Optimal variables for retrieval products of vertical profile measurements

Simone Ceccherini

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

Consiglio Nazionale delle Ricerche

IFAC

esa

ASI · CNR · INAF EARTH · MOON · MARS

INTRODUCTION

In the last years the number of satellite instruments that are sounding the atmosphere is increasing at a high rate, therefore, it is very likely that more instruments will simultaneously measure vertical profiles corresponding to locations that are close in time and space. In this case, the different retrieved profiles can be combined into a single product that includes all the available information and this combination is referred to as *data fusion* [1,2]. Consequently, users will find more convenient to consult fused products rather than individual measurements. In the light of the increased requirement of fused products, **we consider the possibility of using new variables representing the retrieval products, with the purpose of simplifying the subsequent fusion processes**.

Generally, in order to make a complete use of the products in further processing such as data fusion or data assimilation, the retrieval products are represented by means of the retrieved profile, the averaging kernel matrix (AKM), the retrieval covariance matrix (CM) and the a priori information used in the retrieval.

We propose new variables [3], calculated starting from these standard retrieval products, to save the information provided by the measurements in the retrieval products. These new variables have several advantages with respect to the standard quantities and we analyze these advantages on a theoretical basis. **The change of the retrieval products is proposed in the perspective of developing a shared formalism, which facilitates the interface between data providers and data users, while ensuring a full exploitation of the available information**.

OPTIMAL ESTIMATION METHOD

The retrieval of the vertical profile of an atmospheric parameter requires the solution of an inverse problem that is often ill-posed and in order to obtain a stable solution, some a priori information has to be added in the retrieval process. A largely used method to retrieve atmospheric parameters using remote sensing is the *optimal estimation method*, where the a priori information is represented by an a priori profile $\boldsymbol{x}_a$ and by an a priori CM $\mathbf{S}_a$ of the unknown parameter and the solution is given by the profile corresponding to the maximum a posteriori probability calculated with the Bayes theorem.

We assume to have retrieved the vertical profile $\hat{\boldsymbol{x}}$ of an atmospheric parameter from a set of observations (radiances) $\boldsymbol{y}$ with the optimal estimation method. We indicate with $\boldsymbol{f}(\boldsymbol{x})$ the forward model, which allows to express the observations $\boldsymbol{y}$ as a function of the true profile $\boldsymbol{x}_t$:

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}_t) + \boldsymbol{\varepsilon}$$

$\boldsymbol{\varepsilon}$: noise errors with CM $\mathbf{S}_y$

$\boldsymbol{x}_t$: true profile

The retrieved profile is characterized by the AKM and CM:

$$\mathbf{A} = \left(\mathbf{F} + \mathbf{S}_a^{-1}\right)^{-1}\mathbf{F} \qquad \mathbf{S} = \left(\mathbf{F} + \mathbf{S}_a^{-1}\right)^{-1}$$

$$\mathbf{F} = \mathbf{K}^T\mathbf{S}_y^{-1}\mathbf{K}$$

$\mathbf{F}$: Fisher information matrix

$\mathbf{K}$: Jacobian of $\boldsymbol{f}(\boldsymbol{x})$

$\mathbf{F}$ quantifies the information provided by the observations $\boldsymbol{y}$ about the retrieved vertical profile.

$\mathbf{F}$ depends on the a priori information used in the retrieval through $\mathbf{K}$ calculated at $\hat{\boldsymbol{x}}$, which depends on $\boldsymbol{x}_a$ and $\mathbf{S}_a$. Therefore, the dependence of $\mathbf{F}$ on the a priori information is due to the second order terms in the expansion of the forward model as a function of the profile $\boldsymbol{x}$ and consequently, when the linear approximation of the forward model is valid, $\mathbf{F}$ is independent of the a priori information.

LINEARIZATION OF THE TRANSFER FUNCTION AND VARIABLES $\boldsymbol{\alpha}$

We can consider the whole measuring system, including both the observing system and the retrieval method, as an operation that transforms the true profile $\boldsymbol{x}_t$ into the retrieved profile $\hat{\boldsymbol{x}}$ and, accordingly, define the retrieved profile $\hat{\boldsymbol{x}}$ as a function of the true profile $\boldsymbol{x}_t$. This function is referred to as *transfer function* and besides being a function of $\boldsymbol{x}_t$ is also a function of the noise errors $\boldsymbol{\varepsilon}$ of the observations $\boldsymbol{y}$. This dependence can be seen recalling that really $\hat{\boldsymbol{x}}$ depends on $\boldsymbol{x}_t$ through the observations $\boldsymbol{y}$, therefore, we can write

$$\hat{\boldsymbol{x}} = \hat{\boldsymbol{x}}(\boldsymbol{y}) = \hat{\boldsymbol{x}}\left(\boldsymbol{f}(\boldsymbol{x}_t) + \boldsymbol{\varepsilon}\right)$$

Expanding the transfer function at the first order around the a priori profile $\boldsymbol{x}_t = \boldsymbol{x}_a$ and zero errors $\boldsymbol{\varepsilon} = \boldsymbol{0}$, we obtain:

$$\hat{\boldsymbol{x}} = \boldsymbol{x}_a + \mathbf{A}\left(\boldsymbol{x}_t - \boldsymbol{x}_a\right) + \mathbf{G}\boldsymbol{\varepsilon}$$

$\mathbf{G}$: Gain matrix

$$\mathbf{G} = \left(\mathbf{F} + \mathbf{S}_a^{-1}\right)^{-1}\mathbf{K}^T\mathbf{S}_y^{-1}$$

We define the vector $\boldsymbol{\alpha}$:

$$\boldsymbol{\alpha} = \hat{\boldsymbol{x}} - \boldsymbol{x}_a + \mathbf{A}\boldsymbol{x}_a$$

$$\Rightarrow \quad \boldsymbol{\alpha} = \mathbf{A}\boldsymbol{x}_t + \mathbf{G}\boldsymbol{\varepsilon}$$

This equation shows that:

- $\boldsymbol{\alpha}$ is the measurement of the true profile made using the rows of $\mathbf{A}$ as weighting functions.
- $\mathbf{G}\boldsymbol{\varepsilon}$ is the vector that includes the errors of this measurement.
- In the linear approximation of the forward model, $\boldsymbol{\alpha}$ (differently from $\hat{\boldsymbol{x}}$) is independent of the a priori profile $\boldsymbol{x}_a$, however, through the expressions of $\mathbf{A}$ and $\mathbf{G}$, maintains the dependence on the a priori CM $\mathbf{S}_a$.

THE NEW VARIABLES $\boldsymbol{\beta}$

We define the vector $\boldsymbol{\beta}$ as:

$$\boldsymbol{\beta} = \mathbf{S}^{-1}\boldsymbol{\alpha} = \mathbf{S}^{-1}\left(\hat{\boldsymbol{x}} - \boldsymbol{x}_a + \mathbf{A}\boldsymbol{x}_a\right)$$

$$\Rightarrow \quad \boldsymbol{\beta} = \mathbf{F}\boldsymbol{x}_t + \boldsymbol{\delta} \qquad \boldsymbol{\delta} = \mathbf{K}^T\mathbf{S}_y^{-1}\boldsymbol{\varepsilon}$$

This equation shows that:

- $\boldsymbol{\beta}$ is the measurement of the true profile made using the rows of $\mathbf{F}$ as weighting functions.
- $\boldsymbol{\delta}$ is the vector that includes the errors of this measurement.
- **In the linear approximation of the forward model, $\boldsymbol{\beta}$ is uniquely determined independently of both $\boldsymbol{x}_a$ and $\mathbf{S}_a$**.

We can calculate the AKM and CM of $\boldsymbol{\beta}$:

$$\mathbf{A}_{\beta} = \frac{\partial\boldsymbol{\beta}}{\partial\boldsymbol{x}_t} = \mathbf{F} \qquad \mathbf{S}_{\beta} = \left\langle\left(\boldsymbol{\beta} - \langle\boldsymbol{\beta}\rangle\right)\left(\boldsymbol{\beta} - \langle\boldsymbol{\beta}\rangle\right)^T\right\rangle = \left\langle\boldsymbol{\delta}\boldsymbol{\delta}^T\right\rangle = \mathbf{K}^T\mathbf{S}_y^{-1}\left\langle\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\right\rangle\mathbf{S}_y^{-1}\mathbf{K} = \mathbf{F}$$

Therefore, both the AKM and the CM of $\boldsymbol{\beta}$ coincide with the Fisher information matrix $\mathbf{F}$.

The dimensions of $\boldsymbol{\beta}$ are the inverse of the dimensions of $\hat{\boldsymbol{x}}$: $[\boldsymbol{\beta}] = [\hat{\boldsymbol{x}}]^{-1}$, therefore, $\boldsymbol{\beta}$ does not represent a profile of the parameter that we aim to retrieve. However, this is not a problem, because the objective of the retrieval products is no longer the graphical representation of the profile, but to efficiently provide all the information of the observations to subsequent data analyses.

POSSIBLE VARIABLES TO SAVE THE INFORMATION IN THE RETRIEVAL PRODUCTS

Variable	$\hat{\boldsymbol{x}}$
AKM	$\mathbf{A}$
CM	$\mathbf{S}$
Dependence on the a priori information	$\boldsymbol{x}_a$ and $\mathbf{S}_a$

Variable	$\boldsymbol{\alpha}$
AKM	$\mathbf{A}$
CM	$\mathbf{G}\mathbf{S}_y\mathbf{G}^t$
Dependence on the a priori information	$\mathbf{S}_a$

Variable	$\boldsymbol{\beta}$
AKM	$\mathbf{F}$
CM	$\mathbf{F}$
Dependence on the a priori information	Independent

ADVANTAGES OF THE USE OF THE VARIABLES $\boldsymbol{\beta}$

Representation of the profile using any constraint

We can express $\hat{\boldsymbol{x}}$ as a function of $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{x}} = \left(\mathbf{F} + \mathbf{S}_a^{-1}\right)^{-1}\left(\boldsymbol{\beta} + \mathbf{S}_a^{-1}\boldsymbol{x}_a\right)$$

This equation can be used to recover the original retrieved profile using the a priori information $\boldsymbol{x}_a$ and $\mathbf{S}_a$ used in the retrieval procedure, but since in the linear approximation of the forward model $\mathbf{F}$ and $\boldsymbol{\beta}$ are independent of the a priori information, in this approximation, **this equation can be used to produce a profile with any a priori information we like**.

Data fusion

The complete data fusion (CDF) formula of N independent measurements $\hat{\boldsymbol{x}}_i$ of the same vertical true profile $\boldsymbol{x}_t$, obtained with the optimal estimation method, is given by [2]:

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

This formula can be expressed using the quantities $\boldsymbol{\beta}$ and $\mathbf{F}$:

$$\boldsymbol{x}_f = \left(\sum_{i=1}^{N}\mathbf{F}_i + \mathbf{S}_a^{-1}\right)^{-1}\left(\sum_{i=1}^{N}\boldsymbol{\beta}_i + \mathbf{S}_a^{-1}\boldsymbol{x}_a\right)$$

This equation shows that $\boldsymbol{\beta}$ and F are the only quantities needed to perform the data fusion of a set of measurements.

Reduction of the data volume

Data volume of the stored quantities when using the standard retrieval products and the new variables for a profile of n components:

Standard products		New variables	
Quantities	Number of values	Quantities	Number of values
$\hat{\boldsymbol{x}}$	n	$\boldsymbol{\beta}$	n
$\mathbf{A}$	n^2	$\mathbf{F}$	$n(n+1)/2$
$\mathbf{S}$	$n(n+1)/2$		
$\boldsymbol{x}_a$	n		
Total number of values			
$(3n^2+5n)/2$		$(n^2+3n)/2$	

Since the main storage requirement is due to the square term, **the use of the variables $\boldsymbol{\beta}$ allows to reduce to about one-third the stored data volume with respect to the use of the standard retrieval products**.

CONCLUSIONS

We propose new variables to represent the retrieval products that have several advantages with respect to the standard retrieval products.

- In the linear approximation of the forward model, they are independent of the a priori information used in the retrieval.
- They can be used to represent the vertical profile with an a priori information selected by the user.
- They can be directly used to perform the data fusion of a set of measurements performed with different instruments.
- They allow to reduce to about one-third the stored data volume with respect to the use of standard retrieval products.

These properties of the new variables make them a perfect retrieval product when further processing is performed by the users.

REFERENCES

[1] S. Ceccherini, B. Carli and P. Raspollini, *Equivalence of data fusion and simultaneous retrieval*, Optics Express, **23**, 8476-8488 (2015), https://doi.org/10.1364/OE.23.008476

[2] S. Ceccherini, N. Zoppetti, and B. Carli, *An improved formula for the complete data fusion*, Atmospheric Measurement Techniques, **15**, 7039-7048 (2022), https://doi.org/10.5194/amt-15-7039-2022

[3] S. Ceccherini, *Optimal Variables for Retrieval Products*, Atmosphere, **15**, 506 (2024), https://doi.org/10.3390/atmos15040506

ACKNOWLEDGMENTS

The results reported in the poster were obtained in the context of the Earth-Moon-Mars (EMM) project, led by INAF in partnership with ASI and CNR, funded under the National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 3.1: "Fund for the realisation of an integrated system of research and innovation infrastructures" - Action 3.1.1 funded by the European Union - NextGenerationEU.

Finanziato dall'Unione europea NextGenerationEU | Ministero dell'Università e della Ricerca | Italiadomani