

# Optimal use of the information provided by indirect measurements of atmospheric vertical profiles

Simone Ceccherini\*, Piera Raspollini and Bruno Carli

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

\*Corresponding author: [S.Ceccherini@ifac.cnr.it](mailto:S.Ceccherini@ifac.cnr.it)

**Abstract:** A procedure for the optimal utilization and representation of the information retrieved from atmospheric observations is discussed. We show that the "measurement-space solution" exclusively contains the information present in the observations, but is not suitable for a representation of the retrieved profile in the form of a graph. The new method of the "null-space regularization", which provides in a complement space an external constraint and makes the solution compliant with this representation, is presented. In this way the measurement and the constraint are separately determined and are given in a vertical grid that can be freely chosen as fine as desired. The method is applied to simulated measurements and is used to define a procedure for data fusion.

©2009 Optical Society of America

**OCIS codes:** (100.3190) Inverse problems; (120.0280) Remote sensing and sensors; (010.1280) Atmospheric composition; (000.3860) Mathematical methods in physics.

---

## References and links

1. C. D. Rodgers, *Inverse Methods for Atmospheric Sounding: Theory and Practice*, Vol. 2 of Series on Atmospheric, Oceanic and Planetary Physics (World Scientific, 2000).
2. S. Twomey, *Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurements* (Elsevier, New York, 1977).
3. B. Carli, P. Raspollini, M. Ridolfi, and B. M. Dinelli, "Discrete representation and resampling in limb-sounding measurements," *Appl. Opt.* **40**, 1261-1268 (2001).
4. A. Tikhonov, "On the solution of incorrectly stated problems and method of regularization," *Dokl. Akad. Nauk. SSSR* **151**, 501-504 (1963).
5. S. Twomey, "On the numerical solution of Fredholm integral equations of the first kind by the inversion of the linear system produced by quadrature," *J. Assoc. Comput. Mach.* **10**, 97-101 (1963).
6. B. Schimpf and F. Schreier, "Robust and efficient inversion of vertical sounding atmospheric high-resolution spectra by means of regularization," *J. Geophys. Res.* **102**, 16037-16055 (1997).
7. T. Steck, "Methods for determining regularization for atmospheric retrieval problems," *Appl. Opt.* **41**, 1788-1797 (2002).
8. S. Ceccherini, "Analytical determination of the regularization parameter in the retrieval of atmospheric vertical profiles," *Opt. Lett.* **30**, 2554-2556 (2005).
9. M. Ridolfi and L. Sgheri, "A self-adapting and altitude-dependent regularization method for atmospheric profile retrievals," *Atmos. Chem. Phys. Discuss* **8**, 18007-18037 (2008).
10. C. D. Rodgers and B. J. Connor, "Intercomparison of remote sounding instruments," *J. Geophys. Res.* **108**, 4116-4130 (2003).
11. S. Ceccherini, B. Carli, E. Pascale, M. Prosperi, P. Raspollini, and B. M. Dinelli, "Comparison of measurements made with two different instruments of the same atmospheric vertical profile," *Appl. Opt.* **42**, 6465-6473 (2003).
12. J. Joiner and A. M. da Silva, "Efficient methods to assimilate remotely sensed data based on information content," *Q. J. R. Meteorol. Soc.* **124**, 1669-1694 (1998).
13. T. von Clarmann and U. Grabowski, "Elimination of hidden a priori information from remotely sensed profile data," *Atmos. Chem. Phys.* **7**, 397-408 (2007).
14. H. Fischer, M. Birk, C. Blom, B. Carli, M. Carlotti, T. von Clarmann, L. Delbouille, A. Dudhia, D. Ehalt, M. Endemann, J. M. Flaud, R. Gessner, A. Kleinert, R. Koopman, J. Langen, M. López-Puertas, P. Mosner, H. Nett, H. Oelhaf, G. Perron, J. Remedios, M. Ridolfi, G. Stiller and R. Zander, "MIPAS: an instrument for atmospheric and climate research," *Atmos. Chem. Phys.* **8**, 2151-2188 (2008).

15. M. Ridolfi, B. Carli, M. Carlotti, T. v. Clarmann, B. M. Dinelli, A. Dudhia, J.-M. Flaud, M. Höpfner, P. E. Morris, P. Raspollini, G. Stiller, and R. J. Wells, "Optimized forward model and retrieval scheme for MIPAS near-real-time data processing," *Appl. Opt.* **39**, 1323-1340 (2000).
  16. P. Raspollini, C. Belotti, A. Burgess, B. Carli, M. Carlotti, S. Ceccherini, B. M. Dinelli, A. Dudhia, J. M. Flaud, B. Funke, M. Höpfner, M. López-Puertas, V. Payne, C. Piccolo, J. J. Remedios, M. Ridolfi and R. Spang, "MIPAS level 2 operational analysis," *Atmos. Chem. Phys.* **6**, 5605-5630 (2006).
  17. A. Dudhia, V. L. Jay, and C. D. Rodgers, "Microwindow selection for high-spectral-resolution sounders," *Appl. Opt.* **41**, 3665-3673 (2002).
- 

## 1. Introduction

Several techniques exist for the retrieval of atmospheric parameters from remote sensing observations [1, 2]. A typical retrieved quantity is the vertical distribution of an atmospheric constituent. Since the observations do not provide enough information to retrieve the continuous function of the vertical profile, the problem arises of how to represent, in terms of which and how many parameters, the finite number of independent pieces of information provided by the observations. The first choice is that of the vertical grid on which to represent the profile. This vertical grid should be as fine as possible in order both to make sure that all the information included in the observations is adequately represented by the profile and to limit the need for interpolation in the subsequent applications of the data. Indeed, operations of interpolation do introduce a loss of information [3]. However, whenever the information obtained from the observations is not sufficient to determine as many parameters as the grid points, a too fine grid makes the inversion problem ill-posed or ill-conditioned. This problem is generally overcome using some external information. In the optimal estimation method [1], according to the Bayesian approach, an a-priori probability distribution function of the atmospheric state is used to fill the gaps present in the observations. In this case the solution is represented by the state corresponding to the maximum of the a-posteriori probability distribution function, which combines the a-priori information with that of the observations. An alternative method consists in applying a smoothness constraint to the profile. This can be done adding to the chi-square function minimized by the retrieval a term that increases when the profile oscillates [4-9]. In this case the information not coming from the observations is in the smoothness constraint. A further method that allows to make the problem well conditioned consists in writing the profile as a linear combination of a finite number of continuous functions (generally triangular functions, boxcars, polynomials, sine and cosine functions etc...) and to determine the coefficients of these functions fitting the observations [1]. In this case the information not coming from the observations consists in the number and in the kind of the chosen functions. In the following discussion the external information not coming from the observations is referred to as "a-priori information" in all cases, even when it is applied in the form of a constraint.

A-priori information makes it possible to obtain a vertical profile on a fine grid, but generates other types of problems in utilizing the data. In validation activities, when the compared profiles use different a-priori information, it is necessary to assess the error due to this difference [10]. The estimation of the error component due to the use of different a-priori information is a difficult task because it requires a reliable knowledge of the climatological variability of the quantity to be compared, which often is not available. Also in data assimilation, which is the combining of diverse data sampled at different times and different locations into atmospheric models, the measurements containing a-priori information pose some problems [1]. Profiles in different geolocations obtained using the same a-priori information are correlated among themselves as well as profiles obtained using nearby retrieved profiles as a-priori. If not taken into account, these correlations can produce a bias in the products of the assimilation. Also in the case of data fusion, which consists in the synergetic use of more than one measurement of the same atmospheric state obtained with different instruments, the presence of a-priori information in the original measurements can produce a bias in the final product.

In order to overcome the problems posed by the a-priori information included in the retrieved profiles some approaches have been proposed. S. Ceccherini et al. [11] described a method to perform comparison of measurements that uses the intersection space of the spaces of the averaging kernels of the measurements to be compared. This approach does not involve degradation of the available information and properly accounts for the a-priori information included in the profiles to be compared. Joiner et al. [12] proposed two methods that allow to reduce or eliminate the a-priori information from the retrieved quantities: the null-space filtering of the retrievals and the partial eigen-decomposition retrieval. Both methods involve the singular value decomposition (SVD) of appropriate retrieval operators and provide the coefficients of the singular vectors as quantities to be assimilated. Clarmann et al. [13] proposed a method that eliminates the a-priori information by means of a resampling on a coarser vertical grid providing a profile with averaging kernel matrix equal to the unity matrix.

In this paper a procedure is discussed for the optimal exploitation and representation of the vertical profile information retrieved from atmospheric observations. With optimal exploitation it is here intended the use of all the information present in the observations without addition of external information and, therefore, implies the use of a retrieval grid as fine as needed without any a-priori information. This objective has already been attained by Joiner et al. [12], but they only obtain an expression suitable for data assimilation which, as we shall see in Section 3, is not suitable for a representation of the retrieved profile in the form of a graph. The combination of an optimal exploitation with an optimal representation of the profile is the first objective of this paper. The second objective is to extend the optimal exploitation and representation procedure to the case of data fusion.

In Section 2 the theory of the proposed new method for the representation of the retrieved profile is described and in Section 3 it is applied to a simulated case. Section 4 describes the proposed procedure for data fusion problem.

## 2. Theory

### 2.1 Terminology and notation

Consistently with the notations adopted by Rodgers [1], we represent the observations (radiance) with a vector  $\mathbf{y}$  of  $m$  elements and the vertical profile of the unknown atmospheric parameter with a vector  $\mathbf{x}$  of  $n$  elements corresponding to a predefined altitude grid. The forward model is a function  $\mathbf{F}(\mathbf{x})$  that provides the value of the observations when the profile  $\mathbf{x}$  is known. Therefore, the relationship between the vectors  $\mathbf{x}$  and  $\mathbf{y}$  is

$$\mathbf{y} = \mathbf{F}(\mathbf{x}) + \boldsymbol{\varepsilon}, \quad (1)$$

where  $\boldsymbol{\varepsilon}$  is the vector containing the experimental errors of the observations, characterized by a variance-covariance matrix (VCM)  $\mathbf{S}_y$  given by the mean value of the product of  $\boldsymbol{\varepsilon}$  times its transposed.

We shall consider the case of a nearly linear relationship between  $\mathbf{y}$  and  $\mathbf{x}$ . In this case  $\mathbf{F}(\mathbf{x})$  can be expanded up to the first order around a specific value of  $\mathbf{x}$ , identified by  $\mathbf{x}_0$  and referred to as the *linearization point*, and Eq. (1) becomes equal to:

$$\mathbf{y} = \mathbf{F}(\mathbf{x}_0) + \mathbf{K}(\mathbf{x} - \mathbf{x}_0) + \boldsymbol{\varepsilon}, \quad (2)$$

where  $\mathbf{K}$  is the Jacobian matrix (that is the partial derivatives of  $\mathbf{F}(\mathbf{x})$  with respect to the elements of  $\mathbf{x}$ ) calculated at  $\mathbf{x}_0$ .

Equation (2) implies that the components of  $\mathbf{y} - \mathbf{F}(\mathbf{x}_0)$  are the scalar products between  $\mathbf{x} - \mathbf{x}_0$  and the rows of  $\mathbf{K}$  plus the error components. It follows that the knowledge of  $\mathbf{y} - \mathbf{F}(\mathbf{x}_0)$  determines the knowledge of the component of  $\mathbf{x} - \mathbf{x}_0$  that lies in the space generated by the rows of  $\mathbf{K}$ . In agreement with [11], and with the conventions that  $\mathbf{y}$  are the observations and  $\mathbf{x}$

are the aimed measurements, we shall refer to this space as the *measurement space*. Notice that this is not the terminology adopted in [1] where this space is called *row space* and the expression *measurement space* is instead used to indicate the space of  $m$  dimensions to which the observations  $\mathbf{y}$  belong.

With the inversion of Eq. (2) one obtains the correction  $\mathbf{x}-\mathbf{x}_0$  that has to be applied to  $\mathbf{x}_0$  in order to determine  $\mathbf{x}$ . Since the correction is made in the measurement space and no information is acquired about the components of  $\mathbf{x}-\mathbf{x}_0$  in the orthogonal space that is complementary to the measurement space, also the determination of  $\mathbf{x}$  is to be considered as only made in the measurement space. However, this statement requires some further explanation. In fact, if  $\mathbf{x}_0$  has a component in the orthogonal space, when  $\mathbf{x}_0$  is combined with  $\mathbf{x}-\mathbf{x}_0$  to give  $\mathbf{x}$  also  $\mathbf{x}$  acquires this component. However such a component cannot be considered to be a result of the retrieval process because it is a contingent quantity that depends on the procedure rather than on the observations. On the basis of these considerations it is correct to say that since  $\mathbf{x}-\mathbf{x}_0$  is determined in the measurement space, also  $\mathbf{x}$  is determined in this space. The orthogonal complement space to the measurement space is referred to as the *null space* [1].

In order to weigh the observations with their errors and to avoid the complication of correlated errors it is useful to consider the quantities  $\mathbf{S}_y^{-1/2}\mathbf{y}$  instead of the observations  $\mathbf{y}$ .

Multiplying both terms of Eq. (2) on the left by  $\mathbf{S}_y^{-1/2}$  we obtain:

$$\mathbf{S}_y^{-1/2}\mathbf{y} = \mathbf{S}_y^{-1/2}\mathbf{F}(\mathbf{x}_0) + \mathbf{S}_y^{-1/2}\mathbf{K}(\mathbf{x} - \mathbf{x}_0) + \mathbf{S}_y^{-1/2}\boldsymbol{\varepsilon}. \quad (3)$$

It is easy to verify that  $\mathbf{S}_y^{-1/2}\mathbf{y}$  is characterized by a VCM that is the unity matrix.

## 2.2 Representation of the profile

The unknown profile  $\mathbf{x}$  is a vector of  $n$  elements and belongs, therefore, to the space  $\mathbb{R}^n$  which can be decomposed in the direct sum of the measurement space and of the null space. If we indicate with  $p$  the dimension of the measurement space then the dimension of the null space is  $n-p$ . Since each vector of  $\mathbb{R}^n$  can be decomposed as the sum of a vector belonging to the measurement space and a vector belonging to the null space, we can write:

$$\mathbf{x} = \mathbf{x}_a + \mathbf{x}_b, \quad (4)$$

where  $\mathbf{x}_a$  and  $\mathbf{x}_b$ , respectively, belong to the measurement space and to the null space. They can be expressed as:

$$\mathbf{x}_a = \mathbf{V}\mathbf{a}, \quad (5)$$

$$\mathbf{x}_b = \mathbf{W}\mathbf{b}, \quad (6)$$

where  $\mathbf{V}$  is a matrix whose columns are an orthonormal basis of the measurement space,  $\mathbf{W}$  is a matrix whose columns are an orthonormal basis of the null space and  $\mathbf{a}$  and  $\mathbf{b}$  are the projections of  $\mathbf{x}$  on these orthonormal bases:

$$\mathbf{a} = \mathbf{V}^T\mathbf{x}, \quad (7)$$

$$\mathbf{b} = \mathbf{W}^T\mathbf{x}, \quad (8)$$

where the superscript  $T$  denotes transposed matrices.

### 2.3 Estimation of the components in the measurement space

The component of  $\mathbf{x}$  in the measurement space is the only quantity that can be derived from the observations. In order to find  $\mathbf{x}_a$  from Eq. (5) we need to identify  $\mathbf{V}$  and  $\mathbf{a}$ . To this purpose we perform the SVD of  $\mathbf{S}_y^{-1/2} \mathbf{K}$ :

$$\mathbf{S}_y^{-1/2} \mathbf{K} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T, \quad (9)$$

where  $\mathbf{U}$  is a matrix of dimension  $m \times p$  whose columns (referred to as left singular vectors) are an orthonormal basis of the space generated by the columns of  $\mathbf{S}_y^{-1/2} \mathbf{K}$ ,  $\mathbf{\Lambda}$  is a nonsingular diagonal matrix of dimension  $p \times p$  and  $\mathbf{V}$  is a matrix of dimension  $n \times p$  whose columns (referred to as right singular vectors) are an orthonormal basis of the space generated by the rows of  $\mathbf{S}_y^{-1/2} \mathbf{K}$ . Since  $\mathbf{S}_y^{-1/2}$  is a nonsingular matrix the space generated by the rows of  $\mathbf{S}_y^{-1/2} \mathbf{K}$  coincides with the space generated by the rows of  $\mathbf{K}$ , therefore the columns of  $\mathbf{V}$  are an orthonormal basis of the measurement space and, among all the possible orthonormal bases of the measurement space, it can be chosen for representing  $\mathbf{x}_a$  with Eq. (5). We can now determine the components of  $\mathbf{x}_a$  relative to this orthonormal basis.

Substituting Eq. (9) in Eq. (3), multiplying both terms on the left by  $\mathbf{\Lambda}^{-1} \mathbf{U}^T$  and using Eq. (7), after some rearrangements, we obtain that  $\hat{\mathbf{a}}$ , i.e. the estimation of  $\mathbf{a}$  deduced from the observations, is given by:

$$\hat{\mathbf{a}} = \mathbf{a} + \boldsymbol{\varepsilon}_a = \mathbf{V}^T \mathbf{x} + \boldsymbol{\varepsilon}_a = \mathbf{V}^T \mathbf{x}_0 + \mathbf{\Lambda}^{-1} \mathbf{U}^T \mathbf{S}_y^{-1/2} (\mathbf{y} - \mathbf{F}(\mathbf{x}_0)), \quad (10)$$

where

$$\boldsymbol{\varepsilon}_a = \mathbf{\Lambda}^{-1} \mathbf{U}^T \mathbf{S}_y^{-1/2} \boldsymbol{\varepsilon} \quad (11)$$

is the error that we make taking  $\hat{\mathbf{a}}$  as the estimation of  $\mathbf{a}$  and is characterized by the diagonal VCM:

$$\mathbf{S}_a = \mathbf{\Lambda}^{-2}. \quad (12)$$

From Eq. (12) we see that the components of the vector  $\hat{\mathbf{a}}$  are independent of each other and are characterized by variances given by the inverse of the squared singular values of  $\mathbf{S}_y^{-1/2} \mathbf{K}$ . Therefore, components corresponding to large singular values are well determined while components corresponding to small singular values are poorly determined.

The definition of  $\mathbf{x}_a$  in terms of  $\mathbf{V}$  and  $\mathbf{a}$ , determined respectively by Eq. (9) and by Eq. (10), will be referred to as the *measurement-space solution* (MSS). This solution is obtained exploiting all the information coming from the observations without any a-priori information. It is, therefore, the optimal quantity to be used for further post-retrieval processing. This conclusion is consistent with the results obtained by [12] where a solution equal to that of Eq. (10) is used for data assimilation.

The MSS, while optimal for further processing because comprehensive, unbiased and with a diagonal VCM, is not suitable for a representation of the retrieved profile in the form of a graph (i.e. for a graphical representation). Indeed, if the dimension  $n$  is chosen with the desirable redundancy, components of the MSS that correspond to small singular values are poorly determined and make this solution ill-conditioned. On the other hand, if the poorly determined components are removed, the size of the null space grows at the expenses of the

measurement space and, as it will be shown in Section 3, the retrieved profile acquires a rather unphysical shape.

As discussed in the introduction, these difficulties have forced alternative solutions, such as the use of a coarse and instrument dependent altitude grid and the use of constraints in the retrieval (regularization [4-9] and optimal estimation [1]).

#### 2.4 Estimation of the components in the null space

In general the information provided by the observations is not sufficient for the selection of a single profile, indeed it selects a set of possible profiles given by Eq. (4) where  $\mathbf{x}_a$  is determined by the observations and  $\mathbf{x}_b$  is given by Eq. (6) with an arbitrary value of the components  $\mathbf{b}$ . The problem is here how to choose these components in the null space. When further processing of the data is planned it is adequate to provide the measurement information in a sub-space and taking advantage of this opportunity it is better to avoid any choice about  $\mathbf{x}_b$ . Indeed, the further processing may provide some estimation about the null-space components and any constraint (even the condition  $\mathbf{b}=0$ ) does introduce a bias. However, if a graphical representation of the profile is requested different considerations apply. In a graphical representation the result is to be presented in a complete space. The representation of the  $\mathbf{x}$  profile with the MSS corresponds to selecting the value zero for  $\mathbf{b}$ . This is a particular choice among the infinitive possible ones, which does not necessarily provide the best graphical representation, as it will be shown in Section 3.

Since we do not expect that real profiles show unphysical oscillations, a good graphical representation can be obtained selecting the smoothest profile compatible with the observations. This can be done choosing a well conditioned MSS (obtained selecting only the well measured components of  $\mathbf{a}$ ) and determining in the null space (that now includes the null space enlarged with the poorly measured components rejected from the measurement space) the value of  $\mathbf{b}$  that minimizes the oscillations. To this purpose, we minimize the quantity:

$$\|\mathbf{L}_1 \mathbf{x}\|^2 = \mathbf{x}^T \mathbf{L}_1^T \mathbf{L}_1 \mathbf{x} = (\mathbf{a}^T \mathbf{V}^T + \mathbf{b}^T \mathbf{W}^T) \mathbf{L}_1^T \mathbf{L}_1 (\mathbf{V} \mathbf{a} + \mathbf{W} \mathbf{b}) \quad (13)$$

as a function of  $\mathbf{b}$ , where  $\mathbf{L}_1$  is the first-derivative matrix [7] with respect to altitude and Eqs. (4)-(6) have been used for the definition of  $\mathbf{x}$ . In these equations the matrices  $\mathbf{V}$  and  $\mathbf{W}$  do not necessarily coincide with the matrices used in Sections 2.2 and 2.3; they are now, respectively, the matrices containing the bases of the well measured subspace of the measurement space and of the new complementary null-space. Setting equal to zero the derivatives of the quantity in Eq. (13) with respect the components of  $\mathbf{b}$  and substituting for  $\mathbf{a}$  our estimation  $\hat{\mathbf{a}}$  it results that the estimation of  $\mathbf{b}$  is given by:

$$\hat{\mathbf{b}} = -(\mathbf{W}^T \mathbf{R} \mathbf{W})^{-1} \mathbf{W}^T \mathbf{R} \mathbf{V} \hat{\mathbf{a}}, \quad (14)$$

where  $\mathbf{R} = \mathbf{L}_1^T \mathbf{L}_1$ . This calculation of the components of  $\mathbf{x}$  in the new null space allows obtaining the smoothest profile compatible with the observations and corresponds to a regularization of the profile; accordingly, it will be referred to as the *null-space regularization* (NSR).

An important feature that makes this operation different from the other constraints that are applied to the retrieved profile is that in the case of the NSR the constraint is only applied to the null-space component and does not affect the measured component.

#### 2.5 Solution and its characterization

An estimation  $\hat{\mathbf{x}}$  of the profile  $\mathbf{x}$  that makes use of both the MSS and the NSR is obtained from Eqs. (4)-(6) and (14) and is equal to:

$$\hat{\mathbf{x}} = \hat{\mathbf{x}}_a + \hat{\mathbf{x}}_b = \mathbf{V}\hat{\mathbf{a}} + \mathbf{W}\hat{\mathbf{b}} = \left[ \mathbf{I} - \mathbf{W}(\mathbf{W}^T \mathbf{R} \mathbf{W})^{-1} \mathbf{W}^T \mathbf{R} \right] \mathbf{V}\hat{\mathbf{a}}, \quad (15)$$

where  $\hat{\mathbf{a}}$  is given by Eq. (10). As it will be shown in Section 3, the solution  $\hat{\mathbf{x}}$  given by Eq. (15) provides a useful graphical representation of the retrieved profile and it will be referred to as *regularized measurement-space solution* (RMSS). The equations that provide the RMSS can be used to determine its VCM and its averaging kernel matrix [1].

From Eqs. (12) and (15) we obtain that the VCM of  $\hat{\mathbf{x}}$  is equal to:

$$\mathbf{S}_x = \left[ \mathbf{I} - \mathbf{W}(\mathbf{W}^T \mathbf{R} \mathbf{W})^{-1} \mathbf{W}^T \mathbf{R} \right] \mathbf{V} \Lambda^{-2} \mathbf{V}^T \left[ \mathbf{I} - \mathbf{W}(\mathbf{W}^T \mathbf{R} \mathbf{W})^{-1} \mathbf{W}^T \mathbf{R} \right]^T. \quad (16)$$

In order to calculate the averaging kernel matrix we have to differentiate  $\hat{\mathbf{x}}$ , given by Eq. (15), with respect to the true profile  $\mathbf{x}$  [1]. Since from Eqs. (2), (9) and (10) it results that:

$$\frac{\partial \hat{\mathbf{a}}}{\partial \mathbf{x}} = \Lambda^{-1} \mathbf{U}^T \mathbf{S}_y^{-1/2} \mathbf{K} = \Lambda^{-1} \mathbf{U}^T \mathbf{U} \Lambda \mathbf{V}^T = \mathbf{V}^T, \quad (17)$$

from Eq. (15) we obtain that the averaging kernel matrix of the estimation of the profile is equal to:

$$\mathbf{A}_x = \frac{\partial \hat{\mathbf{x}}}{\partial \mathbf{x}} = \left[ \mathbf{I} - \mathbf{W}(\mathbf{W}^T \mathbf{R} \mathbf{W})^{-1} \mathbf{W}^T \mathbf{R} \right] \mathbf{V} \mathbf{V}^T. \quad (18)$$

### 3. Application of the method to a simulated case

In this section we provide some examples of the new procedure in the case of simulated measurements of the MIPAS (Michelson Interferometer for Passive Atmospheric Sounding) instrument [14] that flies aboard Envisat (ENVironmental SATellite). MIPAS is a Fourier-transform spectrometer operating in the middle infrared that observes the atmospheric emission at the limb for the retrieval of the vertical profiles of several minor atmospheric constituents. The code adopted by the European Space Agency for the operational retrieval [15,16] uses a non-linear least-squares fit of the observed spectra with forward model simulations to retrieve the vertical profiles of pressure, temperature, water vapor, ozone, nitric acid, methane, nitrous oxide and nitrogen dioxide between 6 and 68 km of altitude.

The simulation is made in the case of a North pole climatological atmosphere of July with a modified ozone volume mixing ratio (VMR) profile  $\mathbf{x}$ . The tangent altitudes for which the spectra are simulated are those corresponding to the MIPAS measurement scenario adopted between July 2002 and March 2004, that is with a 3 km step between 6 and 42 km, with a 5 km step between 42 and 52 km and with an 8 km step between 52 and 68 km (for a total of 17 tangent altitudes). The simulated observations  $\mathbf{y}$  are obtained adding a realistic random noise to the radiances calculated with the forward model. The microwindow approach, described in [17], is adopted for the retrieval and of the 17 spectra only a subset of 4231 spectral points that contain the maximum information on ozone profile is used. The variances of the random noise and the apodization that is applied to the observed spectra allow building the VCM  $\mathbf{S}_y$ .

The method described in Section 2 requires the definition of a predefined vertical grid on which to represent the ozone profile  $\mathbf{x}$  and of a linearization point  $\mathbf{x}_0$  close enough to the true profile in such a way that the linear approximation of the forward model is appropriate. The predefined grid can be chosen as fine as wished and, thanks to this freedom, it can be determined on the basis of the application rather than according to the vertical resolution of the measurements. In this example we have chosen a vertical grid of 1 km step between 0 and 100 km. The linearization point  $\mathbf{x}_0$  has been obtained interpolating at the predefined grid the ozone profile retrieved by the operational retrieval at the tangent altitudes grid. Accordingly,

we have calculated the forward model and the Jacobian in the linearization point (respectively  $\mathbf{F}(\mathbf{x}_0)$  and  $\mathbf{K}$ ). With these quantities we have all the ingredients needed to apply the method described in Section 2.

The SVD of  $\mathbf{S}_y^{-1/2} \mathbf{K}$  (a matrix of dimension 4231x101) determines 98 singular values different from zero that indicate a measurement space of dimension  $p=98$ . The large rank of the matrix comes out from the SVD operation; however only few components correspond to large singular values (we recall that, according to Eq. (12), components corresponding to small singular values are determined with a large error). All the 98 measured components with their errors can be considered with no harm in successive data assimilation or data fusion operations (because in these applications the components will be weighted according to their errors), however if we want to make a graphical representation of the profile it is preferable not to include components with large errors because they introduce large uncertainties in the values of the profile. In this case, as suggested in Subsection 2.4, it is preferable to reduce the dimension  $p$  of the measurement space and to consider the components affected by a large error as belonging to the null space.

The result of this operation is reported in panel (a) of Fig. 1 which shows the  $\hat{\mathbf{x}}$  profile (red line) obtained with the procedure described in Section 2 in the case of considering only the 18 largest singular values. In the same panel the two components  $\hat{\mathbf{x}}_a$  (blue line) and  $\hat{\mathbf{x}}_b$  (green line) in the measurement space and in the null space, respectively, of  $\hat{\mathbf{x}}$  are shown, illustrating at each altitude the relative importance of MSS and NSR contributions. In panel (b) the differences between  $\hat{\mathbf{x}}$  and the *true profile*  $\mathbf{x}_{\text{true}}$  used to generate the observations (red line) and between  $\hat{\mathbf{x}}_a$  and  $\mathbf{x}_{\text{true}}$  (blue line) are shown on an expanded scale.

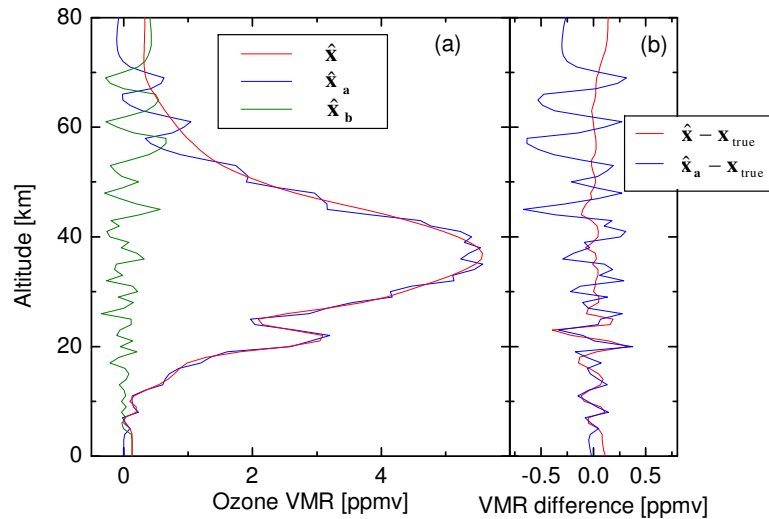


Fig. 1. (a)  $\hat{\mathbf{x}}$  profile (red line) obtained with the procedure described in Section 2 in the case of considering only the 18 largest singular values. Also the two components of  $\hat{\mathbf{x}}$  in the space measurement (blue line) and in the null space (green line) are shown. (b) Differences between  $\hat{\mathbf{x}}$  and  $\mathbf{x}_{\text{true}}$  (red line) and between  $\hat{\mathbf{x}}_a$  and  $\mathbf{x}_{\text{true}}$  (blue line).



The  $\hat{\mathbf{x}}_a$  profile obtained with the MSS method is characterized by oscillations and the addition of the  $\hat{\mathbf{x}}_b$  component obtained with the NSR method is able to remove them and to decrease significantly the difference between the retrieved profile and the true profile.

As expected, the contribution of  $\hat{\mathbf{x}}_b$  is predominant at altitudes larger than 68 km (the highest measured tangent altitude), below 6 km (the lowest measured tangent altitude) and at some altitudes between 52 and 68 km where the step of the measured tangent altitudes is the largest (8 km).

The largest values of the residual difference are observed at around 24 km of altitude, where a better vertical resolution would be needed in order to resolve the variation of the profile that occurs on a smaller scale with respect to both the step of the measurement tangent altitudes and the field of view of the instrument (that are equal to about 3 km).

In order to identify the optimal number of the largest singular values to consider for a good graphical representation of the profile the trade-off between the noise error and the smoothing error [1] has to be analyzed. A too large number of singular values implies that the noise affecting the observations can propagate with a large amplification into the retrieved profile, with the consequence of a too large retrieval error. A too small number of singular values reduces the capability of the retrieved profile to reproduce the shape of the real profile. These effects are verified in Fig. 2 where the true ozone profile (black line) is compared with the profile  $\hat{\mathbf{x}}$  obtained with the procedure described in Section 2 in the case of considering the 10 (blue line) and the 30 (red line) largest singular values.

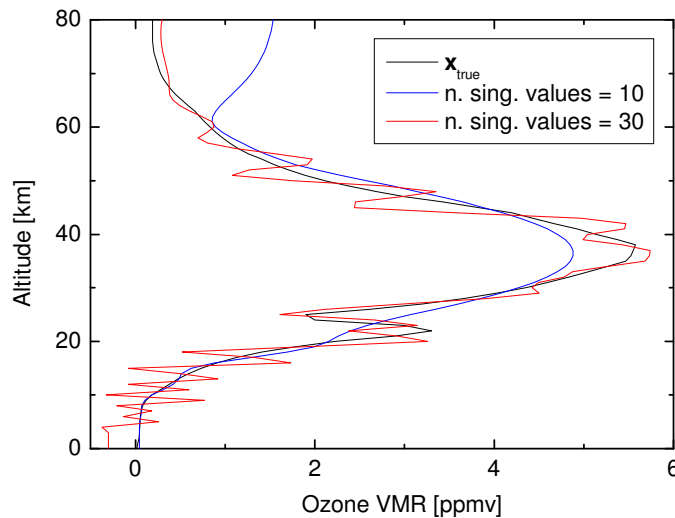


Fig. 2. True ozone profile (black line) and  $\hat{\mathbf{x}}$  profile obtained with the procedure described in Section 2 in the case of considering the 10 (blue line) and the 30 (red line) largest singular values.

The use of only the 10 largest singular values does not allow reproducing the feature at 24 km of altitude and also the correct value of the maximum of ozone, on the other hand the use of the 30 largest singular values determines the presence of large oscillations due to the noise amplification introduced by the ill-conditioned components.

The trade-off associated with the choice of how many largest singular values to maintain in the MSS solution is further analyzed in Fig. 3 which shows the comparison between the noise error (black line) and the smoothing error (red line), averaged on all the altitudes, as a function of the number of considered singular values. The noise errors have been calculated taking the root square of the diagonal elements of  $\mathbf{S}_x$  given by Eq. (16), while the smoothing errors have been calculated as the absolute value of the differences between the profile obtained with the procedure described in Section 2, where  $\mathbf{a}$  is calculated projecting directly the true profile on the basis represented by the columns of  $\mathbf{V}$ , and the true profile.

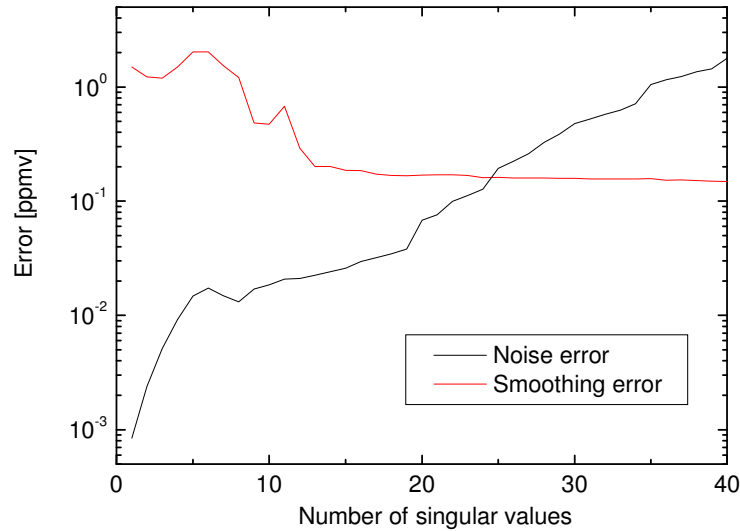


Fig. 3. Comparison between the noise error (black line) and the smoothing error (red line), averaged on all the altitudes, as a function of the number of considered singular values.

As expected, the noise error increases with the number of considered singular values while the smoothing error decreases with it. Above a certain value of the number of singular values a slow decrease of the smoothing error is observed as a consequence of the fact that in the true profile there are features at high altitude that are not caught by the singular values considered in the plot. A number of considered singular values that corresponds to a good compromise between the two errors is that for which the sum of the two errors is minimum. In our case the minimum is obtained considering 18 singular values and the corresponding  $\hat{\mathbf{x}}$  profile is shown in Fig. 1. It is interesting to notice that the value 18 is larger than the number of measured tangent altitudes, which is equal to 17, but not significantly different. This confirms that also in the operational retrieval [15, 16] a good optimization has been made with the choice of a retrieval grid with points on the 17 tangent altitudes.

In Fig. 4 the difference between  $\hat{\mathbf{x}}$  and  $\mathbf{x}_{\text{true}}$  (black line) is compared with the difference between the profile obtained by the operational retrieval interpolated on the predefined grid (that in our case coincides with the linearization point  $\mathbf{x}_0$ ) and  $\mathbf{x}_{\text{true}}$  (red line). The two methods (RMSS and operational retrieval) have comparable quality even if the first better reproduces the values at high altitude and the feature at 24 km while the second does better reproduce the values below 20 km. The differences are due to the fact that the two methods provide a slightly different compromise between vertical resolution and precision; however the general agreement shows that the NSR has provided an efficient regularization at least as good as the constraint of a limited number of retrieval points.

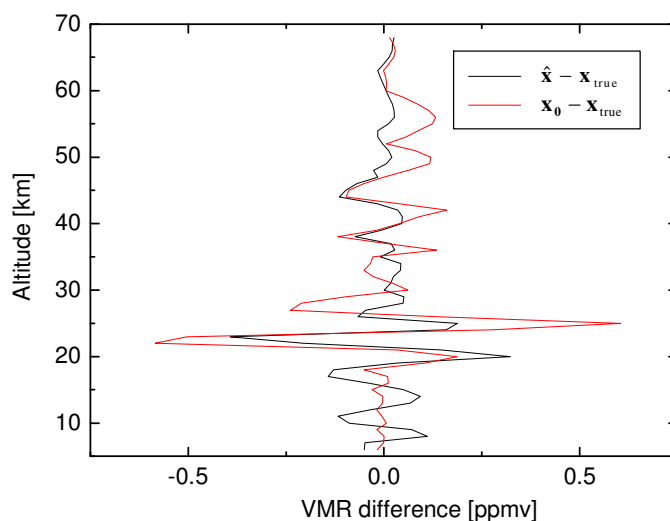


Fig. 4. Difference between  $\hat{\mathbf{x}}$  and  $\mathbf{x}_{\text{true}}$  (black line) compared with the difference between the profile obtained by the operational retrieval interpolated on the predefined grid (that in our case coincides with the linearization point  $\mathbf{x}_0$ ) and  $\mathbf{x}_{\text{true}}$  (red line).

In Subsection 2.3 we have seen that, when the retrieved information is used for subsequent operations, it is better not to provide information about the null-space components, while no constraint (apart from the opportunity of limiting the amount of data) exists as far as the number of provided measurement-space components is concerned. In this case the MSS solution is the ideal product, since it contains all the needed information without useless null space assumptions. On the other hand, as shown in Fig. 3, when the retrieved information is used for the graphical representation of the profile, the conflicting requirements of representation in a complete space and of minimum error impose that some criterion is used for the determination of the optimal number of measurement-space components and that the NSR is applied to the complementary null space.

With the algebra discussed in Section 2 a single procedure can be used to provide the optimal profile for further operation (MSS) as well as the optimal profile for a graphical representation (RMSS). Both profiles can be represented on the same predefined grid and no measurement constraint exists about the choice of this grid. The possibility of using a fine predetermined grid is a major advantage, because it can prevent the approximations that are introduced by interpolations.

#### 4. Application to data fusion

The theory developed in Section 2 also provides a framework for the handling of the data fusion problem. When several independent measurements of the same profile are available there is the problem of how to combine the different measurements in order to exploit all the available information for a more comprehensive and accurate description of the atmospheric state. Generally the different measurements are represented on different grids (defined by the different observation and retrieval methods) and their combination implies some interpolation that produces a loss of information [3]. Furthermore, it is possible that the measurements contain a-priori information that may introduce a bias in the fused data. Interpolations and biases are the two main problems encountered in data fusion. The method described in Section 2 removes these two problems. Indeed, the MSS provides a representation of the profile without a bias due to a-priori information and on a predetermined grid as fine as one likes so

that interpolation is no longer needed. A further advantage of the MSS solution is that it does not depend on procedure-dependent quantities, such as  $\mathbf{x}_0$ , so that it is not necessary to maintain memory of them.

In the light of these advantages of the MSS, in this Section we analyze how the MSS can be used for the fusion of two independent indirect measurements of the same profile  $\mathbf{x}$ . The results of this section can be easily extended to an arbitrary number of independent measurements.

Since the predetermined grid can be chosen freely and be common to the two measurements, we can assume that the two measurement spaces are subspaces of the same space  $\mathbb{R}^n$ . The two MSSs are characterized by the matrices  $\mathbf{V}_1$  and  $\mathbf{V}_2$  that identify the measurement spaces (of dimensions  $p_1$  and  $p_2$  respectively) and by the two vectors  $\hat{\mathbf{a}}_1$  and  $\hat{\mathbf{a}}_2$  (with their VCM  $\mathbf{S}_{a_1}$  and  $\mathbf{S}_{a_2}$ ). Each MSS provides the elements of the relationship shown in Eq. (10). We can write the two relationships in the compact form:

$$\begin{pmatrix} \hat{\mathbf{a}}_1 \\ \hat{\mathbf{a}}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T \end{pmatrix} \mathbf{x} + \begin{pmatrix} \boldsymbol{\varepsilon}_{a_1} \\ \boldsymbol{\varepsilon}_{a_2} \end{pmatrix}, \quad (19)$$

where the notation  $\begin{pmatrix} \mathbf{P} \\ \mathbf{Q} \end{pmatrix}$  means the matrix (vector) obtained arranging the rows of the matrix (vector)  $\mathbf{Q}$  below the rows of the matrix (vector)  $\mathbf{P}$  and  $\boldsymbol{\varepsilon}_{a_1}$  and  $\boldsymbol{\varepsilon}_{a_2}$  contain the errors with which  $\mathbf{a}_1 = \mathbf{V}_1^T \mathbf{x}$  and  $\mathbf{a}_2 = \mathbf{V}_2^T \mathbf{x}$  are estimated by  $\hat{\mathbf{a}}_1$  and  $\hat{\mathbf{a}}_2$ .

Eq. (19) implies that  $\begin{pmatrix} \hat{\mathbf{a}}_1 \\ \hat{\mathbf{a}}_2 \end{pmatrix}$  are the measurements of  $\mathbf{x}$  in the space generated by the columns of  $\mathbf{V}_1$  and of  $\mathbf{V}_2$ . This space is in the union space of the two measurement spaces and the profile that determined by these measurements is the result of the data fusion. This problem is equal to that faced in Section 2 consisting in finding a representation of the profile  $\mathbf{x}$  of  $\mathbb{R}^n$  when the component in a subspace is measured; accordingly, also the same procedure is followed in order to determine the solution. We can represent the profile  $\mathbf{x}$  in the form of the summation of a vector of the union space of the two measurement spaces and of a vector of the orthogonal complement space to this space (which coincides with the intersection space of the two null spaces of the original measurements). The vector in the union space is the MSS of the data fusion problem and its identification implies the determination of the elements of the following relationship:

$$\hat{\mathbf{a}}_{12} = \mathbf{V}_{12}^T \mathbf{x} + \boldsymbol{\varepsilon}_{a_{12}}, \quad (20)$$

where  $\hat{\mathbf{a}}_{12}$  is the estimation of the components of  $\mathbf{x}$  in the basis of the union space represented by the matrix  $\mathbf{V}_{12}$ , and  $\boldsymbol{\varepsilon}_{a_{12}}$  is its error.

As in Section 2 we transform the vector  $\begin{pmatrix} \hat{\mathbf{a}}_1 \\ \hat{\mathbf{a}}_2 \end{pmatrix}$  in such a way that the VCM becomes the unity matrix. This can be obtained multiplying Eq. (19) by the matrix  $\begin{pmatrix} \mathbf{S}_{a_1}^{-1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{a_2}^{-1/2} \end{pmatrix}$ , obtaining:

$$\begin{pmatrix} \mathbf{S}_{\mathbf{a}_1}^{-1/2} \hat{\mathbf{a}}_1 \\ \mathbf{S}_{\mathbf{a}_2}^{-1/2} \hat{\mathbf{a}}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{S}_{\mathbf{a}_1}^{-1/2} \mathbf{V}_1^T \\ \mathbf{S}_{\mathbf{a}_2}^{-1/2} \mathbf{V}_2^T \end{pmatrix} \mathbf{x} + \begin{pmatrix} \mathbf{S}_{\mathbf{a}_1}^{-1/2} \boldsymbol{\varepsilon}_{\mathbf{a}_1} \\ \mathbf{S}_{\mathbf{a}_2}^{-1/2} \boldsymbol{\varepsilon}_{\mathbf{a}_2} \end{pmatrix}. \quad (21)$$

If we perform the SVD of the kernel of Eq. (21) we obtain:

$$\begin{pmatrix} \mathbf{S}_{\mathbf{a}_1}^{-1/2} \mathbf{V}_1^T \\ \mathbf{S}_{\mathbf{a}_2}^{-1/2} \mathbf{V}_2^T \end{pmatrix} = \mathbf{U}_{12} \boldsymbol{\Lambda}_{12} \mathbf{V}_{12}^T, \quad (22)$$

where  $\mathbf{U}_{12}$  is a matrix of dimension  $(p_1+p_2) \times p_{12}$ , with  $p_{12} \leq (p_1+p_2)$ ,  $\boldsymbol{\Lambda}_{12}$  is a nonsingular diagonal matrix of dimension  $p_{12} \times p_{12}$  and  $\mathbf{V}_{12}$  is a matrix of dimension  $n \times p_{12}$  whose columns are an orthonormal basis of the space generated by the rows of  $\begin{pmatrix} \mathbf{S}_{\mathbf{a}_1}^{-1/2} \mathbf{V}_1^T \\ \mathbf{S}_{\mathbf{a}_2}^{-1/2} \mathbf{V}_2^T \end{pmatrix}$ . Since

$\mathbf{S}_{\mathbf{a}_1}^{-1/2}$  and  $\mathbf{S}_{\mathbf{a}_2}^{-1/2}$  are nonsingular matrices, the space generated by the rows of  $\begin{pmatrix} \mathbf{S}_{\mathbf{a}_1}^{-1/2} \mathbf{V}_1^T \\ \mathbf{S}_{\mathbf{a}_2}^{-1/2} \mathbf{V}_2^T \end{pmatrix}$  coincides with the space generated by the columns of  $\mathbf{V}_1$  and of  $\mathbf{V}_2$ .

Consequently the columns of  $\mathbf{V}_{12}$  are an orthonormal basis of the union space of the two measurement spaces.

The knowledge of  $\mathbf{V}_{12}$  can be used to determine the vector  $\hat{\mathbf{a}}_{12}$ . Substituting Eq. (22) in Eq. (21) and multiplying both terms on the left by  $\boldsymbol{\Lambda}_{12}^{-1} \mathbf{U}_{12}^T$  we obtain:

$$\hat{\mathbf{a}}_{12} = \mathbf{V}_{12}^T \mathbf{x} + \boldsymbol{\varepsilon}_{\mathbf{a}_{12}} = \boldsymbol{\Lambda}_{12}^{-1} \mathbf{U}_{12}^T \begin{pmatrix} \mathbf{S}_{\mathbf{a}_1}^{-1/2} \hat{\mathbf{a}}_1 \\ \mathbf{S}_{\mathbf{a}_2}^{-1/2} \hat{\mathbf{a}}_2 \end{pmatrix}, \quad (23)$$

where  $\boldsymbol{\varepsilon}_{\mathbf{a}_{12}}$  is given by:

$$\boldsymbol{\varepsilon}_{\mathbf{a}_{12}} = \boldsymbol{\Lambda}_{12}^{-1} \mathbf{U}_{12}^T \begin{pmatrix} \mathbf{S}_{\mathbf{a}_1}^{-1/2} \boldsymbol{\varepsilon}_{\mathbf{a}_1} \\ \mathbf{S}_{\mathbf{a}_2}^{-1/2} \boldsymbol{\varepsilon}_{\mathbf{a}_2} \end{pmatrix} \quad (24)$$

and is characterized by the VCM:

$$\mathbf{S}_{\mathbf{a}_{12}} = \boldsymbol{\Lambda}_{12}^{-2}. \quad (25)$$

The solution of the data fusion problem consists in the determination of the components of  $\mathbf{x}$  in the union space of the two measurement spaces and in performing a regularisation in the orthogonal complement space of the union space. Eq. (23) and  $\mathbf{V}_{12}$  obtained from Eq. (22) provide the MSS of  $\mathbf{x}$  in the union space of the two measurements. The undetermined component of  $\mathbf{x}$  lies in the orthogonal complement space of the union space, and can be determined following the procedure described in subsection 2.4. Finally the complete expression of the solution and its characterisation can be obtained as described in subsection 2.5.

Since the space characterized by  $V_{12}$  exclusively includes the two original measurement spaces, the MSS in the union space has been obtained by utilizing all the information provided by the two sets of observations and without any a-priori information. This characteristic makes the proposed method for data fusion optimal and equivalent to the simultaneous analysis of the two sets of observations.

## 5. Conclusions

A procedure for the optimal exploitation and representation of the vertical profile retrieved from atmospheric observations has been discussed. The optimal exploitation implies to extract all the information contained in the observations and to make it fully available to subsequent application, such as data comparison, data assimilation and data fusion. To this purpose the retrieved profile must be without any a-priori information and be represented on a vertical grid as fine as needed both for its exhaustive description and for the prevention of subsequent interpolation approximations. We have shown how this result is obtained with the *measurement-space solution* (MSS) which gives the retrieved profile in terms of its components in the subspace defined by the rows of the jacobian matrix. The determination of the MSS coincides with the method proposed by Joiner et al. [12] for data assimilation and leaves completely undetermined the component of the retrieved profile in the null space (that is the complement space to the measurement space).

The MSS, while being optimal for further processing because comprehensive, unbiased by a-priori information and with a diagonal VCM, is not suitable for a graphical representation of the retrieved profile. Indeed, using simulations applied to a test case of ozone retrieval with the MIPAS instrument, we have shown that if the dimension of the vertical retrieval grid is chosen with the desirable redundancy, components of the MSS that correspond to small singular values are poorly determined and the retrieval problem may become ill-conditioned. On the other hand, if the poorly determined components are removed, the size of the null space grows at the expenses of the measurement space and the retrieved profile acquires a rather unphysical shape. This unphysical shape is frequently observed in retrieval techniques and is due to the fact that in a graphical representation the result is presented in a complete space. In a complete space the representation of the profile with only the MSS component corresponds to the selection of the value zero for the null-space component. This is a particular choice among the infinitive possible ones, which does not necessarily provide the best graphical representation.

The above considerations underline the importance of choosing an adequate null-space component when a graphical representation is made. To this purpose we have presented the *null-space regularization* (NSR) method for the calculation of the null-space component which allows to obtain the smoothest profile compatible with the observations. This is a regularization and is different from the other regularizations since it does not affect the measured component and only determines the one that has not been measured. The fact that two components (the MSS and the NSR) are determined respectively in the measurement space and in the null space provides a clear and easily traceable distinction between the measurement and the constraint. Simulations have shown that MSS plus NSR, that is referred to as the *regularized measurement-space solution* (RMSS), provides a good graphical representation of the retrieved profile.

Finally, the procedure developed for the calculation of the RMSS has been applied to data fusion problems. A method has been presented that allows, from the MSS of two or more independent measurements, the calculation of the fused profile with the determination of its MSS in the union space (that is the space obtained merging the measurement spaces of the original measurements). The components of the profile in the orthogonal complement of the union space can be calculated with the NSR method. The sum of the MSS component with the NSR component provides the RMSS of the fused profile. This RMSS is the optimal data fusion in the sense that the result is the same as it would have been obtained with the simultaneous analysis of the measurements.

These results show that the MSS, combined with the NSR, provides both an optimal exploitation and an optimal representation of the vertical profile retrieved from atmospheric observations.