired in the geolocation 43.45° of latitude and 10.77° of longitude, at 8:45:56 UTC of 18 October 2021.

In Fig. 1 we report the retrieved ozone profile of this measurement obtained with the Fast Optimal Retrieval on Layers for IASI (Hurtmans et al., 2012, Astoreca et al., 2014). This product was downloaded from the webpage “IASI Combined Sounding Products – Metop”.

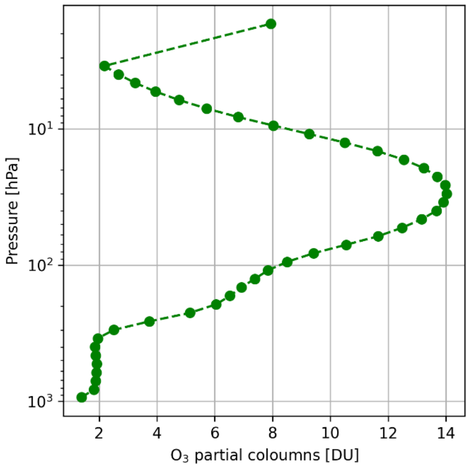

Figure 1: Retrieved ozone profile of the IASI measurement.

In Fig. 2, we report the eigenvalues of $\mathbf{S}_i$ and $\mathbf{S}_{ni}$ for this IASI measurement.

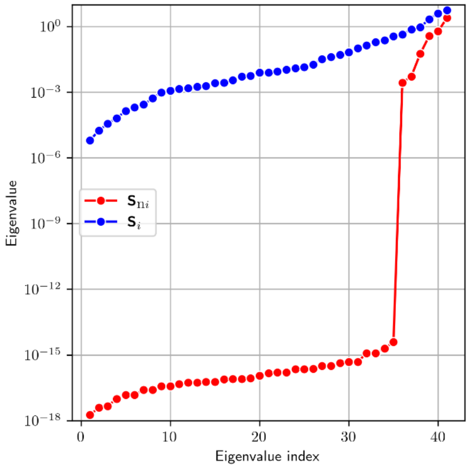

Figure 2: Eigenvalues of the CMs $\mathbf{S}_i$ and $\mathbf{S}_{ni}$ of the IASI measurement.

We can see that the eigenvalues of $\mathbf{S}_i$ are all different from zero, on the other hand, only 6 eigenvalues of $\mathbf{S}_{ni}$ have large values, while the others have values smaller than the numeric noise. The distribution of the eigenvalues of $\mathbf{S}_{ni}$ is due to the fact that the AKM and the retrieval error CM provided to the users are compressed (Astoreca et al., 2017) and are reconstructed using the 6 largest eigenvalues of the Fisher information matrix.

This product is used to perform a consistency check using the two CDF formulas, as described below.

The CDF formula can also be used to estimate, in the linear approximation, how the retrieved profile $\hat{\mathbf{x}}_i$ changes when the a priori profile $\mathbf{x}_{ai}$ and its CM $\mathbf{S}_{ai}$ are changed. This operation, explained in detail in Ceccherini et al. (2014), consists in using the CDF formula with a single input retrieved profile $\hat{\mathbf{x}}_i$, obtained with its a priori profile $\mathbf{x}_{ai}$ and a priori CM $\mathbf{S}_{ai}$, and with the application of a new constraint $\mathbf{x}_{ai}'$ and $\mathbf{S}_{ai}'$. The new profile $\hat{\mathbf{x}}_i'$, that is the original measurement with a new constraint, can be obtained using either CDF(2021) or CDF(2015):

$$\mathbf{x}_{i\ \mathrm{CDF(2021)}}' = \left(\mathbf{S}_i^{-1}\mathbf{A}_i + \mathbf{S}_{\mathrm{ai}}'^{-1}\right)^{-1}\left(\mathbf{S}_i^{-1}\boldsymbol{\alpha}_i + \mathbf{S}_{\mathrm{ai}}'^{-1}\mathbf{x}_{\mathrm{ai}}'\right) \tag{33}$$

$$\mathbf{x}_{i\ \mathrm{CDF(2015)}}' = \left(\mathbf{A}_i^T\mathbf{S}_{\mathrm{n}i}^{\#}\mathbf{A}_i + \mathbf{S}_{\mathrm{ai}}'^{-1}\right)^{-1}\left(\mathbf{A}_i^T\mathbf{S}_{\mathrm{n}i}^{\#}\boldsymbol{\alpha}_i + \mathbf{S}_{\mathrm{ai}}'^{-1}\mathbf{x}_{\mathrm{ai}}'\right), \tag{34}$$

where in the expression derived from CDF(2015) we have used the generalized inverse matrices of $\mathbf{S}_{\mathrm{n}i}$ to deal with the most general case in which $\mathbf{S}_{\mathrm{n}i}$ is singular.

When in Eqs. (33) and (34) we use a new constrain that is equal to the original one: $\mathbf{x}_{\mathrm{ai}}' = \mathbf{x}_{\mathrm{ai}}$ and $\mathbf{S}_{\mathrm{ai}}' = \mathbf{S}_{\mathrm{ai}}$, the formulas should provide the retrieved profile $\hat{\mathbf{x}}_i$. This is a check that we use to validate the self-consistency of the input data and that we can here use to assess the differences between the two CDF formulas.

Substituting $\boldsymbol{\alpha}_i$ from Eq. (6) in Eq. (33) and using Eqs. (2) and (16), we obtain that actually

$$\mathbf{x}_{i\ \mathrm{CDF(2021)}}'\left[\mathbf{x}_{\mathrm{ai}}' = \mathbf{x}_{ai},\ \ \mathbf{S}_{\mathrm{ai}}' = \mathbf{S}_{ai}\right] = \hat{\mathbf{x}}_i\,. \tag{35}$$

On the other hand, substituting $\boldsymbol{\alpha}_i$ from Eq. (6) in Eq. (34) we obtain:

$$\mathbf{x}_{i\ \mathrm{CDF(2015)}}'\left[\mathbf{x}_{\mathrm{ai}}' = \mathbf{x}_{ai},\ \ \mathbf{S}_{\mathrm{ai}}' = \mathbf{S}_{ai}\right] = \hat{\mathbf{x}}_i + \left[\left(\mathbf{A}_i^T\mathbf{S}_{\mathrm{n}i}^{\#}\mathbf{A}_i + \mathbf{S}_{\mathrm{ai}}^{-1}\right)^{-1}\mathbf{A}_i^T\mathbf{S}_{\mathrm{n}i}^{\#} - \mathbf{I}\right]\left(\hat{\mathbf{x}}_i - \mathbf{x}_{\mathrm{ai}}\right), \tag{36}$$

where **I** is the identity matrix. The second term of Eq. (36) measures the error made using the generalized inverse and, using Eqs.(1, 3) and (16), we see that, in the case that $\mathbf{S}_{ni}$ is nonsingular, is equal to zero.

We have calculated the difference $\mathbf{x}_{i\ \mathrm{CDF(2015)}}'\left[\mathbf{x}_{\mathrm{ai}}' = \mathbf{x}_{ai},\ \ \mathbf{S}_{\mathrm{ai}}' = \mathbf{S}_{ai}\right] - \hat{\mathbf{x}}_i$ for several values of the threshold used to determine the eigenvalues that are neglected in the calculation of the generalized inverse matrix of $\mathbf{S}_{\mathrm{n}i}$. In Fig. 3 we report the consistency test provided by this difference in the case of three values of the threshold that correspond to selecting, respectively, the 5, 6 and 7 largest eigenvalues.

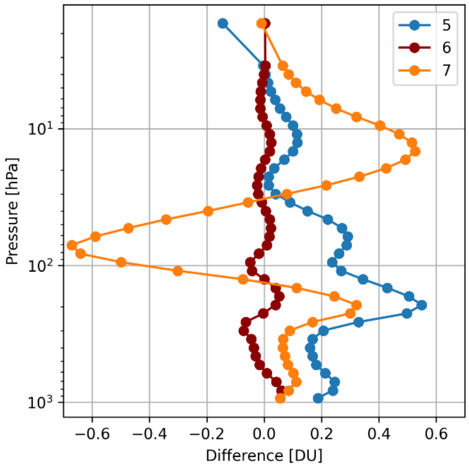

Figure 3: Results of the consistency test with CDF(2015) considering only the 5, 6 and 7 largest eigenvalues in the calculation of the generalized inverse matrix of $\mathbf{S}_{\mathrm{n}i}$.

We can see that the smallest differences are obtained for the case of 6 eigenvalues, as expected from the distribution of the eigenvalues. The case of 5 eigenvalues is affected by the loss of useful information, on the other hand the case of 7 eigenvalues is affected by the amplification of the numeric noise. In this case, the choice of the threshold value can be simply done looking at Fig. 2, where the abrupt variation of the eigenvalues values clearly indicates the threshold. In a general case, in which the variation of the eigenvalues is smooth, this test can be used to define the threshold for the eigenvalues choosing the value that minimizes the difference $\mathbf{x}'_{i\,\mathrm{CDF(2015)}}\left[\mathbf{x}_{ai}'=\mathbf{x}_{ai},\ \mathbf{S}_{ai}'=\mathbf{S}_{ai}\right]-\hat{\mathbf{x}}_i$.

Using the optimum number of eigenvalues for CDF(2015), in Fig. 4 we compare the differences $\mathbf{x}'_{i\,\mathrm{CDF(2021)}}\left[\mathbf{x}_{\mathrm{ai}}'=\mathbf{x}_{\mathrm{ai}},\ \mathbf{S}_{\mathrm{ai}}'=\mathbf{S}_{\mathrm{ai}}\right]-\hat{\mathbf{x}}_i$ and $\mathbf{x}'_{i\,\mathrm{CDF(2015)}}\left[\mathbf{x}_{\mathrm{ai}}'=\mathbf{x}_{\mathrm{ai}},\ \mathbf{S}_{\mathrm{ai}}'=\mathbf{S}_{\mathrm{ai}}\right]-\hat{\mathbf{x}}_i$ of the consistency test for the two CDF formulas with the retrieval error of the profile estimated by the square root of the diagonal elements of the CM $\mathbf{S}_i$.

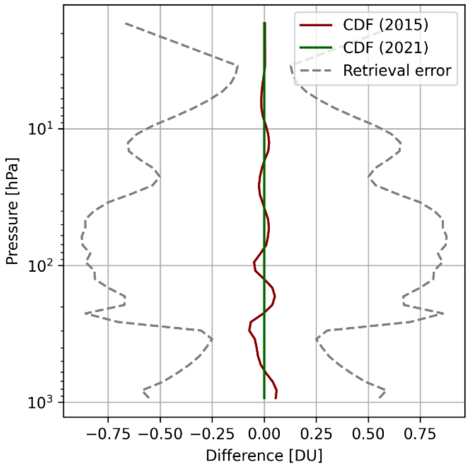

Figure 4: Results of the consistency test applied to the IASI measurement for the two formulas CDF(2015) and CDF(2021) compared with the retrieval error of the profile.

As expected the consistency test provides zero differences using CDF(2021) and detectable differences, although much smaller than the retrieval errors, are present when using CDF(2015). These differences are an estimate of the errors introduced by CDF(2015) in the fusion process with respect to the results of CDF(2021).

The errors introduced by CDF(2015) depend on the compression used to represent the matrices in the files provided to the users. If less compression was applied to the data a greater number of eigenvalues could be considered in the calculation of the generalized inverse matrix of $\mathbf{S}_{ni}$ and the errors introduced by CDF(2015) would be further reduced.

When no compression is applied, the errors introduced by CDF(2015) are due to the numerical precision with which the data are provided, because the eigenvalues smaller than the numerical precision of the largest eigenvalue will usually only contribute to the noise of the generalized inverse. Therefore, less compression and improved numerical precision can reduce the approximation introduced by CDF(2015).

4 Conclusions

The original formula CDF(2015) of the CDF requires the calculation of the inverse matrices of the noise CMs $\mathbf{S}_{\mathrm{n}i}$ of the input profiles and, therefore, can be rigorously applied only when these CMs are nonsingular. In the other cases, the CDF(2015) can still be used replacing the inverse matrices of the noise CMs with the generalized inverse matrices, but the result is an approximation. Furthermore, a variable exists in this operation and a threshold has to be identified for the choice of how many eigenvalues are used in the calculation of the generalized inverse matrices.

A new formula CDF(2021) has been presented that contains the inverse matrices of the retrieval error CMs (the CMs that include both the noise and the smoothing errors), instead of the inverse matrices of the noise CMs. Since the retrieval error CMs are always nonsingular matrices, the new formula can be used without resorting to generalized inverse matrices.

We deduced the analytical relationship between the two formulas and observed that the quality of the approximation provided by the old formula depends on how much $\mathbf{S}_{ni}\mathbf{S}_{ni}^{\#}$ is close to the identity matrix.

Furthermore, we have obtained the expression of the operational CDF(2021), which can handle interpolation and coincidence errors. The operational CDF(2021) is indispensable for the application of the CDF to real measurements, which are often measured on different vertical grids and at either different times or locations.

Finally, we have introduced a consistency check that can be used to define the threshold for the eigenvalues of the noise CMs and applied it to a real IASI measurement to evaluate the errors made using CDF(2015) instead of CDF(2021). We observed that in practice the errors introduced by the use of CDF(2015) are much smaller than the retrieval errors and depend on the data compression and numerical precision with which the data are provided to the users.

The use of the new CDF(2021) and operational CDF(2021) is recommended for data fusion processing, but the errors made with the old CDF(2015) do not appear to be important, even in the case of a significant data compression.

Data availability. The IASI data used in the paper are available at the webpage "IASI Combined Sounding Products – Metop": https://navigator.eumetsat.int/product/EO:EUM:DAT:METOP:IASSND02, last access 27 April 2022. The results of the analysis performed on these data are available from the authors upon request.

Author contributions. SC derived the new formula of the CDF and extended it to the case in which coincidence and interpolation errors are present. He wrote the draft version of the paper. NZ contributed to the method extension, implemented the formulas in a Python code and performed the test on the IASI measurement. BC contributed to the interpretation of the results and made heavy revisions giving a coherent structure to the paper. All the authors revised the paper.

Competing interests. The authors declare that they have no conflict of interest.

References

Astoreca, R., Coheur, P., Hurtmans, D., Hadji-Lazaro, J., George, M., Clerbaux, C.: Product User Manual Near real-time IASI CO, The EUMETSAT Network of Satellite Application Facilities, pp. 28, https://www.eumetsat.int/media/41287, last access 9 May 2022, 2017

Astoreca, R., Hurtmans, D., Coheur, P.-F.: Fast Optimal Retrieval on Layers for IASI, Algorithm Theoretical Basis Document, The EUMETSAT Network of Satellite Application Facilities, pp. 14, https://acsaf.org/docs/atbd/Algorithm_Theoretical_Basis_Document_IASI_CO_Feb_2014.pdf, last access: 27 April 2022, 2014

Ceccherini, S.: Comment on "Synergetic use of IASI and TROPOMI space borne sensors for generating a tropospheric methane profile product", Atmos. Meas. Tech. Discuss. [preprint], https://doi.org/10.5194/amt-2021-98, accepted for publication in Atmos. Meas. Tech., 2021.

Ceccherini, S., Carli, B., Raspollini, P.: Quality quantifier of indirect measurements, Opt. Express, 20, 5151-5167, 2012.

Ceccherini, S., Carli, B., Raspollini, P.: The average of atmospheric vertical profiles, Opt. Express, 22, 24808-14816, 2014.

Ceccherini, S., Carli, B., and Raspollini, P.: Equivalence of data fusion and simultaneous retrieval, Opt. Express, 23, 8476-8488, 2015.

https://doi.org/10.5194/amt-2022-182
Preprint. Discussion started: 29 August 2022

Ceccherini, S., Carli, B., Tirelli, C., Zoppetti, N., Del Bianco, S., Cortesi, U., Kujanpää, J., and Dragani, R.: Importance of interpolation and coincidence errors in data fusion, Atmos. Meas. Tech., 11, 1009-1017, https://doi.org/10.5194/amt-11-1009-2018, 2018.

Clerbaux, C., Boynard, A., Clarisse, L., George, M., Hadji-Lazaro, J., Herbin, H., Hurtmans, D., Pommier, M., Razavi, A., Turquety, S., Wespes, C., and Coheur, P.-F.: Monitoring of atmospheric composition using the thermal infrared IASI/MetOp sounder, Atmos. Chem. Phys., 9, 6041–6054, doi:10.5194/acp-9-6041-2009, 2009.

Fisher, R. A.: The logic of inductive inference, J. Roy. Stat. Soc., 98, 39-54, 1935.

Hurtmans, D., Coheur, P., Wespes, C., Clarisse, L., Scharf , O., Clerbaux, C., Hadji -Lazaro, J.,George, M. & Turquety, S.: FORLI radiative transfer and retrieval code for IASI. J. Quant. Spectrosc. Radiat. Transfer, 113, 1391-1408, https://doi.org/10.1016/j.jqsrt.2012.02.036, 2012

IASI Combined Sounding Products – Metop: https://navigator.eumetsat.int/product/EO:EUM:DAT:METOP:IASSND02, last access 27 April 2022, 2010.

Kalman, R. E.: Algebraic aspects of the generalized inverse of a rectangular matrix, Proceedings of Advanced Seminar on Generalized Inverse and Applications, M. Z. Nashed, Academic, San Diego, Calif., 111-124, 1976.

Rodgers, C.D., Inverse Methods for Atmospheric Sounding: Theory and Practice, Vol. 2 of Series on Atmospheric, Oceanic and Planetary Physics, World Scientific, Singapore, 2000.

Tirelli, C., Ceccherini, S., Zoppetti, N., Del Bianco, S., Gai, M., Barbara, F., Cortesi, U., Kujanpää, J., Huan, Y., and Dragani, R.: Data fusion analysis of Sentinel-4 and Sentinel-5 simulated ozone data, J. Atmos. Ocean. Tech., 37, 573–587, https://doi.org/10.1175/JTECH-D-19-0063.1, 2020.

Zoppetti, N., Ceccherini, S., Carli, B., Del Bianco, S., Gai, M., Tirelli, C., Barbara, F., Dragani, R., Arola, A., Kujanpää, J., van Peet, J. C. A., van der A, R., and Cortesi, U.: Application of the Complete Data Fusion algorithm to the ozone profiles measured by geostationary and low-Earth-orbit satellites: a feasibility study, Atmos. Meas. Tech., 14, 2041–2053, https://doi.org/10.5194/amt-14-2041-2021, 2021